Essays in Macro-finance

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Dmitry Kuvshinov

aus Novosibirsk, Russland

Bonn, 2019

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Moritz Schularick
Zweitreferent:	Prof. Dr. Gernot Müller
Tag der mündlichen Prüfung:	26. Juni 2019

Acknowledgements

This dissertation has benefited from help and support of many different people. First and foremost, I would like to deeply thank my main supervisor, Moritz Schularick, for his invaluable feedback, guidance and support at every stage of my thesis, and for getting me interested in working with long-run data in the first place. A special thank you also goes out to my second supervisor Gernot Müller, who has offered many highly perceptive and valuable comments as well as fruitful and enjoyable research collaboration. Many thanks also to my other collaborators Òscar Jordà, Katharina Knoll, Alan Taylor, Martin Wolf and Kaspar Zimmermann, who have shared their helpful advice and feedback throughout a multitude of research projects and ideas. A special mention goes to Kaspar, who has not only been a fantastic person to work with, but also a great friend and an ever-present discussant and sounding board for all my research ideas and papers. All these people have also offered continuous help and support during the job search process at the end of my PhD, for which I am immensely grateful.

I would also like to thank other members of the current and former Bonn faculty – in particular Christian Bayer, Benjamin Born, Pavel Brendler, Narly Dwarkasing, Moritz Kuhn, Isabel Schnabel and Donghai Zhang – for their advice and help at various stages of my research. Special gratitude is reserved for my fellow former and current PhD students and junior department members Thilo Albers, Branislav Albert, Luis Bauluz, Yao Chen, Filip Novokhmet, Björn Richter and Felix Ward. I have taken great pleasure in talking about research and less serious things with them, and receiving helpful tips and advice throughout my PhD. A special mention goes out to Franciso Amaral, who has both helped me with the toughest data collection tasks for my various research projects, and provided helpful feedback on my research along the way. I also offer a warm thanks to my friends Zvonimir Bašić, Jonas Radbruch, Felix Schran and Louis Strang, without whom my time in Bonn would have been rather boring and mundane, and – even though this last claim is somewhat more contentions – also far less productive.

My work has also benefited from excellent research assistance in assembling the various new historical data series. I would like to thank Felix Rhiel, Lucie Stoppok, Maira Talha, Thomas Schwarz and Yevhenii Usenko for fantastic help in this matter.

I am grateful for financial support received from the Bundesministerium für Bildung und Forschung (BMBF) and the Deutscher Akademischer Austauschdienst (DAAD). I would especially like to thank Heide Baumung for making sure that I did actually receive this financial support and helping me deal with countless administrative matters throughout my years in Bonn.

Last but definitely not least, I offer great thanks to my fantastic family for toughing it out together throughout my research process, providing perspective and support during the low points and helping me celebrate the high points. A multitude of thanks are due to my mother Natalia and mother-in-law Gisela, who have both offered valuable experience from their own PhD years and provided great help and support with family matters throughout all stages of my PhD. The most important thank-you of all is, however, reserved for my wonderful wife Kristina and equally wonderful children Julia and Victor – although this definitely does not mean that they have to read any of my thesis.

Contents

List of Figures

List of Tables

Introduction

Macro-financial risk is all around us. It manifests itself in both day-to-day eventualities of how much money we earn through working or investing, and the more rare crisis events which can carry severe consequences of the loss of a job or an entire life's savings. The understanding of what drives these macro-financial fluctuations is, however, still incomplete. This incompleteness manifests itself in the multitude of asset pricing puzzles, such as the "equity premium puzzle" of Mehra and Prescott (1985) – the fact that investing in equities yields a high level of return relative to government bonds – and the "volatility puzzle" of Shiller (1981) – the fact that equity returns are much more volatile than the underlying fundamentals (corporate earnings or dividends). An active debate continues about whether these puzzles are a result of rational actions or behavioural biases, to the extent that the 2014 Economics Nobel prize was awarded to advocates of both behavioural and rational asset pricing theories (Fama, 2014; Shiller, 2014). It also remains unclear exactly how the behaviour of asset prices is linked to macroeconomic outcomes, with macroeconomic factors found to do poorly in explaining the level and variation in financial return premia, and asset price variation playing only a limited role in the business cycle dynamics (Campbell, 1999).

One key reason for these gaps in knowledge is the lack of reliable long-run macro-financial data. Most of the empirical regularities classified as puzzles are limited to the US equity market, with little known about other asset classes or countries. These knowledge gaps are especially important since macro-financial risk is highly dependent on rare extreme events such as financial crises and growth disasters (Barro, 2006; Nakamura, Steinsson, Barro, and Ursúa, 2013). A better understanding of the nature of macro-financial risk and its underlying drivers therefore requires long spans of data, which would ideally go beyond the well-explored territory of US equities.

This thesis seeks to advance our knowledge of what lies behind the risk and volatility in financial and macroeconomic outcomes, and how the two are related. The first contribution of the thesis lies in constructing new long-run data series on asset prices, returns and other financial indicators, which allow us to establish some basic facts about the empirical properties of macro-financial risk (Chapters 1–3). These new data allow us to study a rich selection of asset classes – from the better-known equity and government bond markets to the less explored markets for housing and corporate bonds – in greater detail, across more countries and over a longer time period than previously possible. The second contribution lies in analysing these data empirically to study the underlying causes and real-world consequences of financial risk and volatility (Chapters 2–4). Third, this thesis seeks to advance our theoretical knowledge of these macro-financial linkages through both examining existing theories (Chapter 2), and proposing new ways of thinking about the real effects of time-varying financial frictions (Chapter 5).

Chapter 1 – "The Rate of Return on Everything, 1870–2015", joint work with Òscar Jordà, Katharina Knoll, Moritz Schularick and Alan Taylor, forthcoming in the *Quartrly Journal of Economics* – deals with a very basic question of what are the returns on the major investable asset classes, the associated risk premiums, and the overall return on wealth itself. Before this work, little was known about the rate of return on the asset which forms the major part of household wealth portfolios – residential real estate. In this Chapter, we present a new database of returns on housing, as well as the three other major financial assets: government bills and bonds, and corporate equities. The data cover 16 advanced economies over the period 1870 to 2015, at annual frequency.

These new data yield several surprising findings. First, the returns to investing in residential real estate turn out to be high, around 7% p.a. in real terms, displaying a similar level but a lower volatility relative to those on equities. We, therefore, uncover a new "housing premium puzzle" which may be even more difficult to resolve than the equity premium puzzle of Mehra and Prescott (1985). Second, returns on safe assets turn out to be very low, but not very safe, with medium-term variability on a comparable scale to risky assets. Coincidentally, it is the safe asset returns that drive much of the medium-run variation in the realised risk premiums. Finally, because of the high returns to holding equities and real estate, the rate of return on total household wealth – Piketty (2014)'s r – is much higher than (roughly double the size of) the growth rate of the economy, g. The high $r - g$ gap points to potentially high equilibrium levels of wealth and wealth inequality in advanced economies.

Chapter 2 – "The Time Varying Risk Puzzle" – asks, given these high returns, what is it that ultimately drives them? More precisely, we have known since the work of Shiller (1981) that returns on equities tend to be excessively volatile – i.e. that stock prices fluctuate much more than the underlying fundamentals or dividends. But how widespread this "volatility puzzle" is, and which forces lie behind it, is subject to much disagreement: a field overview paper by Cochrane (2017) mentions some 10 groups of candidate theories. In this Chapter, I build on the database for housing and equities established in Chapter 1, and add to it data on long-run returns for the third major risky asset class – corporate bonds. I then use these data to study the existence and underlying drivers of excess volatility in these broad multi-asset cross-country long-run series. It turns out that excess volatility is alive and well in these new data: returns on all three major risky asset classes are predictable, with this return predictability extending over the variety of time periods and across individual countries. The “excess volatility” puzzle of Shiller, therefore, applies not only to US equities, but also to international equity, housing and corporate bond markets throughout the last century and a half.

Chapter 2 also documents one crucial irregularity in these cross-asset volatility patterns. Even though there is excess volatility within each asset class – meaning that the discount rate on each asset varies over time – the discount rates across asset classes do not co-move. In return predictability regressions, equity valuations predict future equity returns, but not future housing or corporate bond returns, and so on. The absence of observed discount rate co-movement is puzzling since most existing explanations for excess financial volatility rely on time variation in a single cross-asset factor – the discount rate – which is common to all major risky asset classes. The lack of co-movement and cross-asset predictability implies that variation in discount rates – through factors such as risk aversion, disaster risk and intermediary risk appetite – is, ultimately, not the key driver of observed asset price movements. This “time varying risk puzzle” is difficult to square with most existing asset pricing theories. I explore several modifications and alternative explanations that can help reconcile theory and data. Factors such as asset-specific risk, investor heterogeneity and market segmentation make only a modest contribution to the lack of discount rate co-movement, and are unlikely to explain the observed patterns in the data. Instead, the data point to volatile expectations as the central source of asset price volatility, in line with behavioural models. The observed expectation volatility, in turn, has real effects at business cycle frequency. Elevated sentiment – or overoptimistic expectations – predict low future GDP growth, and sentiment reversals often mark the onset of financial crises.

Chapter 3 – “The Big Bang: Stock Market Capitalization in the Long Run”, written jointly with Kaspar Zimmermann – shows that the effects of excess volatility and time varying risk premiums reach out beyond the realm of asset prices and returns. In this Chapter, we study the evolution of stock market capitalization – the total valuation of all listed equity in each of the 17 advanced economies – over the last century and a half. We combine newly collected stock capitalization data with statistics on asset returns from Chapters 1 and 2 to study not only the long-run trends, but also the underlying determinants of stock market size. It turns out that market capitalization – the usual metric for the country’s financial development – is actually primarily determined by changes in prices, rather than quantities of listed equity. We use the method of Piketty and Zucman (2014) to decompose the changes in listed equity wealth into quantity and price effects, and show that net equity issuance stays roughly constant through time at close to 1% GDP, or 4% of the previous year’s market capitalization, whereas real equity prices show large swings not only at short, but also long frequencies.

The most curious of these long-run equity price swings has taken place over the last 30 years. Between 1870 and 1985, the stock market cap to GDP ratio in our sample was roughly flat at one-third of GDP. In the 1980s and 1990s, the stock market underwent a sustained “big bang” expansion during which the market cap to GDP ratio tripled to 100%, and remained high thereafter. This structural increase in market cap has taken place in every country in our sample, and is driven by a sharp and persistent rise in equity prices. A large part of this price increase is, in turn, attributable to a structurally lower equity risk premium. We also show that the changes in the stock market cap to GDP ratio are themselves a reliable proxy for the time variation in the equity premium: high capitalization predicts future low equity returns, outperforming standard metrics such as the dividend-price ratio, and sharp run-ups in stock market cap share many characteristics with stock market bubbles. This suggests that rather than being associated with political norms and financial efficiency, stock market capitalization serves as a “Buffet indicator” of investor risk appetite.

Chapter 4 – “Sovereigns Going Bust: Estimating the Cost of Default”, joint work with Kaspar Zimmermann forthcoming in the *European Economic Review* – studies the macroeconomic implications of a specific type of financial risk: that of a sovereign bankruptcy. Theoretical models tell us that in order for sovereign debt to exist at all, defaulting on such debt obligations should incur a penalty. Since legal penalties on sovereign debt contracts are generally not enforceable, this penalty ought to take the form of macroeconomic risk, for example through a reduction in GDP. But there is very little agreement about the empirical magnitude and underlying drivers of such macroeconomic penalties. Our paper seeks to fill this gap by estimating the GDP cost of sovereign default using a novel empirical method – regression-adjusted inverse propensity score weighting (IPSWRA) of Jordà and Taylor (2016) – applied to an up-to-date comprehensive dataset of sovereign defaults and variables that influence the decision to default across 112 countries over years 1970–2010. Introducing the IPSWRA methodology to the sovereign default literature allows us to obtain a best-practice estimate of the cost, as well as a deeper understanding of why these costs come about in the first place.

We find that sovereign default is costly: it reduces GDP by roughly 3–4% in the short to medium run, with the cost lasting for the best part of 10 years. These GDP costs can, in turn, be traced back to disruptions in banking sector activity and international trade. We show that the default is accompanied by sharp declines in investment and credit, and a sizeable external adjustment through a substantial decline in imports. The default cost also becomes much higher – peaking at 9.5% of GDP – if the sovereign default is followed by a systemic banking crisis, but is ameliorated for economies under floating exchange rate regimes. Our findings suggest that financial autarky, trade frictions and sovereign-banking spillovers play a key role in generating the cost of default. These empirical results lend support to recent advances in incorporating these channels into theoretical models of sovereign debt, such as the work of Mendoza and Yue (2012), Gennaioli, Martin, and Rossi (2014) and Na, Schmitt-Grohé, Uribe, and Yue (2018)

The final Chapter 5 – “Deleveraging, Deflation and Depreciation in the Euro Area”, joint work with Gernot Müller and Martin Wolf published in the *European Economic Review* – conducts a further investigation of the macroeconomic effects of time varying financial risk. Many of the manifestations of financial risk discussed in Chapters 1–3 bring about a need to deleverage, for example as firm networth falls (Kiyotaki and Moore, 1997), or banks’ borrowing constraints tighten (Gertler and Kiyotaki, 2010). In this paper, we study the real effects of such deleveraging shocks in a currency union setting. To do this, we build on the Eggertsson and Krugman (2012) closed economy set-up and construct a New Keynesian model of a currency union with internal and external debt, using it to study the impact of deleveraging in one part of the union on relative prices and output of different union members. We show that deleveraging in one part of the union brings about deflationary spillovers which put downward pressure on prices in the other member countries, and hinder the relative price adjustment in the aftermath of this shock. This leads to a situation where an asymmetric deleveraging shock affecting only some union members brings about a potentially severe recession in one part of the union, accompanied by low inflation and stable real exchanges rates across the entire union. These deflationary spillovers come about through trade among individual member countries, and only arise when the central bank is unable to cushion the aggregate effects of the shock at the union level due to a zero lower bound constraint. Under these conditions, greater wage flexibility would make the deleveraging crisis worse, rather than better – an instance of Eggertsson and Krugman (2012)’s “paradox of flexibility” in a multi-country setting.

Our model is able to replicate the key stylised features of the recent euro-area crisis. In the data, deleveraging has been concentrated among a subset of the eurozone members: Greece, Italy, Ireland, Portugal and Spain. These countries have also undergone a substantial and protracted output decline. Nevertheless, relative price adjustment between these “stressed” economies and other euro area members has been slow, and union-wide inflation subdued – something that our model would attribute to the presence of the above-mentioned deflationary spillovers. Chapter 5 shows that financial risks – such as those discussed in Chapters 1–3 – can have pronounced macroeconomic effects, and helps us understand the exact dynamics of how these risks have played out in the recent euro area crisis.

References

Barro, Robert J. 2006. “Rare Disasters and Asset Markets in the Twentieth Century.” *Quarterly Journal of Economics* 121 (3): 823–866.

Campbell, John Y. 1999. “Asset Prices, Consumption, and the Business Cycle.” In. Vol. 1, Handbook of Macroeconomics. Elsevier. Chapter 19, 1231–1303.

Cochrane, John H. 2017. “Macro-Finance.” *Review of Finance* 21 (3): 945–985.

Eggertsson, Gauti B., and Paul Krugman. 2012. “Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach.” *Quarterly Journal of Economics* 127 (3): 1469–1513.

Fama, Eugene F. 2014. “Two Pillars of Asset Pricing.” *American Economic Review* 104 (6): 1467–1485.

Gennaioli, Nicola, Alberto Martin, and Stefano Rossi. 2014. “Sovereign Default, Domestic Banks, and Financial Institutions.” *Journal of Finance* 69 (2): 819–866.

Gertler, Mark, and Nobuhiro Kiyotaki. 2010. “Chapter 11 - Financial Intermediation and Credit Policy in Business Cycle Analysis.” In. Edited by Benjamin M. Friedman and Michael Woodford. Vol. 3, Handbook of Monetary Economics. Elsevier, 547–599.

Jordà, Òscar, and Alan M. Taylor. 2016. “The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy.” *Economic Journal* 126 (590): 219–255.

Kiyotaki, Nobuhiro, and John Moore. 1997. “Credit Cycles.” *Journal of Political Economy* 105 (2): 211–248.

Mehra, Rajnish, and Edward C. Prescott. 1985. “The Equity Premium: A Puzzle.” *Journal of Monetary Economics* 15 (2): 145–161.

Mendoza, Enrique G, and Vivian Z Yue. 2012. “A General Equilibrium Model of Sovereign Default and Business Cycles.” *Quarterly Journal of Economics* 127 (2): 889–946.

Na, Seunghoon, Stephanie Schmitt-Grohé, Martín Uribe, and Vivian Yue. 2018. “The Twin Ds: Optimal Default and Devaluation.” *American Economic Review* 108 (7): 1773–1819.

Nakamura, Emi, Jón Steinsson, Robert Barro, and José Ursúa. 2013. “Crises and Recoveries in an Empirical Model of Consumption Disasters.” *American Economic Journal: Macroeconomics* 5 (3): 35–74.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Harvard University Press.

Piketty, Thomas, and Gabriel Zucman. 2014. “Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010.” *Quarterly Journal of Economics* 129 (3): 1255–1310.

Shiller, Robert J. 1981. “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71 (3): 421–436.

Shiller, Robert J. 2014. “Speculative Asset Prices.” *American Economic Review* 104 (6): 1486–1517.

Chapter 1

The Rate of Return on Everything, 1870–2015*

Joint with Òscar Jordà, Katharina Knoll, Moritz Schularick and Alan Taylor

1.1 Introduction

What is the rate of return in an economy? It is a simple question, but hard to answer. The rate of return plays a central role in current debates on inequality, secular stagnation, risk premiums, and the decline in the natural rate of interest, to name a few. A main contribution of our paper is to introduce a large new dataset on the total rates of return for all major asset classes, including housing—the largest, but oft ignored component of household wealth. Our data cover most advanced economies—sixteen in all—starting in the year 1870.

* This work is part of a larger project kindly supported by research grants from the Bundesministerium für Bildung und Forschung (BMBF) and the Institute for New Economic Thinking (INET). We are indebted to a large number of researchers who helped with data on individual countries. We are especially grateful to Francisco Amaral for outstanding research assistance, and would also like to thank Felix Rhiel, Mario Richarz, Thomas Schwarz, Lucie Stoppok, and Yevhenii Usenko for research assistance on large parts of the project. For their helpful comments we thank the editors and referees, along with Roger Farmer, John Fernald, Philipp Hofflin, David Le Bris, Clara Martínez-Toledano, Emi Nakamura, Thomas Piketty, Matthew Rognlie, Jón Steinsson, Johannes Stroebel, and Stijn Van Nieuwerburgh. We likewise thank conference participants at the NBER Summer Institute EFG Program Meeting, the Brevan Howard Centre for Financial Analysis at Imperial College Business School, the CEPR Housing Conference, and CEPR ESSIM at the Norges Bank, as well as seminar participants at Banca d'Italia, the Bank of England, Reserve Bank of New Zealand, Cornell University, New York University, University of Illinois at Urbana-Champaign, University of Chicago Booth School of Business, UC Berkeley, UCLA Anderson, Research Center SAFE, SciencesPo, and the Paris School of Economics. All errors are our own. The views expressed herein are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of San Francisco, the Board of Governors of the Federal Reserve System, or the Deutsche Bundesbank.

Although housing wealth is on average roughly one half of national wealth in a typical economy (Piketty, 2014), data on total housing returns (price appreciation plus rents) has been lacking (Shiller, 2000, provides some historical data on house prices, but not on rents). In this paper we build on more comprehensive work on house prices (Knoll, Schularick, and Steger, 2017) and newly constructed data on rents (Knoll, 2017) to enable us to track the total returns of the largest component of the national capital stock.

We further construct total returns broken down into investment income (yield) and capital gains (price changes) for four major asset classes, two of them typically seen as relatively risky—equities and housing—and two of them typically seen as relatively safe—government bonds and short-term bills. Importantly, we compute actual asset returns taken from market data and therefore construct more detailed series than returns inferred from wealth estimates in discrete benchmark years for a few countries as in Piketty (2014).

We also follow earlier work in documenting annual equity, bond, and bill returns, but here again we have taken the project further. We re-compute all these measures from original sources, improve the links across some important historical market discontinuities (e.g., market closures and other gaps associated with wars and political instability), and in a number of cases we access new and previously unused raw data sources. In all cases, we have also brought in auxiliary sources to validate our data externally, and 100+ pages of online material documents our sources and methods. Our work thus provides researchers with the first broad non-commercial database of historical equity, bond, and bill returns—and the only database of housing returns—with the most extensive coverage across both countries and years.[1]

Our paper aims to bridge the gap between two related strands of the academic literature. The first strand is rooted in finance and is concerned with long-run returns on different assets. Dimson, Marsh, and Staunton (2009) probably marked the first comprehensive attempt to document and analyze long-run returns on investment for a broad cross-section of countries. Meanwhile, Homer and Sylla (2005) pioneered a multi-decade project to document the history of interest rates.

The second related strand of literature is the analysis of comparative national balance sheets over time, as in Goldsmith (1985a). More recently, Piketty and Zucman (2014) have brought together data from national accounts and other sources tracking the development of national wealth over long time periods. They also calculate rates of return on capital by dividing aggregate capital income in the national accounts by the aggregate value of capital, also from national accounts.

Our work is both complementary and supplementary to theirs. It is complementary as the asset price perspective and the national accounts approach are ultimately

1. For example, our work complements and extends the database on equity returns by Dimson, Marsh, and Staunton (2009). Their dataset is commercially available, but has a shorter coverage and does not break down the yield and capital gain components.

tied together by accounting rules and identities. Using market valuations, we are able to corroborate and improve the estimates of returns on capital that matter for wealth inequality dynamics. Our long-run return data are also supplementary to the work of Piketty and Zucman (2014) in the sense that we greatly extend the number of countries for which we can calculate real rates of return back to the late nineteenth century.

The new evidence we gathered can shed light on active research debates, reaching from asset pricing to inequality. For example, in one contentious area of research, the accumulation of capital, the expansion of capital's share in income, and the growth rate of the economy relative to the rate of return on capital all feature centrally in the current debate sparked by Piketty (2014) on the evolution of wealth, income, and inequality. What do the long-run patterns on the rates of return on different asset classes have to say about these possible drivers of inequality?

In many financial theories, preferences over current versus future consumption, attitudes toward risk, and covariation with consumption risk all show up in the premiums that the rates of return on risky assets carry over safe assets. Returns on different asset classes and their correlations with consumption sit at the core of the canonical consumption-Euler equation that underpins textbook asset pricing theory (see, e.g., Mehra and Prescott, 1985). But tensions remain between theory and data, prompting further explorations of new asset pricing paradigms including behavioral finance. Our new data add another risky asset class to the mix, housing, and with it, new challenges.

In another strand of research triggered by the financial crisis, Summers (2014) seeks to revive the secular stagnation hypothesis first advanced in Alvin Hansen's (1939) AEA Presidential Address. Demographic trends are pushing the world's economies into uncharted territory as the relative weight of borrowers and savers is changing and with it the possibility increases that the interest rate will fall by an insufficient amount to balance saving and investment at full employment. What is the evidence that this is the case?

Lastly, in a related problem within the sphere of monetary economics, Holston, Laubach, and Williams (2017) show that estimates of the natural rate of interest in several advanced economies have gradually declined over the past four decades and are now near zero. What historical precedents are there for such low real rates that could inform today's policymakers, investors, and researchers?

The common thread running through each of these broad research topics is the notion that the rate of return is central to understanding long-, medium-, and short-run economic fluctuations. But which rate of return? And how do we measure it? For a given scarcity of funding supply, the risky rate is a measure of the profitability of private investment; in contrast, the safe rate plays an important role in benchmarking compensation for risk, and is often tied to discussions of monetary policy settings and the notion of the natural rate. Below, we summarize our main findings.

Main findings. We present four main findings:

1. **On risky returns, r^{risky}**

   In terms of total returns, residential real estate and equities have shown very similar and high real total gains, on average about 7% per year. Housing outperformed equities before WW2. Since WW2, equities have outperformed housing on average, but had much higher volatility and higher synchronicity with the business cycle. The observation that housing returns are similar to equity returns, but much less volatile, is puzzling. Like Shiller (2000), we find that long-run capital gains on housing are relatively low, around 1% p.a. in real terms, and considerably lower than capital gains in the stock market. However, the rental yield component is typically considerably higher and more stable than the dividend yield of equities so that total returns are of comparable magnitude.

   Before WW2, the real returns on housing and equities (and safe assets) followed remarkably similar trajectories. After WW2 this was no longer the case, and across countries equities then experienced more frequent and correlated booms and busts. The low covariance of equity and housing returns reveals that there could be significant aggregate diversification gains (i.e., for a representative agent) from holding the two asset classes.

2. **On safe returns, r^{safe}**

   We find that the real safe asset return (bonds and bills) has been very volatile over the long-run, more so than one might expect, and oftentimes even more volatile than real risky returns. Each of the world wars was (unsurprisingly) a moment of very low safe rates, well below zero. So was the 1970s stagflation. The peaks in the real safe rate took place at the start of our sample, in the interwar period, and during the mid-1980s fight against inflation. In fact, the long decline observed in the past few decades is reminiscent of the secular decline that took place from 1870 to WW1. Viewed from a long-run perspective, the past decline and current low level of the real safe rate today is not unusual. The puzzle may well be why was the safe rate so high in the mid-1980s rather than why has it declined ever since.

   Safe returns have been low on average in the full sample, falling in the 1%–3% range for most countries and peacetime periods. While this combination of low returns and high volatility has offered a relatively poor risk-return tradeoff to investors, the low returns have also eased the pressure on government finances, in particular allowing for a rapid debt reduction in the aftermath of WW2.

3. **On the risk premium, $r^{risky} - r^{safe}$**

   Over the very long run, the risk premium has been volatile. Our data uncover substantial swings in the risk premium at lower frequencies that sometimes endured for decades, and which far exceed the amplitudes of business-cycle swings.

In most peacetime eras, this premium has been stable at about 4%–5%. But risk premiums stayed curiously and persistently high from the 1950s to the 1970s, long after the conclusion of WW2. However, there is no visible long-run trend, and mean reversion appears strong. Interestingly, the bursts of the risk premium in the wartime and interwar years were mostly a phenomenon of collapsing safe returns rather than dramatic spikes in risky returns.

In fact, the risky return has often been smoother and more stable than the safe return, averaging about 6%–8% across all eras. Recently, with safe returns low and falling, the risk premium has widened due to a parallel but smaller decline in risky returns. But these shifts keep the two rates of return close to their normal historical range. Whether due to shifts in risk aversion or other phenomena, the fact that safe returns seem to absorb almost all of these adjustments seems like a puzzle in need of further exploration and explanation.

4. **On returns minus growth, $r^{wealth} - g$**

Piketty (2014) argued that, if investors' return to wealth exceeded the rate of economic growth, rentiers would accumulate wealth at a faster rate and thus worsen wealth inequality. Using a measure of portfolio returns to compute "*r* minus *g*" in Piketty's notation, we uncover an important finding. Even calculated from more granular asset price returns data, the same fact reported in Piketty (2014) holds true for more countries, more years, and more dramatically: namely "$r \gg g$."

In fact, the only exceptions to that rule happen in the years in or around wartime. In peacetime, *r* has always been much greater than *g*. In the pre-WW2 period, this gap was on average 5% (excluding WW1). As of today, this gap is still quite large, about 3%–4%, though it narrowed to 2% in the 1970s before widening in the years leading up to the Global Financial Crisis.

One puzzle that emerges from our analysis is that while "*r* minus *g*" fluctuates over time, it does not seem to do so systematically with the growth rate of the economy. This feature of the data poses a conundrum for the battling views of factor income, distribution, and substitution in the ongoing debate (Rognlie, 2015). The fact that returns to wealth have remained fairly high and stable while aggregate wealth increased rapidly since the 1970s, suggests that capital accumulation may have contributed to the decline in the labor share of income over the recent decades (Karabarbounis and Neiman, 2014). In thinking about inequality and several other characteristics of modern economies, the new data on the return to capital that we present here should spur further research.

1.2 A new historical global returns database

In this section, we will discuss the main sources and definitions for the calculation of long-run returns. A major innovation is the inclusion of housing. Residential real

Table 1.1. Data coverage

Country	Bills	Bonds	Equity	Housing
Australia	1870–2015	1900–2015	1870–2015	1901–2015
Belgium	1870–2015	1870–2015	1870–2015	1890–2015
Denmark	1875–2015	1870–2015	1873–2015	1876–2015
Finland	1870–2015	1870–2015	1896–2015	1920–2015
France	1870–2015	1870–2015	1870–2015	1871–2015
Germany	1870–2015	1870–2015	1870–2015	1871–2015
Italy	1870–2015	1870–2015	1870–2015	1928–2015
Japan	1876–2015	1881–2015	1886–2015	1931–2015
Netherlands	1870–2015	1870–2015	1900–2015	1871–2015
Norway	1870–2015	1870–2015	1881–2015	1871–2015
Portugal	1880–2015	1871–2015	1871–2015	1948–2015
Spain	1870–2015	1900–2015	1900–2015	1901–2015
Sweden	1870–2015	1871–2015	1871–2015	1883–2015
Switzerland	1870–2015	1900–2015	1900–2015	1902–2015
UK	1870–2015	1870–2015	1871–2015	1896–2015
USA	1870–2015	1871–2015	1872–2015	1891–2015

estate is the main asset in most household portfolios, as we shall see, but so far very little has been known about long-run returns on housing. Our data on housing returns will cover capital gains, and imputed rents to owners and renters, the sum of the two being total returns.[2] Equity return data for publicly-traded equities will then be used, as is standard, as a proxy for aggregate business equity returns.[3]

The data include nominal and real returns on bills, bonds, equities, and residential real estate for Australia, Belgium, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. The sample spans 1870 to 2015. Table 1.1 summarizes the data coverage by country and asset class.

Like most of the literature, we examine returns to national aggregate holdings of each asset class. Theoretically, these are the returns that would accrue for the hypothetical representative-agent investor holding each country's portfolio. An advantage of this approach is that it captures indirect holdings much better, although

2. Since the majority of housing is owner-occupied, and housing wealth is the largest asset class in the economy, owner-occupier returns and imputed rents also form the lion's share of the total return on housing, as well as the return on aggregate wealth.

3. Moskowitz and Vissing-Jørgensen (2002) compare the returns on listed and unlisted U.S. equities over the period 1953–1999 and find that in aggregate, the returns on these two asset classes are similar and highly correlated, but private equity returns exhibit somewhat lower volatility. Moskowitz and Vissing-Jørgensen (2002) argue, however, that the risk-return tradeoff is worse for private compared to public equities, because aggregate data understate the true underlying volatility of private equity, and because private equity portfolios are typically much less diversified.

it leads to some double-counting thereby boosting the share of financial assets over housing somewhat. The differences are described in Appendix 1.A.15.[4]

1.2.1 The composition of wealth

Figure 1.1 shows the decomposition of economy-wide investible assets and capital stocks, based on data for five major economies at the end of 2015: France, Germany, Japan, UK and US.[5] Investible assets shown in the left panel of Figure 1.1 (and in Table 1.A.23) exclude assets that relate to intra-financial holdings and cannot be held directly by investors, such as loans, derivatives (apart from employee stock options), financial institutions' deposits, insurance and pension claims. Other financial assets mainly consist of corporate bonds and asset-backed securities. Other non-financial assets are other buildings, machinery & equipment, agricultural land, and intangible capital. The capital stock is business capital plus housing. Other capital is mostly made up of intangible capital and agricultural land. Data are sourced from national accounts and national wealth estimates published by the countries' central banks and statistical offices.[6]

Housing, equity, bonds, and bills comprise over half of all investible assets in the advanced economies today, and nearly two-thirds if deposits are included. The right-hand side panel of Figure 1.1 shows the decomposition of the capital stock into housing and various other non-financial assets. Housing is about one half of the outstanding stock of capital. In fact, housing and equities alone represent over half of total assets in household balance sheets (see Figures 1.A.5 and 1.A.6).

The main asset categories *outside* the direct coverage of this study are: commercial real estate, business assets, and agricultural land; corporate bonds; pension and insurance claims; and deposits. But most of these assets represent claims of, or are closely related to, assets that we do cover. For example, pension claims tend to be invested in stocks and bonds; listed equity is a levered claim on business assets of firms; land and commercial property prices tend to co-move with residential property prices; and deposit rates are either included in, or very similar to, our bill rate measure.[7]

Our data also exclude foreign assets. Even though the data on foreign asset holdings are relatively sparse, the evidence that we do have—presented in Appendix 1.A.15.4—suggests that foreign assets have, through history, only accounted

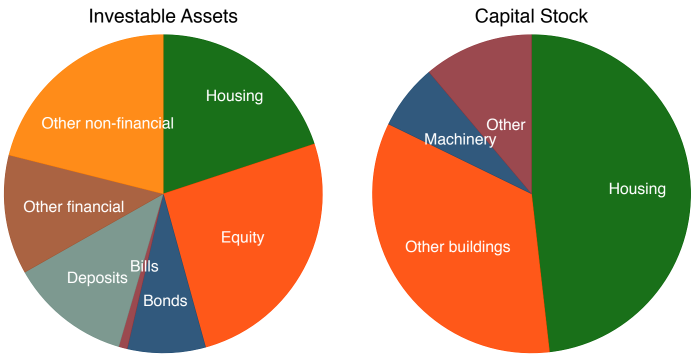
4. Within country heterogeneity is undoubtedly important, but clearly beyond the scope of a study covering nearly 150 years of data and 16 advanced economies.

5. Individual country data are shown Appendix Tables 1.A.23 and 1.A.24.

6. Both decompositions also exclude human capital, which cannot be bought or sold. Lustig, Van Nieuwerburgh, and Verdelhan (2013) show that for a broader measure of aggregate wealth that includes human capital, the size of human wealth is larger than of non-human wealth, and its return dynamics are similar to those of a long-term bond.

7. Moreover, returns on commercial real estate are captured by the levered equity returns of the firms that own this real estate, and hence are indirectly proxied by our equity return data.

Figure 1.1. Composition of investible assets and capital stock in the major economies



Note: Composition of total investible assets and capital stock. Average of the individual asset shares of France, Germany, Japan, UK, and US, as of end-2015. Investible assets are defined as the gross total of economy-wide assets excluding loans, derivatives, financial institutions' deposits, insurance, and pension claims. Other financial assets mainly consist of corporate bonds and asset-backed securities. Other non-financial assets are other buildings, machinery & equipment, agricultural land, and intangible capital. The capital stock is business capital plus housing. Other capital is mostly made up by intangible capital and agricultural land. Data are sourced from national accounts and national wealth estimates published by the countries' central banks and statistical offices.

for a small share of aggregate wealth, and the return differentials between domestic and foreign asset holdings are, with few exceptions, not that large. Taken together, this means that our dataset almost fully captures the various components of the return on overall household wealth.

1.2.2 Historical returns data

Bill returns. The canonical risk-free rate is taken to be the yield on Treasury bills, i.e., short-term, fixed-income government securities. The yield data come from the latest vintage of the long-run macrohistory database (Jordà, Schularick, and Taylor, 2017).[8] Whenever data on Treasury bill returns were unavailable, we relied on either money market rates or deposit rates of banks from Zimmermann (2017). Since short-term government debt was rarely used and issued in the earlier historical period, much of our bill rate data before the 1960s actually consist of deposit rates.[9]

Bond returns. These are conventionally the total returns on long-term government bonds. Unlike earlier cross-country studies, we focus on the bonds listed and traded

8. www.macrohistory.net/data

9. In general, it is difficult to compute the total returns on deposits because of uncertainty about losses during banking crises, and we stick to the more easily measurable government bill and bond returns where these data are available. Comparisons with the deposit rate data in Zimmermann (2017), however, indicate that the interest rate differential between deposits and our bill series is very small, with deposit rates on average roughly 0.7 percentage points below bills—a return close to zero in real terms. The returns on government bills and deposits are also highly correlated over time.

on local exchanges and denominated in local currency. This focus makes bond returns more comparable with the returns of bills, equities, and housing. Moreover, this results in a larger sample of bonds, and on bonds that are more likely to be held by the representative household in the respective country. For some countries and periods we have made use of listings on major global exchanges to fill gaps where domestic markets were thin, or local exchange data were not available (for example, Australian bonds listed in New York or London). Throughout the sample we target a maturity of around 10 years. For the second half of the 20th century, the maturity of government bonds is generally accurately defined. For the pre-WW2 period we sometimes had to rely on data for perpetuals, i.e., very long-term government securities (such as the British consol). Although as a convention we refer here to government bills and bonds as "safe" assets, both are naturally exposed to inflation and default risk, for example. In fact, real returns on these assets fluctuate substantially over time as we shall see (specifically, Sections 1.5 and 1.6).

Equity returns. These returns come from a broad range of sources, including articles in economic and financial history journals, yearbooks of statistical offices and central banks, stock exchange listings, newspapers, and company reports. Throughout most of the sample, we aim to rely on indices weighted by market capitalization of individual stocks, and a stock selection that is representative of the entire stock market. For some historical time periods in individual countries, however, we also make use of indices weighted by company book capital, stock market transactions, or weighted equally, due to limited data availability.

Housing returns. We combine the long-run house price series introduced by Knoll, Schularick, and Steger (2017) with a novel dataset on rents drawn from the unpublished PhD thesis of Knoll (2017). For most countries, the rent series rely on the rent components of the cost of living of consumer price indices constructed by national statistical offices. We then combine them with information from other sources to create long-run series reaching back to the late 19th century. To proxy the total return on the residential housing stock, our returns include both rented housing and owner-occupied properties.[10] Specifically, wherever possible we use house price and rental indices that include the prices of owner-occupied properties, and the imputed rents on these houses. Imputed rents estimate the rent that an owner-occupied house would earn on the rental market, typically by using rents of similar houses that are rented. This means that, in principle, imputed rents are similar to market rents, and are simply adjusted for the portfolio composition of owner-occupied as opposed to rented housing. Imputed rents, however, are not directly observed and hence less

10. This is in line with the treatment of housing returns in the existing literature on returns to aggregate wealth—see, for example, Piketty, Saez, and Zucman (2018) and Rognlie (2015).

precisely measured than market rents, and are typically not taxed.[11] To the best of our knowledge, we are the first to calculate total returns to housing in the literature for as long and as comprehensive a cross section of economies as we report.

Composite returns. We compute the rate of return on safe assets, risky assets, and aggregate wealth, as weighted averages of the individual asset returns. To obtain a representative return from the investor's perspective, we use the outstanding stocks of the respective asset in a given country as weights. To this end, we make use of new data on equity market capitalization (from Kuvshinov and Zimmermann, 2018) and housing wealth for each country and period in our sample, and combine them with existing estimates of public debt stocks to obtain the weights for the individual assets. A graphical representation of these asset portfolios, and further description of their construction is provided in the Appendix 1.A.15.3. Tables 1.B.1 and 1.B.2 present an overview of our four asset return series by country, their main characteristics and coverage. The paper comes with an extensive data appendix that specifies the sources we consulted and discusses the construction of the series in greater detail (see the Data Appendix, Sections 1.B.2, 1.B.3, and 1.B.4 for housing returns, and Section 1.B.5 for equity and bond returns).

1.2.3 Calculating returns

The total annual return on any financial asset can be divided into two components: the capital gain from the change in the asset price P, and a yield component Y, that reflects the cash-flow return on an investment. The total nominal return R for asset j in country i at time t is calculated as:

$$\text{Total return:} \quad R^j_{i,t} \;=\; \frac{P^j_{i,t} - P^j_{i,t-1}}{P^j_{i,t-1}} + Y^j_{i,t}\,. \tag{1.1}$$

Because of wide differences in inflation across time and countries, it is helpful to compare returns in real terms. Let $\pi_{i,t} = (CPI_{i,t} - CPI_{i,t-1})/CPI_{i,t-1}$ be the realized consumer price index (CPI) inflation rate in a given country i and year t. We calculate inflation-adjusted *real returns* r for each asset class as,

$$\text{Real return:} \quad r^j_{i,t} \;=\; (1 + R^j_{i,t})/(1 + \pi_{i,t}) - 1\,. \tag{1.2}$$

These returns will be summarized in period average form, by country, or for all countries.

Investors must be compensated for risk to invest in risky assets. A measure of this "excess return" can be calculated by comparing the real total return on the risky asset

11. We discuss the issues around imputed rents measurement, and our rental yield series more generally in Section 1.3.3.

with the return on a risk-free benchmark—in our case, the government bill rate, $r^{bill}_{i,t}$. We therefore calculate the excess return *ER* for the risky asset *j* in country *i* as

$$\text{Excess return:} \quad ER^{j}_{i,t} = r^{j}_{i,t} - r^{bill}_{i,t}. \tag{1.3}$$

In addition to individual asset returns, we also present a number of weighted "composite" returns aimed at capturing broader trends in risky and safe investments, as well as the "overall return" or "return on wealth." Appendix 1.A.15.3 provides further details on the estimates of country asset portfolios from which we derive country-year specific weights.

For safe assets, we assume that total public debt is divided equally into bonds and bills since there are no data on their market shares (only for total public debt) over our full sample. As a result, we compute the safe asset return as:

$$\text{Safe return:} \quad r^{safe}_{i,t} = \frac{r^{bill}_{i,t} + r^{bond}_{i,t}}{2}. \tag{1.4}$$

The risky asset return is calculated as a weighted average of the returns on equity and on housing. The weights *w* represent the share of asset holdings of equity and of housing stocks in the respective country *i* and year *t*, scaled to add up to 1. We use stock market capitalization and housing wealth to calculate each share and hence compute risky returns as:

$$\text{Risky return:} \quad r^{risky}_{i,t} = r^{equity}_{i,t} \times w^{equity}_{i,t} + r^{housing}_{t} \times w^{housing}_{i,t}. \tag{1.5}$$

The difference between our risky and safe return measures then provides a proxy for the aggregate risk premium in the economy:

$$\text{Risk premium:} \quad RP_{i,t} = r^{risky}_{i,t} - r^{safe}_{i,t}. \tag{1.6}$$

The "return on wealth" measure is a weighted average of returns on risky assets (equity and housing) and safe assets (bonds and bills). The weights *w* here are the asset holdings of risky and safe assets in the respective country *i* and year *t*, scaled to add to 1.[12]

$$\text{Return on wealth:} \quad r^{wealth}_{i,t} = r^{risky}_{i,t} \times w^{risky}_{i,t} + r^{safe}_{i,t} \times w^{safe}_{i,t}. \tag{1.7}$$

Finally, we also consider returns from a global investor perspective in Appendix 1.A.9. There we measure the returns from investing in local markets in U.S.

12. For comparison, Appendix 1.A.16 provides information on the equally-weighted risky return, and the equally-weighted rate of return on wealth, both calculated as simple averages of housing and equity, and housing, equity and bonds respectively.

dollars (USD). These returns effectively subtract the depreciation of the local exchange rate vis-a-vis the dollar from the nominal return:

$$\text{USD return:} \quad R_{i,t}^{j,USD} \;=\; (1 + R_{i,t}^{j})/(1 + \hat{s}_{i,t}) - 1\,, \tag{1.8}$$

where $\hat{s}_{i,t}$ is the rate of depreciation of the local currency versus the U.S. dollar in year t.

The real USD returns are then computed net of U.S. inflation $\pi_{US,t}$:

$$\text{Real USD return:} \quad r_{i,t}^{j,USD} \;=\; (1 + R_{i,t}^{j,USD})/(1 + \pi_{US,t}) - 1\,. \tag{1.9}$$

1.2.4 Constructing housing returns using the rent-price approach

This section briefly describes our methodology to calculate total housing returns. We provide further details as needed later in Section 1.3.3 and in Appendix 1.B.2. We construct estimates for total returns on housing using the rent-price approach. This approach starts from a benchmark rent-price ratio (RI_0/HPI_0) estimated in a baseline year ($t = 0$). For this ratio we rely on net rental yields from the Investment Property Database (IPD).[13] We can then construct a time series of returns by combining separate information from a country-specific house price index series (HPI_t/HPI_0) and a country-specific rent index series (RI_t/RI_0). For these indices we rely on prior work on housing prices (Knoll, Schularick, and Steger, 2017) and new data on rents (Knoll, 2017). This method assumes that the indices cover a representative portfolio of houses. Under this assumption, there is no need to correct for changes in the housing stock, and only information about the growth rates in prices and rents is necessary.

Hence, a time series of the rent-price ratio can be derived from forward and back projection as

$$\frac{RI_t}{HPI_t} \;=\; \left[\frac{\left(RI_t/RI_0\right)}{\left(HPI_t/HPI_0\right)}\right]\frac{RI_0}{HPI_0}\,. \tag{1.10}$$

In a second step, returns on housing can then be computed as:

$$R_{t+1}^{housing} \;=\; \frac{RI_{t+1}}{HPI_t} + \frac{HPI_{t+1} - HPI_t}{HPI_t}\,. \tag{1.11}$$

Our rent-price approach is sensitive to the choice of benchmark rent-price-ratios and cumulative errors from year-by-year extrapolation. We verify and adjust rent-price approach estimates using a range of alternative sources. The main source for

13. These net rental yields use rental income net of maintenance costs, ground rent, and other irrecoverable expenditure. These adjustments are discussed exhaustively in the next section. We use net rather than gross yields to improve comparability with other asset classes.

comparison is the balance sheet approach to rental yields, which calculates the rent-price ratio using national accounts data on total rental income and housing wealth. The "balance sheet" rental yield RY_t^{BS} is calculated as the ratio of total net rental income to total housing wealth:

$$RY_t^{BS} = \text{Net rental income}_t/\text{Housing Wealth}_t, \tag{1.12}$$

This balance sheet rental yield estimate can then be added to the capital gains series in order to compute the total return on housing from the balance sheet perspective. We also collect additional point-in-time estimates of net rental yields from contemporary sources such as newspaper advertisements. These measures are less sensitive to the accumulated extrapolation errors in equation (1.10), but are themselves measured relatively imprecisely.[14] Wherever the rent-price approach estimates diverge from these historical sources, we make adjustments to benchmark the rent-price ratio estimates to these alternative historical measures of the rental yield. We also construct two additional housing return series—one benchmarked to all available alternative yield estimates, and another using primarily the balance sheet approach. The results of this exercise are discussed in Section 1.3.3. Briefly, all the alternative estimates are close to one another, and the differences have little bearing on any of our results.

1.3 Rates of return: Aggregate trends

Our headline summary data appear in Table 1.2 and Figure 1.2. The top panel of Table 1.2 shows the full sample (1870–2015) results whereas the bottom panel of the table shows results for the post-1950 sample. Note that here, and throughout the paper, rates of return are always annualized. Units are always expressed in percent per year, for raw data as well as for means and standard deviations. All means are arithmetic means, except when specifically referred to as geometric means.[15] Data are pooled and equally-weighted, i.e., they are raw rather than portfolio returns. We will always include wars so that results are not polluted by bias from omitted

14. We discuss the advantages and disadvantages of these different approaches in Section 1.3.3. Broadly speaking, the balance sheet approach can be imprecise due to measurement error in total imputed rent and national housing wealth estimates. Newspaper advertisements are geographically biased and only cover gross rental yields, so that the net rental yields have to be estimated.

15. In what follows we focus on conventional average annual real returns. In addition, we often report period-average geometric mean returns corresponding to the annualized return that would be achieved through reinvestment or compounding. For any sample of years T, geometric mean returns are calculated as

$$\left(\prod_{t\in T}(1+r_{i,t}^j)\right)^{\frac{1}{T}} - 1.$$

Note that the arithmetic period-average return is always larger than the geometric period-average return, with the difference increasing with the volatility of the sequence of returns.

disasters. We do, however, exclude hyperinflation years (but only a few) in order to focus on the underlying trends in returns, and to avoid biases from serious measurement errors in hyperinflation years, arising from the impossible retrospective task of matching within-year timing of asset and CPI price level readings which can create a spurious, massive under- or over-statement of returns in these episodes.[16]

The first key finding is that residential real estate, not equity, has been the best long-run investment over the course of modern history. Although returns on housing and equities are similar, the volatility of housing returns is substantially lower, as Table 1.2 shows. Returns on the two asset classes are in the same ballpark—around 7%—but the standard deviation of housing returns is substantially smaller than that of equities (10% for housing versus 22% for equities). Predictably, with thinner tails, the compounded return (using the geometric average) is vastly better for housing than for equities—6.6% for housing versus 4.7% for equities. This finding appears to contradict one of the basic tenets of modern valuation models: higher risks should come with higher rewards.

Differences in asset returns are not driven by unusual events in the early pre-WW2 part of the sample. The bottom panel of Table 1.2 makes this point. Compared to the full sample results in the top panel, the same clear pattern emerges: stocks and real estate dominate in terms of returns. Moreover, average returns post–1950 are similar to those for the full sample even though the postwar subperiod excludes the devastating effects of the two world wars. Robustness checks are reported in Figures 1.A.1, 1.A.2, and 1.A.3. Briefly, the observed patterns are not driven by the smaller European countries in our sample. Figure 1.A.1 shows average real returns weighted by country-level real GDP, both for the full sample and post–1950 period. Compared to the unweighted averages, equity performs slightly better, but the returns on equity and housing remain very similar, and the returns and riskiness of all four asset classes are very close to the unweighted series in Table 1.2.

The results could be biased due to the country composition of the sample at different dates given data availability. Figure 1.A.2 plots the average returns for sample-consistent country groups, starting at benchmark years—the later the benchmark year, the more countries we can include. Again, the broad patterns discussed above are largely unaffected.

We also investigate whether the results are biased due to the world wars. Figure 1.A.3 plots the average returns in this case. The main result remains largely unchanged. Appendix Table 1.A.3 also considers the risky returns during wartime in more detail, to assess the evidence for rare disasters in our sample. Returns during both wars were indeed low and often negative, although returns during WW2 in a number of countries were relatively robust.

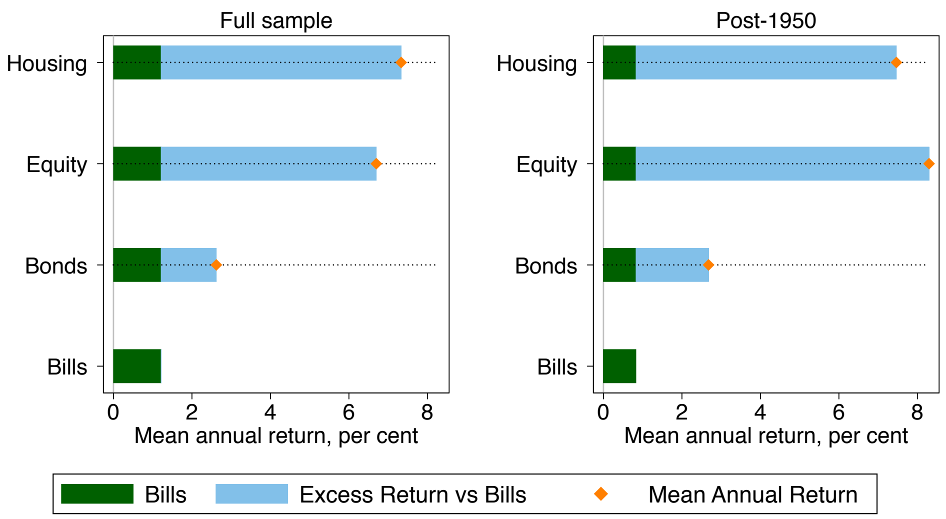
16. Appendix 1.A.7 and Table 1.A.12 do, however, provide some rough proxies for returns on different asset classes during hyperinflations.

Table 1.2. Global real returns

	Real returns				Nominal Returns			
	Bills	Bonds	Equity	Housing	Bills	Bonds	Equity	Housing
Full sample:								
Mean return p.a.	1.03	2.53	6.88	7.06	4.58	6.06	10.65	11.00
Standard deviation	6.00	10.69	21.79	9.93	3.32	8.88	22.55	10.64
Geometric mean	0.83	1.97	4.66	6.62	4.53	5.71	8.49	10.53
Mean excess return p.a.	.	1.51	5.85	6.03				
Standard deviation	.	8.36	21.27	9.80				
Geometric mean	.	1.18	3.77	5.60				
Observations	1767	1767	1767	1767	1767	1767	1767	1767
Post-1950:								
Mean return p.a.	0.88	2.79	8.30	7.42	5.39	7.30	12.97	12.27
Standard deviation	3.42	9.94	24.21	8.87	4.03	9.81	25.03	10.14
Geometric mean	0.82	2.32	5.56	7.08	5.31	6.88	10.26	11.85
Mean excess return p.a.	.	1.91	7.42	6.54				
Standard deviation	.	9.21	23.78	9.17				
Geometric mean	.	1.51	4.79	6.18				
Observations	1022	1022	1022	1022	1022	1022	1022	1022

Note: Annual global returns in 16 countries, equally weighted. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for all four asset returns. Excess returns are computed relative to bills.

Figure 1.2. Global real rates of return



Notes: Arithmetic average real returns p.a., unweighted, 16 countries. Consistent coverage within each country: each country-year observation used to compute the average has data for all four asset returns.

Finally, our aggregate return data take the perspective of a domestic investor in a representative country. Appendix Table 1.A.14 instead takes the perspective of a global USD-investor, and assesses the USD value of the corresponding returns. The magnitude and ranking of returns are similar to those reported in Table 1.2, although the volatilities are substantially higher. This is to be expected given that the underlying asset volatility is compounded by the volatility in the exchange rate. We also find somewhat higher levels of USD returns, compared to those in local currency.

What comes next in our discussion of raw rates of return? We will look more deeply at risky rates of return, and delve into their time trends and the decomposition of housing and equity returns into the capital gain and yield components in greater detail in Section 1.4. We will do the same for safe returns in Section 1.5. But first, to justify our estimates, since these are new data, we have to spend considerable time to explain our sources, methods, and calculations. We next compare our data to other literature in Section 1.3.1. We subject the equity returns and risk premium calculation to a variety of accuracy checks in Section 1.3.2. We also subject the housing returns and risk premium calculation to a variety of accuracy checks in Section 1.3.3. Section 1.3.4 then discusses the comparability of the housing and equity return series. For the purposes of our paper, these very lengthy next four subsections undertake the necessary due diligence and discuss the various quality and consistency checks we undertook to make our data a reliable source for future analysis—and only after that is done do we proceed with analysis and interpretation based on our data.

However, we caution that all these checks may be as exhausting as they are exhaustive and a time-constrained reader eager to confront our main findings may jump to the end of this section and resume where the analytical core of the paper begins at the start of Section 1.4 on page 42.

1.3.1 Comparison to existing literature

Earlier work on asset returns has mainly focused on equities and the corresponding risk premium over safe assets (bills or bonds), starting with Shiller's analysis of historical US data (Shiller, 1981), later extended to cover post-1920 Sweden and the UK (Campbell, 1999), and other advanced economies back to 1900 (Dimson, Marsh, and Staunton, 2009), or back to 1870 (Barro and Ursúa, 2008). The general consensus in this literature is that equities earn a large premium over safe assets. The cross-country estimates of this premium vary between 7% in Barro and Ursúa (2008) and 6% in Dimson, Marsh, and Staunton (2009) using arithmetic means. Campbell (1999) documents a 4.7% geometric mean return premium instead.

We find a similarly high, though smaller, equity premium using our somewhat larger and more consistent historical dataset. Our estimate of the risk premium stands at 5.9% using arithmetic means, and 3.8% using geometric means (see Table 1.2). This is lower than the estimates by Campbell (1999) and Barro and Ursúa (2008). The average risk premium is similar to that found by Dimson, Marsh, and Staunton (2009), but our returns tend to be slightly lower for the overlapping time period.[17] Details aside, our data do confirm the central finding of the literature on equity market returns: stocks earn a large premium over safe assets.

Studies on historical housing returns, starting with the seminal work of Robert Shiller (see Shiller, 2000, for a summary), have largely focused on capital gains. The rental yield component has received relatively little attention, and in many cases is missing entirely. Most of the literature pre-dating our work has therefore lacked the necessary data to calculate, infer, or discuss the total rates of return on housing over the long run. The few studies that take rents into account generally focus on the post-1970s US sample, and often on commercial real estate. Most existing evidence either places the real estate risk premium between equities and bonds, or finds that it is similar to that for equities (see Ruff, 2007; Francis and Ibbotson, 2009; Ilmanen, 2011; Favilukis, Ludvigson, and Van Nieuwerburgh, 2017). Some studies have even found that over the recent period, real estate seems to outperform equities in risk-adjusted terms (Shilling, 2003; Cheng, Lin, and Liu, 2008).

The stylized fact from the studies on long-run housing capital appreciation is that over long horizons, house prices only grow a little faster than the consumer price index. But again, note that this is *only* the capital gain component in (1.1). Low levels of real capital gains to housing was shown by Shiller (2000) for the US, and is also true, albeit to a lesser extent, in other countries, as documented in Knoll, Schularick, and Steger (2017). Our long-run average capital appreciation data for the US largely come from Shiller (2000), with two exceptions. For the 1930s, we use the more representative index of Fishback and Kollmann (2015) that documents a larger fall in prices during the Great Depression. From 1975 onwards, we use a Federal Housing Finance Agency index, which has a slightly broader coverage. Appendix 1.A.13 compares our series with Shiller's data and finds that switching to Shiller's index has no effect on our results for the US. See also the Online Appendix of Knoll, Schularick, and Steger (2017) for further discussion.

However, our paper turns this notion of low housing returns on its head—because we show that including the yield component in (1.1) in the housing return calculation generates a housing risk premium roughly as large as the equity risk premium. Prior to our work on historical rental yields this finding could not be known. Coincidentally, in our long-run data we find that most of the real equity return also comes from the dividend yield rather than from real equity capital gains which are

17. Our returns are substantially lower for France and Portugal (see Appendix Table 1.A.18). These slightly lower returns are largely a result of more extensive consistency and accuracy checks that eliminate a number of upward biases in the series, and better coverage of economic disasters. We discuss these data quality issues further in Section 1.3.2. Appendix 1.A.12 compares our equity return estimates with the existing literature on a country basis.

low, especially before the 1980s. Thus the post-1980 observation of large capital gain components in both equity and housing total returns is completely unrepresentative of the normal long-run patterns in the data, another fact underappreciated before now.

Data on historical returns for all major asset classes allow us to compute the return on aggregate wealth (see equation 1.7). In turn, this return can be decomposed into various components by asset class, and into capital gains and yields, to better understand the drivers of aggregate wealth fluctuations. This connects our study to the literature on capital income, and the stock of capital (or wealth) from a national accounts perspective. Even though national accounts and national balance sheet estimates have existed for some time (see Kuznets, 1941; Goldsmith, 1985b), it is only recently that scholars have systematized and compared these data to calculate a series of returns on national wealth.[18]

Piketty, Saez, and Zucman (2018) compute balance sheet returns on aggregate wealth and for individual asset classes using post-1913 US data. Balance sheet return data outside the US are sparse, although Piketty and Zucman (2014) provide historical estimates at benchmark years for three more countries, and, after 1970, continuous data for an additional five countries. Appendix 1.A.18 compares our market-based return estimates for the US with those of Piketty, Saez, and Zucman (2018). Housing returns are very similar. However, equity returns are several percentage points above our estimates, and those in the market-based returns literature more generally. Part of this difference reflects the fact that balance sheet returns are computed to measure income before corporate taxes, whereas our returns take the household perspective and are therefore net of corporate tax. Another explanation for the difference is likely to come from some measurement error in the national accounts data.[19] When it comes to housing, our rental yield estimates are broadly comparable and similar to those derived using the balance sheet approach, for a broad selection of countries and historical time spans.[20]

Our dataset complements the market-based return literature by increasing the coverage in terms of assets, return components, countries, and years; improving data consistency and documentation; and making the dataset freely available for

18. The return on an asset from a national accounts perspective, or the "balance sheet approach" to returns, r_t^{BS} is the sum of the yield, which is capital income (such as rents or profits) in relation to wealth, and capital gain, which is the change in wealth not attributable to investment. See Appendix 1.A.18 and equation (1.A.1) for further details.

19. See Appendix 1.A.18 for more detail. In brief, the main conceptual difference between the two sets of estimates, once our returns are grossed up for corporate tax, is the inclusion of returns on unlisted equities in the national accounts data. But existing evidence suggests that these return differentials are not large (Moskowitz and Vissing-Jørgensen, 2002), and the size of the unlisted sector not sufficiently large to place a large weight of this explanation, which leads us to conjecture that there is some measurement error in the national income and wealth estimates that is driving the remaining difference.

20. See Section 1.3.3 and Appendix 1.B.2 for more detail.

future research. This comprehensive coverage can also help connect the market-based return estimates to those centered around national accounts concepts. We hope that eventually, this can improve the consistency and quality of both market-based returns and national accounts data.

1.3.2 Accuracy of equity returns

The literature on equity returns has highlighted two main sources of bias in the data: weighting and sample selection. Weighting biases arise when the stock portfolio weights for the index do not correspond with those of a representative investor, or a representative agent in the economy. Selection biases arise when the selection of stocks does not correspond to the portfolio of the representative investor or agent. This second category also includes issues of survivorship bias and missing data bias arising from stock exchange closures and restrictions.

We consider how each of these biases affect our equity return estimates in this section. An accompanying Appendix Table 1.B.2 summarizes the construction of the equity index for each country and time period, with further details provided in Appendix 1.B.5.

Weighting bias. The best practice when weighting equity indices is to use market capitalization of individual stocks. This approach most closely mirrors the composition of a hypothetical representative investor's portfolio. Equally-weighted indices are likely to overweight smaller firms, which tend to carry higher returns and higher volatility.

The existing evidence from historical returns on the Brussels and Paris stock exchanges suggests that using equally-weighted indices biases returns up by around 0.5 percentage points, and their standard deviation up by 2–3 percentage points (Le Bris and Hautcoeur, 2010; Annaert, Buelens, Cuyvers, De Ceuster, Deloof, et al., 2011). The size of the bias, however, is likely to vary across across markets and time periods. For example, Grossman (2017) shows that the market-weighted portfolio of UK stocks outperformed its equally-weighted counterpart over the period 1869–1929.

To minimize this bias, we use market-capitalization-weighted indices for the vast majority of our sample (see Appendix Table 1.B.2 and Appendix 1.B.5). Where market-capitalization weighting was not available, we have generally used alternative weights such as book capital or transaction volumes, rather than equally-weighted averages. For the few equally-weighted indices that remain in our sample, the overall impact on aggregate return estimates ought to be negligible.

Selection and survivorship bias. Relying on an index whose selection does not mirror the representative investor's portfolio carries two main dangers. First, a small sample may be unrepresentative of overall stock market returns. And second, a sample that is selected ad-hoc, and especially ex-post, is likely to focus on surviving firms, or successful firms, thus overstating investment returns. This second bias extends not only to stock prices but also to dividend payments, as some historical studies only consider dividend-paying firms.[21] The magnitude of survivorship bias has generally been found to be around 0.5 to 1 percentage points (Nielsen and Risager, 2001a; Annaert, Buelens, and De Ceuster, 2012), but in some time periods and markets it could be larger (see Le Bris and Hautcoeur, 2010, for France).

As a first best, we always strive to use all-share indices that avoid survivor and selection biases. For some countries and time periods where no such indices were previously available, we have constructed new weighted all-share indices from original historical sources (e.g., early historical data for Norway and Spain). Where an all-share index was not available or newly constructed, we have generally relied on "blue-chip" stock market indices. These are based on an ex-ante value-weighted sample of the largest firms on the market. It is updated each year and tends to capture the lion's share of total market capitalization. Because the sample is selected ex-ante, it avoids ex-post selection and survivorship biases. And because historical equity markets have tended to be quite concentrated, "blue-chip" indices have been shown to be a good proxy for all-share returns (see Annaert, Buelens, Cuyvers, De Ceuster, Deloof, and De Schepper, 2011). Finally, we include non-dividend-paying firms in the dividend yield calculation.

Stock market closures and trading restrictions. A more subtle form of selection bias arises when the stock market is closed and no market price data are available. One way of dealing with closures is to simply exclude them from the baseline return comparisons. But this implicitly assumes that the data are "missing at random"—i.e., that stock market closures are unrelated to underlying equity returns. Existing research on rare disasters and equity premiums shows that this is unlikely to be true (Nakamura, Steinsson, Barro, and Ursúa, 2013). Stock markets tend to be closed precisely at times when we would expect returns to be low, such as periods of war and civil unrest. Return estimates that exclude such rare disasters from the data will thus overstate stock returns.

To guard against this bias, we include return estimates for the periods of stock market closure in our sample. Where possible, we rely on alternative data sources to fill the gap, such as listings of other exchanges and over-the-counter transactions—for example, in the case of WW1 Germany we use the over-the-counter index from Ronge (2002) and for WW2 France we use the newspaper index from Le Bris and Hautcoeur (2010). In cases where alternative data are not available, we interpolate the prices of securities listed both before and after the exchange was closed to es-

21. As highlighted by Brailsford, Handley, and Maheswaran (2012), this was the case with early Australian data, and the index we use scales down the series for dividend-paying firms to proxy the dividends paid by all firms, as suggested by these authors.

Table 1.3. Geometric annual average and cumulative total equity returns in periods of stock market closure

Episode	Real returns		Nominal returns		Real capitalization	
	Geometric average	Cumulative	Geometric average	Cumulative	Geometric average	Cumulative
Spanish Civil War, 1936–40	-4.01	-15.09	9.03	41.32	-10.22	-35.04
Portuguese Revolution, 1974–77	-54.98	-90.88	-44.23	-82.65	-75.29	-98.49
Germany WW1, 1914–18	-21.67	-62.35	3.49	14.72		
Switzerland WW1, 1914–16	-7.53	-14.50	-0.84	-1.67	-8.54	-16.34
Netherlands WW2, 1944–46	-12.77	-20.39	-5.09	-8.36		

Note: Cumulative and geometric average returns during periods of stock market closure. Estimated by interpolating returns of shares listed both before an after the exchange was closed. The change in market capitalization compares the capitalization of all firms before the market was closed, and once it was opened, and thus includes the effect of any new listings, delistings and bankruptcies that occured during the closure.

timate the return (if no dividend data are available, we also assume no dividends were paid).[22]

Even though this approach only gives us a rough proxy of returns, it is certainly better than excluding these periods, which effectively assumes that the return during stock market closures is the same as that when the stock markets are open. In the end, we only have one instance of stock market closure for which we are unable to estimate returns—that of the Tokyo stock exchange in 1946–1947. Appendix 1.A.8 further assesses the impact of return interpolation on the key moments of our data and finds that, over the full sample, it is negligible.

Table 1.3 shows the estimated stock returns during the periods of stock exchange closure in our sample. The first two columns show average and cumulative real returns, and the third and fourth columns show the nominal returns. Aside from the case of WW1 Germany, returns are calculated by comparing the prices of shares listed both before and after the market closure. Such a calculation may, however, overstate returns because it selects only those companies that "survived" the closure. As an additional check, the last two columns of Table 1.3 show the inflation-adjusted change in market capitalization of stocks before and after the exchange was closed. This serves as a lower bound for investor returns because it would be as if we assumed that all delisted stocks went bankrupt (i.e., to a zero price) during the market closure.

Indeed, the hypothetical investor returns during the periods of market closure are substantially below market averages. In line with Nakamura, Steinsson, Barro,

22. For example, the Swiss stock exchange was closed between July 1914 and July 1916. Our data for 1914 capture the December 1913–July 1914 return, for 1915 the July 1914–July 1916 return, and for 1916 the July 1916–December 1916 return. For the Spanish Civil war, we take the prices of securities in end-1936 and end-1940, and apportion the price change in-between equally to years 1937–1939.

and Ursúa (2013), we label these periods as "rare disasters." The average per-year geometric mean return ranges from a modestly negative –4% p.a. during the Spanish Civil War, to losses of roughly 55% p.a. during the three years after the Portuguese Carnation Revolution. Accounting for returns of delisted firms is likely to bring these estimates down even further, as evinced by the virtual disappearance of the Portuguese stock market in the aftermath of the revolution.

Having said this, the impact of these rare events on the average cross-country returns (shown in Table 1.2) is small, around –0.1 percentage points, precisely because protracted stock market closures are very infrequent. The impact on country-level average returns is sizeable for Portugal and Germany (around –1 percentage point), but small for the other countries (–0.1 to –0.4 percentage points). Appendix 1.A.7 provides a more detailed analysis of returns during consumption disasters. On average, equity returns during these times are low, with an average cumulative real equity return drop of 6.7% during the disaster years.

Lastly, Nakamura, Steinsson, Barro, and Ursúa (2013) also highlight a more subtle bias arising from asset price controls. This generally involves measures by the government to directly control transaction prices, as in Germany during 1943–47, or to influence the funds invested in the domestic stock market (and hence the prices) via controls on spending and investment, as in France during WW2 (Le Bris, 2012). These measures are more likely to affect the timing of returns rather than their long-run average level, and should thus have little impact on our headline estimates. For example, Germany experienced negative nominal and real returns despite the WW2 stock price controls; and even though the policies it enacted in occupied France succeeded in generating high nominal stock returns, the real return on French stocks during 1940–44 was close to zero. Both of these instances were also followed by sharp drops in stock prices when the controls were lifted.[23]

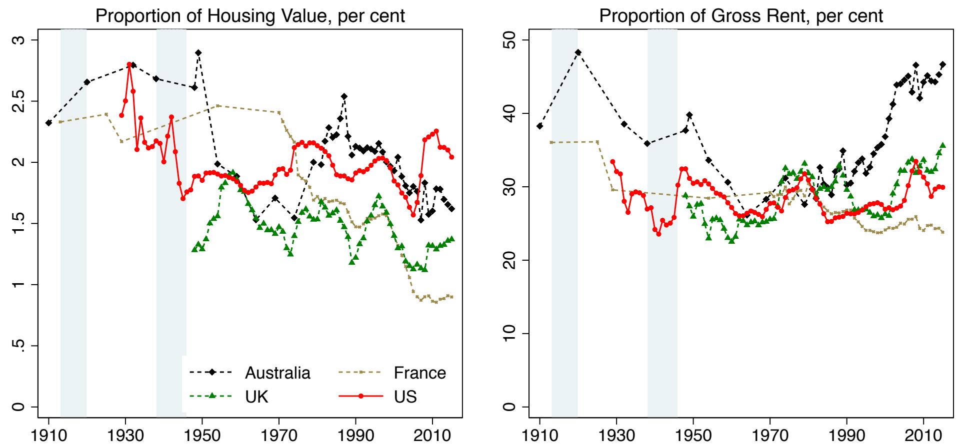
1.3.3 Accuracy of housing returns

The biases that affect equity returns—weighting and selection—can also apply to returns on housing. There are also other biases that are specific to housing return estimates. These include costs of running a housing investment, and the benchmarking of rent-price ratios to construct the historical rental yield series. We discuss each of these problematic issues in turn in this section. Our focus throughout is mainly on rental yield data, as the accuracy and robustness of the house price series has been extensively discussed in Knoll, Schularick, and Steger (2017) in their online appendix.

Maintenance costs. Any homeowner incurs costs for maintenance and repairs which lower the rental yield and thus the effective return on housing. We deal with

23. The losses in the German case are difficult to ascertain precisely because the lifting of controls was followed by a re-denomination that imposed a 90% haircut on all shares.

Figure 1.3. Costs of running a housing investment



Note: Total costs include depreciation and all other housing-related expenses excluding interest, taxes and utilities (mainly maintenance and insurance payments). Costs are estimated as the household consumption of the relevant intermediate housing input, or fixed housing capital, in proportion to total housing wealth (left panel), or total gross rent (right panel).

this issue by the choice of the benchmark rent-price ratios. Specifically, we anchor to the Investment Property Database (IPD) whose rental yields reflect net income—net of property management costs, ground rent, and other irrecoverable expenditure—as a percentage of the capital employed. The rental yields calculated using the rent-price approach detailed in Section 1.2.4 are therefore net yields. To enable a like-for-like comparison, our historical benchmark yields are calculated net of estimated running costs and depreciation. Running costs are broadly defined as housing-related expenses excluding interest, taxes, and utilities—i.e., maintenance costs, management, and insurance fees.

Applying the rent-price approach to net yield benchmarks assumes that running costs remain stable relative to gross rental income over time within each country. To check this, Figure 1.3 presents historical estimates of running costs and depreciation for Australia, France, UK, and US, calculated as the sum of the corresponding housing expenditures and fixed capital consumption items in the national accounts. The left-hand panel presents these as a proportion of total housing value, and the right-hand panel as a proportion of gross rent. Relative to housing value, costs have been stable over the last 40 years, but were somewhat higher in the early-to-mid 20th century. This is to be expected since these costs are largely related to structures, not land, and structures constituted a greater share of housing value in the early 20th century (Knoll, Schularick, and Steger, 2017). Additionally, structures themselves may have been of poorer quality in past times. When taken as a proportion of gross rent, however, as shown in the right-hand panel of Figure 1.3, housing costs have been relatively stable, or at least not higher historically than they are today. This is likely because both gross yields and costs are low today, whereas historically both yields and costs were higher, with the two effects more or less cancelling out. This suggests that the historical rental yields that we have calculated using the rent-price approach are a good proxy for net yields.

Rental yield benchmarking. To construct historical rental yield series using the rent-price approach, we start with a benchmark rent-price ratio from the Investment Property Database (IPD), and extend the series back using the historical rent and house price indices (see Section 1.2.4).[24] This naturally implies that the level of returns is sensitive to the choice of the benchmark ratio. Moreover, past errors in rent and house price indices can potentially accumulate over time and may cause one to substantially over- or understate historical rental yields and housing returns. If the historical capital gains are overstated, the historical rental yields will be overstated too.

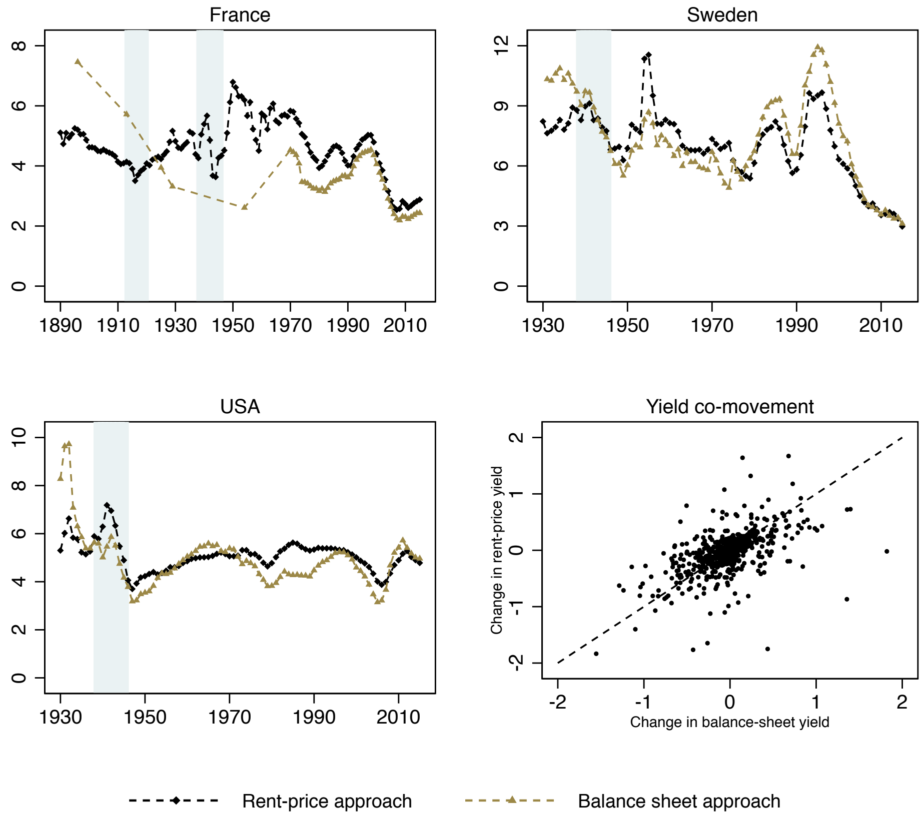
To try to avert such problems, we corroborate our rental yield estimates using a wide range of alternative historical and current-day sources. The main source of these independent comparisons comes from estimates using the balance sheet approach and national accounts data. As shown in equation 1.12, the "balance sheet" rental yield is the ratio of nationwide net rental income to total housing wealth. Net rental income is computed as gross rents paid less depreciation, maintenance and other housing-related expenses (excluding taxes and interest), with all data taken from the national accounts. The balance sheet approach gives us a rich set of alternative rental yield estimates both for the present day and even going back in time to the beginning of our sample in a number of countries. The second source for historical comparisons comes from advertisements in contemporary newspapers and various other contemporary publications. Third, we also make use of alternative current-day benchmarks based on transaction-level market rent data, and the rental expenditure and house price data from numbeo.com.[25] For all these measures, we adjust gross yields down to obtain a proxy for net rental yields.

Historical sources offer point-in-time estimates which avoid the cumulation of errors, but can nevertheless be imprecise. The balance sheet approach relies on housing wealth and rental income data, both of which are subject to potential measurement error. For housing wealth, it is inherently difficult to measure the precise value of all dwellings in the economy. Rental income is largely driven by the imputed rents of homeowners, which have to be estimated from market rents by matching the market rent to owner-occupied properties based on various property characteristics.

24. For Australia and Belgium, we instead rely on yield estimates from transaction-level data (Fox and Tulip (2014) and Numbeo.com, which are more in line with current-day and alternative historical estimates than IPD.

25. The high-quality transaction level data are available for Australia and the US, from Fox and Tulip (2014) (sourced from RP Data) and Giglio, Maggiori, and Stroebel (2015) (sourced from Trulia) respectively. We use the Fox and Tulip (2014) yield as benchmark for Australia. For the US, we use IPD because it is in line with several alternative estimates, unlike Trulia data which are much higher. See Appendix 1.B.2 for further details.

Figure 1.4. Comparison of rent-price and balance-sheet approach historical rental yields



Note: The rent-price approach uses the baseline estimates in this paper. The balance sheet approach estimates the net yield in each year as total rental expenditure less housing running costs and depreciation, in proportion to total housing wealth.

This procedure can suffer from errors both in the survey data on property characteristics and market rents, and the matching algorithm.[26] Newspaper advertisements are tied to a specific location, and often biased towards cities. And transaction-level or survey data sometimes only cover the rental sector, rather than both renters and homeowners.

Given the potential measurement error in all the series, our final rental yield series uses data from both the rent-price approach and the alternative benchmarks listed above. More precisely, we use the following method to construct our final "best-practice" rental yield series. If the rent-price approach estimates are close to alternative measures, we keep the rent-price approach data. This is the case for most historical periods in our sample. If there is a persistent level difference between the

26. For example, in the UK a change to imputation procedures in 2016 and the use of new survey data resulted in historical revisions which almost tripled imputed rents (see Office for National Statistics, 2016). We use a mixture of the old and new/revised data for our historical estimates.

Table 1.4. Impact of using different rental yield benchmarks

	Equity	Housing				
		Baseline	Low initial bench-mark	High intial bench-mark	Historical bench-marks	Balance sheet approach
Mean return p.a.	6.88	7.06	6.29	7.89	6.83	6.30
Standard deviation	21.79	9.93	9.89	10.03	9.93	9.95
Geometric mean	4.66	6.62	5.85	7.45	6.39	5.86
Observations	1767	1767	1767	1767	1767	1767

Note: Average total real returns across 16 countries, equally weighted.

rent-price approach and alternative estimates, we adjust the benchmark yield to better match the historical and current data across the range of sources. This is the case for Australia and Belgium. If the levels are close for recent data but diverge historically, we adjust the historical estimates to match the alternative benchmarks. For most countries such adjustments are small or only apply to a short time span, but for Finland and Spain they are more substantial. Appendix 1.B.2 details the alternative sources and rental yield construction, including any such adjustments, for each country.

How large is the room for error in our final housing return series? To get a sense of the differences, Figure 1.4 compares the rent-price approach of net rental yield estimates (black diamonds) with those using the balance sheet approach (brown triangles). The first three panels show the time series of the two measures for France, Sweden, and US, and the bottom-right panel shows the correlation between changes in rent-price and balance sheet yields in nine countries (Australia, Denmark, France, Germany, Italy, Japan, Sweden, UK, and US).[27] The level of the rent-price ratio using the two approaches is similar, both in the modern day and historically.[28] The two yield measures also follow a very similar time series pattern, both in the three countries depicted in panels 1–3, and the broader sample of countries summarized in the bottom-right panel.

Table 1.4 provides a more comprehensive comparison. Columns 1 and 2 present the arithmetic and geometric mean, and the standard deviation, for the baseline measures of equity and housing annual real total returns in our sample (also shown in Table 1.2). Column 3 instead uses the lowest possible initial benchmark for the housing series.[29] The resulting returns are around 0.8 percentage points (hence-

27. We limit our analysis to countries where the balance sheet approach data goes back at least several decades.

28. For France, the historical data disagree somewhat, with balance sheet approach estimates both above and below the rent-price approach for some years. We further confirm the housing return series for France using returns on housing investment trusts, documented in the subsequent sections.

29. For example, the balance sheet approach yield in 2013 Danish data is lower than the IPD yield; hence column 3 uses the 2013 balance sheet yield as the initial benchmark. For countries where

forth, pps) lower, in both arithmetic and geometric mean terms. Column 4 instead uses the highest available benchmark, thus raising housing returns by 0.8 pps. Column 5 uses historical benchmarks for all rental yield series before 1980, i.e., we use these benchmarks as the main source for the yields, and only use the rent-price approach for interpolation.[30] This makes very little difference to the returns, lowering them by around 0.2 pps. The last column 6 instead uses the balance sheet approach as the baseline estimate, both for the current and historical period. It then uses the rent-price approach to fill the gaps and interpolate between the balance sheet estimates. [31] Finally, we compute the total balance sheet return on housing as the sum of capital gains and the balance sheet yield.[32] The resulting return is 0.8 pps lower than our baseline estimates.

Taken together, this analysis suggests that the potential margins for error are small. Even under the more stringent checks, housing returns remain within a percentage point of our baseline estimates. The average return is always similar to equities in arithmetic mean terms, and always above equities when using the geometric mean.

Geographic coverage and selection biases. Our data aim to approximate the return on a representative agent's housing portfolio. Selection bias means that the selection of properties in our dataset does not mirror the balance sheet of the representative agent. The main reason for this bias is selective geographical coverage. Housing returns can vary a lot by location, and our data are based on a sample of housing transactions.

To make our samples as representative as possible, we strive to attain a broad geographic coverage for both house price and rental data. Knoll, Schularick, and Steger (2017) discuss the potential location biases in house price data, but find that the house price trends in their, and hence our, dataset should be representative of country-level trends. When it comes to rents, the benchmark IPD yields are based on portfolios of institutional investors, which are slightly biased towards cities. This would lead to lower yields than the national average. On the other hand, investors may select higher-yielding properties within any given city. Comparisons with aggregate balance sheet approach data and alternative estimates indicate that, on average, IPD yields tend to be representative at country level. Further, IPD yields are

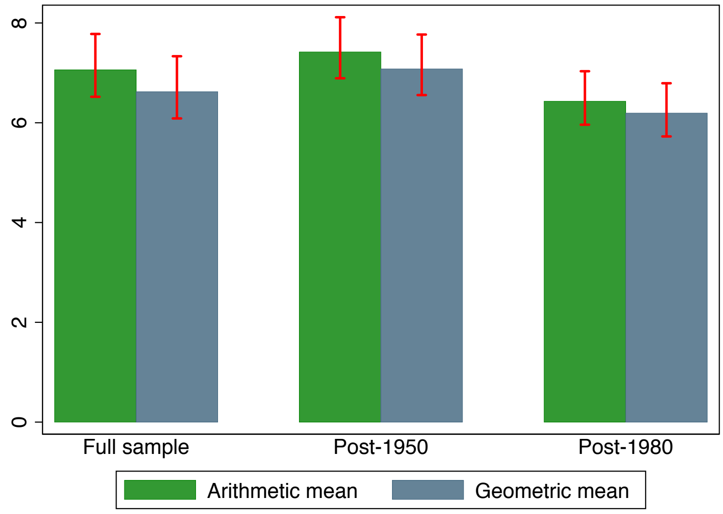
we benchmark to historical rental yields, we use the same historical benchmark for all three series. For example, for Australia, we use a historical benchmark yield in 1949. So the "high" housing return series uses the high rental yield benchmark for 1950–2015, and the historical benchmark for 1900–1949.

30. For example, the series for Denmark is benchmarked to the lower balance sheet approach yield estimates for 1890–1910 and 1950–1970, and newspaper estimates for 1920–1940 (also see Appendix Figure 1.B.3).

31. Newspaper yield estimates are used as additional benchmarks for interpolation

32. This means that we use market-based house price data for capital gains, which is also common practice in the balance sheet approach computation, due to the large potential for error when estimating housing capital gains as a residual between wealth changes and investment. Piketty, Saez, and Zucman (2018) use Shiller (2000) house price data for the balance sheet return computation.

Figure 1.5. Sensitivity of housing returns to a rent-price location correction



Note: Bars show the arithmetic- and geometric- average housing returns for selected sub-periods. Error bars show the impact on historical returns of increasing or reducing the benchmark price/rent ratio by ± 3, which broadly captures the difference between in- and out-of-city-center locations.

capitalization weighted, which again better captures the yield on a representative portfolio. Finally, we aim for national coverage with the historical rental indices used for extrapolation, and historical balance sheet benchmarks.

Despite this, it is likely that both our house price and rental data are somewhat biased towards cities and urban areas, especially for historical periods—simply because urban housing data are more widely available and researched. Even though this would affect the level of capital gain and yield, it should have little influence on total returns, since cities tend to have higher capital gains, but lower rental yields.[33] Additionally, Knoll, Schularick, and Steger (2017) show that the rural-urban divide has a relatively small impact on capital gains. Relatedly, we can establish some bounds on how much our rental yields can vary with the choice of location. In 2013, Numbeo.com data suggest that price-rent ratios in and out of city centres differ by less than 3 times annual rent. The rental yield is the inverse of these price-rent ratios. This motivates us to construct a lower bound rent-price ratio as $RP_{low} = 1/(1/RP_{actual} + 3)$ and an upper bound rent-price ratio as $RP_{high} = 1/(1/RP_{actual} - 3)$ for each country in 2013 to estimate upper and lower bounds of our housing returns depending on the choice of location. Given the currently high price-rent ratios, these adjustments have a relatively small impact on our data. Figure 1.5 shows that increasing or reducing the price-rent ratio by 3 changes annual return estimates by about ±1 pps per year relative to our preferred baseline.

33. Eisfeldt and Demers (2015) study the geographical distribution of returns on single family rentals in the US from 1970s to today and find that lower capital gain areas tend to have much higher rental yields, and there is very little geographic variation in total returns.

This suggests that the level of housing returns in our dataset should be representative of a country-wide portfolio. Still, it could be that returns on locations not included in our sample display higher volatility. For example, the post-1975 US indices are based on conforming mortgages and may exclude the more volatile segments of the market. To assess the likely magnitude of this bias, Table 1.5 compares the recent level and volatility of the US conforming mortgage based OFHEO house price indices with those that cover other segments of the market as well, which are sourced from Zillow.[34] Comparing columns 2 and 3 of Table 1.5, the nationwide moments of the data are similar across the two measures—but, as expected, the OFHEO data display slightly higher real capital gains and slightly lower volatility, because they have a less comprehensive coverage of the areas that were hit hardest by the subprime crisis, which receives a relatively high weight in the 1995–2015 US sample used here.

Columns 3–5 of Table 1.5 also show that the volatility of the housing series increases as we move from the aggregate portfolio (column 2) to the subnational and local level. The standard deviation of Zipcode-level housing returns is roughly one-third higher than that in the national data. If investors owned one undiversified house whose price tracks the neighborhood index, the risk and return characteristics of this portfolio would be somewhat closer to those of the aggregate equity index, although the gap would still be large.

Of course, it is much more difficult to invest in a diversified housing portfolio than a well-diversified equity portfolio. That being said, Benhabib and Bisin (2016) show that most equity is also held in an undiversified manner. The data regarding returns on individual housing and private equity returns are, however, at this point in time, very sparse. To understand exactly how these risk-return characteristics play out at a disaggregated level, a more detailed study of individual portfolios and returns is necessary. This would be a worthy goal of future research.

Another selection bias comes about from the fact that rent data largely come from the rental market, whereas the majority of housing stock is held by owner-occupiers. To guard against this, we rely on rental indices that, whenever possible, account for the growth of imputed rents. Additionally, we benchmark our series to the balance sheet yield data that are constructed to cover the total housing stock. Still, imputed rents are measured with error, and may not reflect the cost that the homeowner would pay on the rental market. If owning is relatively cheaper than renting—for example, due to tax exemptions or long-run house price appreciation—homeowners would purchase larger or better houses than they would rent, and imputed rents would overstate the value of housing services accruing to homeowners. On the other hand, buying a house is subject to credit constraints, which means

34. As we show later in Section 1.4.3, almost all the volatility in housing returns comes about from house prices. Therefore for the analysis of volatility, we focus on house prices rather than rental yields.

Table 1.5. Level and volatility of real housing capital gains at different levels of aggregation

	Baseline	Zillow			
	National	National	State	County	Zipcode
Mean real capital gain p.a.	1.42	0.79	1.07	0.53	0.92
Standard deviation	4.67	5.67	6.05	6.28	7.46

Note: US data, 1995–2015. Average annual real capital gain and standard deviation of house prices. Baseline data are sourced from the OFHEO index. Zillow data are sourced from the Zillow Home Value Index which covers around 95% of the US housing stock, and are averages of monthly values. National data are the returns and volatility of prices for a nationwide housing, and the other figures cover a representative state, county or zipcode level porftolio respectively.

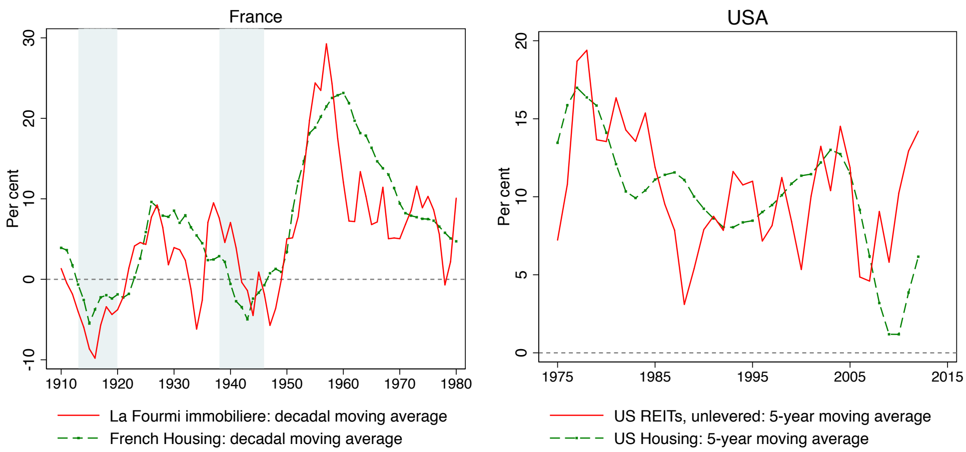
that renters can afford better houses than homeowners all else equal. In this case, imputed rents would understate the flow value of housing services. Overall, the direction of any potential bias is unclear and leaves much scope for future work.

Finally, the portfolio selection in the price and rent series is subject to survivorship bias. In general, our price and rental yield estimates aim to capture transaction or appraisal values, and rental costs on a broad and impartially selected portfolio of properties. Some survivorship bias may, however, enter the series for the following reasons. First, indices that rely on an ex-post selection of cities may inadvertently choose the more "successful" cities over the less successful ones. Second, houses that decline in value are likely to lose liquidity and be sold less frequently, hence carrying a lower weight in the index. And third, chain-linking historical house price and rent indices to compute annual returns will generally ignore the impact of large destructions of the housing stock, in particular those occurring during wartime.

Several factors suggest that the impact of survivorship bias on housing returns should be limited. First, Figure 1.5 and Knoll, Schularick, and Steger (2017) show that any location-specific bias in our estimates is likely to be small. Second, if the magnitude of survivorship bias is similar to that in equity markets (Section 1.3.2), then the bias is also unlikely to be large. Third, the low liquidity and weight of houses with declining prices is in some ways similar to the documented negative returns on delisted equities (Shumway, 1997; Shumway and Warther, 1999), which in general cannot be incorporated into the stock return series due to the lack of data. Therefore this bias should be less of a concern when comparing housing and equity returns. Finally, similarly to the stock market closures discussed in Section 1.3.2, even though capital stock destruction during wars can have a substantial impact on returns in specific years, it is unlikely to profoundly affect cross-country long-run returns due to the rarity of such events.[35] And as Figure 1.9 later shows, the main

35. As a reasonable upper bound, existing estimates suggest that around 33%–40% of the German housing stock was destroyed by Allied bombing during WW2 (Diefendorf, 1993; Akbulut-Yuksel, 2014), but this would lower the full-sample country-specific average annual real total return by only about 0.30 pps per year.

Figure 1.6. Returns on housing compared to real estate investment funds



Note: Total real return on housing, and shares of housing investment firms in France and US. Moving averages. Following Giacomini, Ling, and Naranjo (2015), we assume a 45% leverage ratio for US REITs.

facts in the data are similar for countries that experienced major war destruction on their own territory versus countries that did not (e.g., Australia, Canada, Sweden, Switzerland, and US). Further, Appendix Table 1.A.5 shows that housing offers a similar return relative to equity on average even after wars are excluded.

Returns on real estate investment trusts. Another way to check our housing returns is to compare them to the historical returns on housing investment trusts. These trusts offer independent estimates of returns. Real estate investment trusts, or REITs, are investment funds that specialize in the purchase and management of residential and commercial real estate. Many of these funds list their shares on the local stock exchange. The return on these shares should closely track total real estate returns. Differences will arise because the REIT portfolio is more geographically concentrated, its assets often contain non-residential property, and share price fluctuations may reflect expectations of future earnings and sentiment, as well as underlying portfolio returns. Further, the REIT portfolio returns should be net of taxes and transaction costs as well as housing running costs, and may therefore be somewhat lower than our housing series. Still, returns on the REIT portfolio should be comparable to housing and can be used to check the general plausibility of our return series.

Figure 1.6 compares our historical housing returns (dashed line) with those on investments in REITs (solid line) in France and US, two countries for which longer-run REIT return data are available. The REIT returns series for France refers to shares of the fund “La Fourmi Immobiliere” (see Simonnet, Gallais-Hamonno, and Arbulu, 1998). The fund acquired a portfolio of 15 properties in Paris between 1900 and 1913, worth around 36 million euros at 2015 prices, and its shares were listed on the Paris stock exchange between 1904 and 1997. We exclude the period after

1985, when "La Fourmi Immobiliere" was taken over by AGF. For the US, we use the FTSE NAREIT residential total return index after 1994, and the general FTSE equity NAREIT before. REIT returns have to be unlevered to capture the returns on the REIT housing portfolio. "La Fourmi Immobiliere" had an unlevered balance sheet structure, hence we do not adjust their returns. We assume a REIT leverage of 45% for the U.S. following Giacomini, Ling, and Naranjo (2015). Returns for France are presented as decadal moving averages, and for the US as five-year moving averages, given the shorter span of the US data.

Comparing the solid and dashed lines in Figure 1.6, the long-run levels of unlevered REIT and housing returns are remarkably similar. The time trend also follows a similar pattern, especially in France. The REIT returns, however, tend to be somewhat more volatile—most likely because they reflect changes in the market's valuations of future earnings, as well as the current portfolio performance. The REIT returns also seem to be affected by the general ups and downs of the stock market: for example, the 1987 "Black Monday" crash and dot-com bust in the U.S., as well as the 1930s Great Depression and 1960s stock crises in France. This suggests that the valuations of the funds' housing portfolios may have been affected by shifts in general stock market sentiment, possibly unrelated to housing market fundamentals.

Overall, the returns on real estate investment funds serve to confirm the general housing return level in our dataset. The comparison also suggests that returns in housing markets tend to be smoother than those in stock markets. The next section examines various factors that can affect the comparability of housing and equity returns more generally.

1.3.4 Comparability of housing and equity returns

Even if the fundamentals driving housing and equity returns (expected dividend/profit, and rental flows) are similar, investor returns for the two asset classes may differ for a number of reasons including taxes, transaction costs, and the financial structure of the investment claim. In this subsection we consider such comparability issues.

Transaction costs. The conventional wisdom is that while bonds and equities can be purchased with low transaction costs and at short notice, the seller of a house typically incurs significant costs. We provide a rough estimate of how transaction costs affect our return estimates for housing. We perform a simple back-of-the-envelope calculation to do this using contemporary data on average holding periods of residential real estate and average transaction costs incurred by the buyer. According to the (OECD, 2012), average round-trip transaction costs across 13 of the 16 countries in our sample amount to about 7.7 percent of the property's value.[36]

36. Data are available for Australia, Belgium, Switzerland, Germany, Denmark, Finland, France, U.K., Japan, the Netherlands, Norway, Sweden, and the U.S. Transaction costs are highest in Belgium

However, these simple cost ratios need to be adjusted for the typical trading frequency of each asset. According to the American Community Survey of 2007, more than 50 percent of U.S. homeowners had lived in their current home for more than 10 years. Current average holding periods are similar in, e.g., the U.K., Australia and the Netherlands. Another way to estimate housing turnover is using housing sales data, which for the US gives us an average holding period of close to 20 years.[37] Either way, accounting for transaction costs would thus lower the average annual return to housing to less than 100 basis points (e.g., 77 basis points per year based on a 7.7% cost incurred every 10 years).

For equities, the cost of each individual transaction is much smaller, but the number of transactions is much higher. Jones (2002) estimates that at the New York Stock exchange over the period 1900–2001, the average transaction cost was around 80 bps (half bid-ask spread of 30 bps plus commission rate of 50 bps), while turnover was roughly 60% per year, resulting in the annual average equity transaction costs of 40 bps. Comparing this number to the back-of-the-envelope housing transaction cost estimates reported above, it seems that even though equity transaction costs are probably somewhat lower, the difference between two asset classes is likely to be small—and no more than 0.5 pps per year.

The fact that housing faces much higher costs per each transaction, however, means that the realized housing transaction costs may understate the "shadow" utility cost which would include the suboptimal allocation choices from staying in the same house and not moving, for example. It might also reduce the volatility of housing returns, making them react more sluggishly to shocks. This means that the relatively modest short-run volatility of housing returns could mask more pronounced fluctuations at lower frequencies. Appendix 1.A.11 and Table 1.A.17 compare equity and housing return volatility over longer horizons of up to 20 years. It turns out that the standard deviation of housing returns is always around one-half that of equity returns, regardless of the time horizon, which suggests that housing returns not only have lower short-run volatility, but also less pronounced swings at all longer horizons.

Leverage. Household-level returns on real estate and equity will be affected by the structure of the household balance sheet, and how these investments are financed. Jordà, Schularick, and Taylor (2016) show that advanced economies in the second half of the 20th century experienced a boom in mortgage lending and borrowing. This surge in household borrowing did not only reflect rising house prices, it also reflected substantially higher household debt levels relative to asset values (and relative to household incomes). The majority of households in advanced economies

amounting to nearly 15 percent of the property value and lowest in Denmark amounting to only 1 percent of the property value.

37. Between April 2017 and March 2018, 5.5 millions existing homes were sold in the US at an average price of $250,000, which amounts to roughly one-twentieth of the total US housing stock.

today hold a leveraged portfolio in their local real estate market. As with any leveraged portfolio, this significantly increases both the risk and the return associated with the investment. And today, unlike in the early twentieth century, houses can be levered much more than equities. The benchmark rent-price ratios from the IPD used to construct estimates of the return to housing refer to rent-price ratios of unleveraged real estate. Consequently, the estimates presented so far constitute only un-levered housing returns of a hypothetical long-only investor, which is symmetric to the way we (and the literature) have treated equities.

However, computing raw returns to housing and equity indices neglects the fact that an equity investment contains embedded leverage. The underlying corporations have balance sheets with both debt and equity liabilities. Thus, reconciliation is needed, and two routes can be taken. For truly comparable raw un-levered returns, equity returns could be de-levered. Alternatively, for truly comparable levered returns, housing returns would have to be levered up to factor in the actual leverage (using mortgages) seen on household balance sheets. Is this a big deal in practice? We argue that it does not bias our conclusions significantly based on some elementary calculations.

Consider, for example, the second reconciliation of levering up housing returns. Let the real long-term mortgage borrowing rate be r_0, and α be the leverage of the average house proxied by total mortgages divided by the value of the housing stock. Then we can solve for levered real housing returns TR' as a function of un-levered real housing returns TR using the formula $TR' = (TR - \alpha r_0))/(1-\alpha)$. In our data, $TR \approx 7.0\%$ and $\alpha \approx 0.2$. Using long bond return as a proxy for r_0 of around 2.5% p.a., this would imply $TR' = 8.1\%$.[38] In other words, for the representative agent the levered housing return is about 110 bps higher than the unlevered housing return (8.1% versus 7%), a small difference. Such adjustments appear to be inconsequential for the main conclusions we present in this paper. In fact, they would bolster one of our central new claims which is that real housing returns at least match or even exceed real equity returns in the long run when the two are compared on an equal footing.

Taxes. When computing equity and housing returns we do not account for taxes. From an investor's perspective accounting for taxes is clearly important. Typically, equity capital gains—and, for some countries and periods, even dividend income—have been subject to a capital gains tax. When dividends are not taxed as capital gains, they tend to be taxed as income. In some countries, housing capital gains are subject to capital gains taxes, but owner-occupied houses in particular have been

38. For evidence on α, the average economy wide housing leverage measured by total mortgages divided by the value of the housing stock, see Jordà, Schularick, and Taylor (2016). If one preferred to use the mortgage rate rather than the long bond in this calculation, the evidence in Zimmermann (2017) points to an average real mortgage rate r_m of around 3% p.a. This would imply $TR' = 8\%$, only slightly lower than the figure quoted in the main text.

granted exemptions in many cases. Imputed rents of homeowners are, unlike dividend income, almost never taxed. Additionally, housing tends to be subject to asset-specific levies in the form of property taxes, documented extensively in Appendix 1.B.6.

For both equities and housing, the level and applicability of taxes has varied over time. For housing, this variation in treatment also extends to assessment rules, valuations, and tax band specifications. As a ballpark estimate, the impact of property taxes would lower real estate returns by less than 1.0 pps per year relative to equity (see Appendix 1.B.6 for further details). The various exemptions for homeowners make the impact of capital gains taxes on real estate returns even harder to quantify but such exemptions also imply that differential tax treatment is unlikely to play an important role in explaining differences in the return between equities and housing.[39]

Since quantifying the time- and country-varying effect of taxes on returns with precision is beyond the scope of this study, throughout this paper we focus on pre-tax returns from an investor perspective. Importantly, these pre-tax returns are net of corporate profit taxes, which are netted out before the cashflow payment to the investor. Studies of returns from an aggregate wealth perspective such as Piketty, Saez, and Zucman (2018) typically compute business equity returns gross of corporate tax. Appendix 1.A.19 discusses the impact of adding back corporate taxes on our return data. Equity returns before corporate tax would be roughly 1 percentage point higher than our baseline estimates (Table 1.A.27). This adjustment is, however, very much an upper bound on the housing-equity return differential for the following reasons. First, as noted above, a true like-for-like comparison should also delever equity returns and compare the returns on business and housing wealth. Appendix 1.A.19 Table 1.A.27 estimates that first adding back corporate taxes, and then delevering equity returns leaves them approximately equal to the baseline estimates that we report. Second, the total tax burden on the pre-corporate-tax equity returns is likely to be higher than on housing, since in light of the various homeowner exemptions, the post-corporate-tax burden on the two assets appears to be roughly similar. Third, the returns on the two asset classes are similar before 1920, when the corporate tax rate was close to zero.

Temporal aggregation and return averaging. The way house and equity price indices are constructed is likely to influence the volatility of the return series. The house price indices used for return calculations (e.g., indices from national statistical agencies) tend to be an *average* of all transactions in a given year, or use a sample of transactions or appraisal values throughout the year. But the equity prices used for

39. Note that whilst this is true for aggregate or owner-occupied housing, the tax burden on landlords is likely to be somewhat higher than that on holders of listed equity, because landlords do not benefit from the homeowner exemptions to property taxes, and their rental income is taxed.

Table 1.6. Impact of using end-of-year versus yearly-average asset prices

	Equity (MSCI index)		Housing (this paper)
	End-of-year	Yearly average	Yearly average
Mean return p.a.	8.70	7.51	6.55
Standard deviation	27.56	22.00	7.45
Observations	694	694	694

Note: Annual global real returns in 16 countries, equally weighted, 1970–2015. End-of-year returns are computed using the return index value for the last day of the year. Yearly average returns are computed using the average index value throughout the year.

return calculations, by the usual convention followed here, on the contrary, compare *end-of-year* prices of shares. The use of end-of-year rather than yearly-average prices mechanically makes equity returns more volatile.

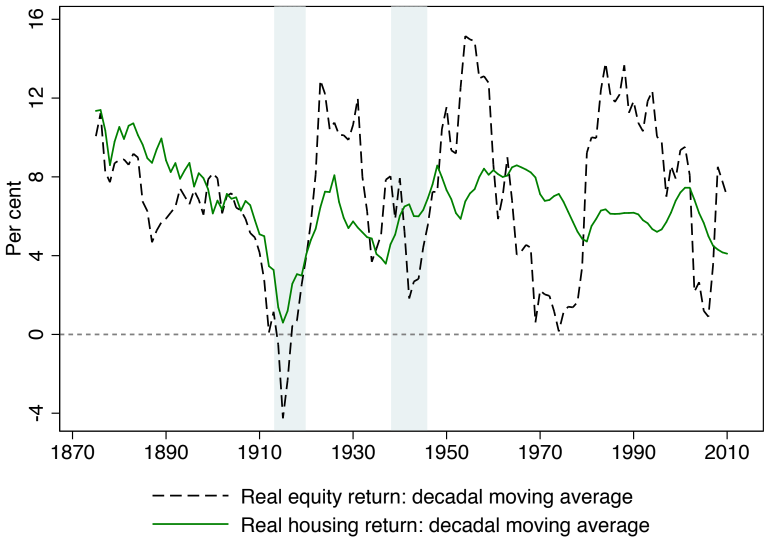
We can assess the magnitude of this effect by constructing an equity return index based on annual averages of daily data, to mimic the way housing returns are computed, and then comparing it to a “normal” return using end-of-year index values. For this robustness exercise we use daily MSCI equity returns data for 1970–2015. Table 1.6 presents the end-of-year and yearly-average real equity returns in the first two columns, and our yearly-average housing returns for the same time period in the third column. Comparing the first two columns shows that yearly averaging lowers the standard deviation of returns by around one-fifth, or 5 pps. It also lowers the average return by around 1 ppt, because the return series is a transformation of the raw price data, and lowering the variance reduces the mean of the return. But the standard deviation of the smoothed yearly-average equity series is still almost three times that of housing over the same time period.

Because historical house price data sometimes rely on relatively few transactions, they are likely to be somewhat less smooth than averages of daily data. Therefore Table 1.6 provides an upper bound of the impact of averaging on our return series. Even taking this upper bound at face value, the averaging of house price indices is likely to explain some, but far from all, of the differences in volatility of equity and housing returns.

1.4 Risky rates of return

At this waystation the lengthy pilgrimage of Section 1.3 ends: the numerous details of how we compiled our data; the many important, but somewhat technical, aspects of data construction; the extensive accuracy checks. In these next sections the pace picks up and the focus turns to analysis and interpretation of the data. We examine broad trends and explore their implications for how we think about macroeconomics and finance, confronting the four big themes laid out in the introduction: the long-

Figure 1.7. Trends in real returns on equity and housing



Note: Mean returns for 16 countries, weighted by real GDP. Decadal moving averages.

run behavior of risky returns, safe returns, risk premia, and "*r* minus *g*." Readers who skipped the better part of Section 1.3 are welcomed to rejoin the flow here.

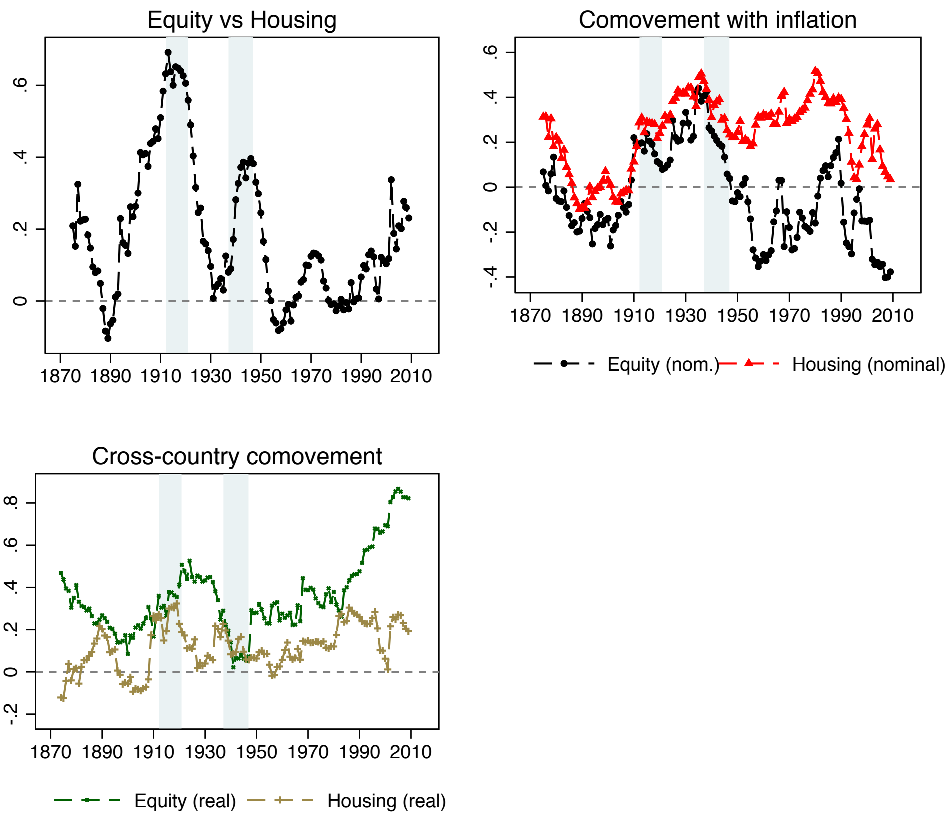
1.4.1 Global returns

We first turn in Figure 1.7 to a more detailed inspection of the returns on the risky assets, equity and housing. The global returns are GDP-weighted averages of the 16 countries in our sample. Although we do not show the unweighted data, the corresponding figure would look very similar. We smooth the data using a decadal moving averages as explained earlier. For example, the observation reported in 1900 is the average of data from 1895 to 1905. Figure 1.7 shows the trends in decadal-average real returns on housing (solid line) and equity (dashed line) for our entire sample. In addition, Figure 1.8 displays the average rolling decadal correlation of annual risky returns between asset classes, across countries, and with inflation.

Risky returns were high in the 1870s and 1880s, fell slowly at first, but then sharply after 1900, with the decade-average real equity returns turning negative during WW1. Risky returns recovered quickly in the 1920s, before experiencing a drop in the the Great Depression, especially for equities. Strikingly, after WW2 the trajectories of returns for the two risky asset classes had similar long-run means but over shorter periods diverged markedly from each other.

Equity returns have experienced many pronounced global boom-bust cycles, much more so than housing returns, with average real returns as high as +16% and as low as –4% over entire decades. Equity returns fell in WW2, boomed in the post-war reconstruction, and fell off again in the climate of general macroeconomic instability in the 1970s. Equity returns bounced back following a wave of deregula-

Figure 1.8. Correlations across risky asset returns



Note: Rolling decadal correlations. The global correlation coefficient is the average of individual countries for the rolling window. Cross-country correlation coefficient is the average of all country pairs for a given asset class. Country coverage differs across time periods.

tion and privatization in the 1980s. The next major event was the Global Financial Crisis, which exacted its toll on equities and to some extent housing, as we shall see.

Housing returns, on the other hand, have remained remarkably stable over almost the entire post-WW2 period. As a consequence, the correlation between equity and housing returns, depicted in the top left panel of Figure 1.8, was highly positive before WW2, but has all but disappeared over the past five decades. The low covariance of equity and housing returns over the long run reveals potential attractive gains from diversification across these two asset classes that economists, up to now, have been unable to measure or analyze.

In terms of relative returns, we see that housing persistently outperformed equity up until WW1, even though both of these asset returns followed a broadly similar temporal pattern. In recent decades, equities have slightly outperformed housing in (arithmetic, not geometric) average, but with much higher volatility and cyclicality. Furthermore, upswings in equity prices have generally not coincided with times of low growth or high inflation, when standard asset pricing theory would say high returns would have been particularly valuable.

The top-right panel of Figure 1.8 examines the correlation between risky rates of return and inflation. It shows that equity co-moved negatively with inflation in the 1970s, while housing provided a more robust hedge against an unusually rapid surge consumer prices. In fact, apart from the interwar period, when the world was gripped by a broad deflationary bias, we find that equity returns have co-moved negatively with inflation in almost all eras. Moreover, the big downswings in equity returns in the two world wars and the 1970s coincided with periods of generally poor economic performance.

In the past two decades equity returns have also become highly correlated across countries, as shown by the sharp rise in the degree of cross-country comovement in the bottom-left panel of Figure 1.8, measured as the average of all country-pair correlations for a given window.[40] A well-diversified global equity portfolio has thus become less of a hedge against country-specific risk (Quinn and Voth, 2008). As is a matter of debate, this may reflect greater freedom to arbitrage and trade across equity markets globally, or an increase in the global shocks to which firms, especially those in the typical equity index, are increasingly exposed. In contrast to equities, cross-country housing returns have remained relatively uncorrelated, perhaps because housing assets remain less globally tradable than equities, or because they are more exposed to idiosyncratic country-level shocks.

1.4.2 Country returns

Next we explore risky returns in individual countries. Table 1.7 shows returns on equities and housing by country for the full sample and for the post-1950 and post-1980 subsamples. Long-run risky asset returns for most countries are close to 6%–8% per year, a figure which we think represents a robust and strong real return to risky capital. Still, the figures also show an important degree of heterogeneity among countries. Many of the countries that experienced large political shocks show lower equity returns. This is the case for Portugal and Spain which both underwent prolonged civil strife, and France which undertook a wave of nationalizations in the aftermath of WW2. French equity returns are also negatively affected by the fallout from the world wars, and the fallout from an oil crisis in the 1960s (for more detail, see Le Bris and Hautcoeur, 2010; Blancheton, Bonin, and Le Bris, 2014). In contrast, real equity returns in Finland have been as high as 10%, on average throughout the

40. We report the average of all country-pair combinations for a given window, calculated as

$$Corr_{i,t} = \frac{\sum_j \sum_{k \neq j} Corr(r_{i,j,t \in T}, r_{i,k,t \in T})}{\sum_j \sum_{k \neq j} 1}$$

for asset i (here: equities or housing), and time window $T = (t-5, t+5)$. Here j and k denote the country pairs, and r denotes real returns, constructed as described in Section 1.2.3.

Table 1.7. Real rates of return on equity and housing

Country	Full Sample		Post 1950		Post 1980	
	Equity	Housing	Equity	Housing	Equity	Housing
Australia	7.79	6.37	7.53	8.29	8.70	7.16
Belgium	6.23	7.89	9.65	8.14	11.49	7.20
Denmark	7.49	8.22	9.73	7.04	13.30	5.14
Finland	10.03	9.58	12.89	11.18	16.32	9.47
France	3.21	6.39	6.01	9.68	9.61	5.78
Germany	7.11	7.82	7.53	5.30	10.07	4.13
Italy	7.25	4.77	6.09	5.55	9.45	4.57
Japan	6.00	6.54	6.21	6.74	5.62	3.58
Netherlands	6.96	7.28	9.19	8.53	11.51	6.41
Norway	5.67	8.03	7.33	9.10	12.22	9.82
Portugal	4.51	6.31	4.84	6.01	8.60	7.15
Spain	5.83	5.21	7.75	5.83	11.96	4.62
Sweden	8.02	8.30	11.37	8.94	15.87	9.00
Switzerland	6.51	5.63	8.37	5.64	9.29	6.19
UK	6.83	5.44	9.10	6.57	9.11	6.81
USA	8.46	6.10	8.89	5.76	9.31	5.86
Average, unweighted	6.67	7.26	8.30	7.47	10.78	6.43
Average, weighted	7.12	6.72	8.19	6.40	9.08	5.50

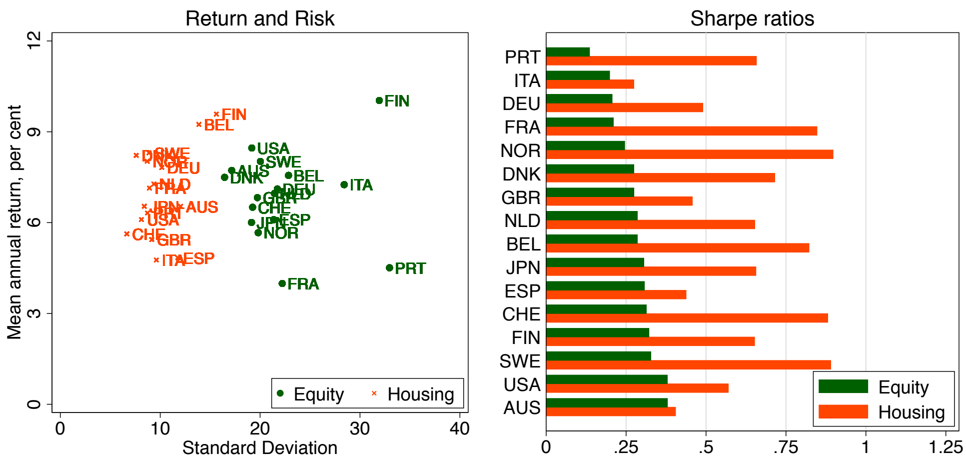
Note: Average annual real returns. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both real housing and equity returns. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

sample. Housing returns also show considerable heterogeneity. Returns on housing have been high on average in the Nordic countries, but low in Italy and Spain. US risky asset returns fall roughly in the middle of the country-specific figures, with equity returns slightly above average, and housing returns slightly below. Our estimates of post-WW2 US housing returns are in line with those in Favilukis, Ludvigson, and Van Nieuwerburgh (2017).[41] The degree of heterogeneity and the relative ranking of returns is broadly similar when comparing the full sample to the post-1950 period.

This country-level evidence reinforces one of our main findings: housing has been as good a long-run investment as equities, and possibly better. Housing has offered a similar return to equity in the majority of countries and time periods. In the long-run, housing outperformed equities in absolute terms in 6 countries, and equities outperformed housing in 5. Returns on the two assets were about the same in the remaining 5 countries. After WW2, housing was the best-performing asset class in 3 countries, and equities in 9.

41. Favilukis, Ludvigson, and Van Nieuwerburgh (2017) estimate a gross nominal return on US housing of 9%–11%, based on three data sources going back to 1950s and 1970s. This implies a net real return of around 5%–7% (once inflation, maintenance and running costs are subtracted), in line with our estimates in Table 1.7.

Figure 1.9. Risk and return of equity and housing



Note: Left panel: average real return p.a. and standard deviation. Right panel: Sharpe ratios, measured as $(\bar{r}_i - \bar{r}_{bill})/\varsigma_i$, where i is the risky asset with $\bar{r}_i$ mean return and ς_i standard deviation. 16 countries. Consistent coverage within each country.

However, although aggregate total returns on equities exceed those on housing for certain countries and time periods, equities do not outperform housing in simple risk-adjusted terms. Figure 1.9 compares the risk and returns of housing and equities for each country. The left panel plots average annual real returns on housing and equities against their standard deviation. The right panel shows the Sharpe ratios for equities and housing for each country in the sample.[42] Housing provides a higher return per unit of risk in each of the 16 countries in our sample, with Sharpe ratios on average more than double those for equities.

1.4.3 Decomposition of returns

To further look into the underlying drivers of housing and equity returns, we decompose them into the capital gain (price) and yield (dividend or rent) components. To be consistent with the data in Section 1.3 and Table 1.2, we decompose *real* total return into *real* capital gain—that is, the price change net of inflation—and dividend or rental yield—that is, the nominal yield as proportion of the previous year's share or house price.[43] Yet caveats arise. In principle, it is not entirely clear whether inflation should be subtracted from the capital gain or yield component. Moreover, firms may buyback shares to generate low-tax capital gains instead of paying out

42. The Sharpe ratio is calculated as $(\bar{r}_i - \bar{r}_{bill})/\sigma_i$, where i is the risky asset (housing or equity) with $\bar{r}_i$ mean return and σ_i standard deviation.

43. The small residual difference between combined capital gain and dividend income, and the equity total return, accounts for gain and loss from capital operations such as stock splits or share buybacks, and income from reinvestment of dividends.

Table 1.8. Total return components for equity and housing

	Equity			Housing		
	Real capital gain	Dividend income	Real total return	Real capital gain	Rental income	Real total return
Full sample:						
Mean return p.a.	2.78	4.17	6.82	1.61	5.50	6.92
Standard deviation	21.37	1.74	21.89	9.87	2.05	10.40
Geometric mean	0.57	4.16	4.58	1.15	5.48	6.43
Observations	1707	1707	1707	1707	1707	1707
Post-1950:						
Mean return p.a.	4.73	3.80	8.36	2.39	5.22	7.38
Standard deviation	23.70	1.81	24.24	8.59	1.93	8.95
Geometric mean	2.03	3.79	5.62	2.06	5.21	7.04
Observations	995	995	995	995	995	995

Note: Average annual real capital gain, dividend or rental income, and total return across 16 countries, unweighted. Standard deviation in parentheses. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both equity and housing returns, capital gains and yields. Dividend and rental income expressed in percent of previous year's asset price.

higher-taxed dividends; thus, the manner of distribution of total returns may not be invariant to circumstances.

Table 1.8 decomposes equity and housing returns into capital gains and dividends or rents, for the full cross-country sample and the period after 1950. Over the full sample, most of the real return is attributable to the yield. Dividends account for roughly 60% of real equity returns, and rents for roughly 80% of real housing returns. In terms of geometric means (Table 1.8, row 3), almost all of both equity and housing returns are attributable to, respectively, dividend and rental income. After 1950, capital gains become more important for both equities and housing. For equities, real capital gains account for the majority of the total return after 1950, and for housing for roughly one-third.

The importance of dividends and rents is partly a matter of convention. Appendix 1.A.14 and Appendix Table 1.A.20 computes the equivalent decomposition for nominal returns, and finds that the capital gain versus dividend/rental income split is then closer to roughly 50/50. Nevertheless, without dividends or rents, the real returns on both assets would be low, especially in geometric mean terms. This is consistent with the existing literature on real house prices: Shiller (2000) documents that house prices in the US moved in line with inflation before the 2000s bubble, and Knoll, Schularick, and Steger (2017) show that real house prices in advanced

Table 1.9. Total return components for equity and housing by country

	Equity				Housing			
	Real capital gain	Dividend income	Real total return	Capital gain share	Real capital gain	Rental income	Real total return	Capital gain share
Australia	3.06	4.90	7.79	0.38	2.53	3.99	6.37	0.24
	(16.30)	(1.08)	(16.94)		(11.94)	(0.92)	(11.92)	
Belgium	2.53	3.83	6.23	0.40	1.95	6.15	7.89	0.14
	(22.92)	(1.64)	(23.61)		(15.05)	(1.46)	(15.51)	
Denmark	2.71	4.95	7.49	0.35	1.26	7.13	8.22	0.08
	(16.14)	(2.09)	(16.45)		(7.02)	(2.42)	(7.60)	
Finland	5.19	5.08	10.03	0.51	2.82	7.14	9.58	0.17
	(31.18)	(1.95)	(31.93)		(14.61)	(2.86)	(15.62)	
France	-0.37	3.73	3.21	0.09	1.55	5.09	6.39	0.13
	(21.57)	(1.33)	(22.14)		(9.39)	(1.14)	(10.03)	
Germany	2.74	4.08	7.11	0.40	1.86	6.03	7.82	0.13
	(20.99)	(1.58)	(21.72)		(9.24)	(2.61)	(10.16)	
Italy	3.78	3.61	7.25	0.51	1.45	3.49	4.77	0.18
	(27.99)	(1.34)	(28.42)		(9.28)	(1.59)	(9.61)	
Japan	3.12	2.65	6.00	0.54	2.00	4.70	6.54	0.18
	(18.94)	(1.77)	(19.15)		(7.99)	(1.24)	(8.41)	
Netherlands	3.38	4.87	8.10	0.41	1.75	5.96	7.51	0.13
	(19.21)	(1.57)	(19.61)		(8.22)	(1.68)	(8.76)	
Norway	1.61	4.21	5.67	0.28	1.49	6.72	8.03	0.10
	(19.33)	(1.60)	(19.82)		(8.26)	(1.19)	(8.70)	
Portugal	2.92	2.28	5.11	0.56	1.13	4.47	5.21	0.12
	(34.34)	(1.22)	(34.73)		(9.26)	(1.98)	(9.37)	
Spain	1.80	4.53	5.83	0.28	1.26	4.16	5.21	0.13
	(20.48)	(2.30)	(21.15)		(11.59)	(1.60)	(12.00)	
Sweden	4.08	4.12	8.02	0.50	1.39	7.12	8.30	0.09
	(19.54)	(1.05)	(20.03)		(8.46)	(1.61)	(8.88)	
Switzerland	3.17	3.20	6.27	0.50	0.81	4.54	5.24	0.08
	(20.61)	(1.46)	(20.73)		(6.50)	(0.62)	(6.74)	
UK	2.48	4.53	6.83	0.35	1.63	3.94	5.44	0.17
	(19.12)	(1.39)	(19.73)		(8.94)	(0.86)	(9.15)	
USA	4.19	4.38	8.46	0.49	0.90	5.33	6.10	0.08
	(18.90)	(1.57)	(19.17)		(7.84)	(0.75)	(8.12)	

Note: Arithmetic average of annual real capital gain, dividend or rental income, and total return, full sample. Standard deviation in parentheses. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both equity and housing returns, capital gains and yields. Dividend and rental income expressed as percentage of previous year's asset price. Capital gain share is the proportion of real total return attributable to real capital gains.

economies were more or less flat before 1950. This is also true in our data: the pre-1950 annual real housing capital gains are just 0.5%. Post-1950 capital gains are somewhat higher at 2.5%, but still only half the magnitude of the rental yields. Adding rents to the equation radically changes the picture, and brings the long-run housing total return close to 7%. Interestingly, the broad picture is similar for equities: the real equity capital gain before 1950 is, on average, just 0.4%, compared to 4.7% per year after 1950. However, the contribution of dividend and rental income means that housing and equity returns were high both before and after 1950.

While most of the return can be attributed to dividends and rents, almost all of the volatility comes from equity and house prices, i.e., the capital gains component. Table 1.8 row 2 shows that both dividends and rents are very stable, with a standard deviation of dividend/rental yields of about 2%. Prices, on the contrary, move around much more, with standard deviation roughly equal to that of total returns (21.4% for equities and 9.9% for housing). The higher variability of equity returns, and the superior risk-adjusted performance of housing can, therefore, largely be attributed to the lower volatility of house prices compared to those of equities. An additional factor is that capital gains—the more volatile component—account for a relatively larger share of equity returns.

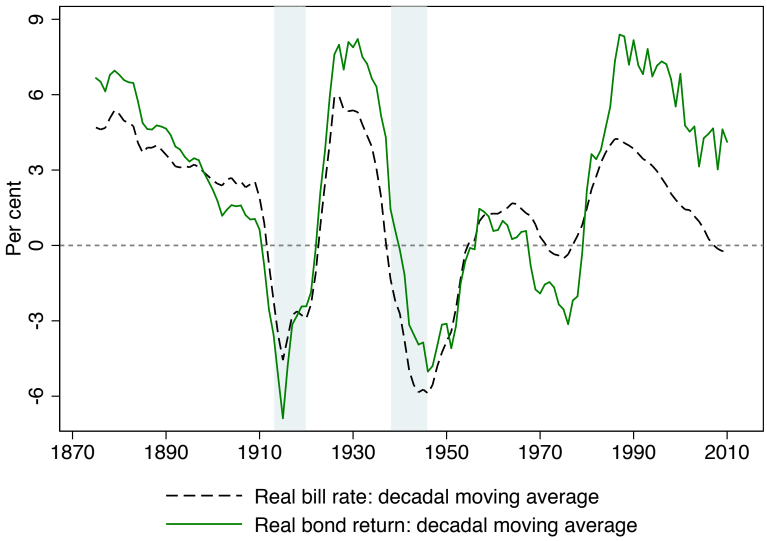
Since aggregate equity prices are subject to large and prolonged swings, a representative investor would have to hold on to his equity portfolio for longer in order to ensure a high real return. Aggregate housing returns, on the other hand, are more stable because swings in aggregate house prices are generally less pronounced. National aggregate housing portfolios have had comparable real returns to national aggregate equity portfolios, but with only half the volatility.

Table 1.9 examines the relative importance of capital gains versus dividends/rents at the country level (figures are arithmetic means and standard deviations). Additionally we report the share of real total return accounted for by capital gains. The fact that the majority of housing returns come from yields is true across all countries. The lowest real capital gains are observed in Switzerland, and the highest in Finland, but none exceed 3% per year in the full sample. Rents are relatively more important in the US, accounting for roughly 90% of returns, but this is not unusual: Denmark, Sweden and Switzerland have the same share of capital gains as the US. For equities, the picture is more mixed. Seven countries, including the US, have a roughly 50/50 split between real capital gain and dividend yield shares. Other countries record low or negative real capital gains over the full sample, and especially so in geometric mean terms (see Appendix Table 1.A.22).

1.5 Safe rates of return

Turning to safe asset returns, Figure 1.10 shows the trends in real returns on government bonds (solid line) and bills (dashed line) since 1870. Again, returns are GDP-weighted averages of the 16 countries in our sample; the corresponding unweighted

Figure 1.10. Trends in real returns on bonds and bills



Note: Mean returns for 16 countries, weighted by real GDP. Decadal moving averages.

figure would look very similar. We smooth the data using a decadal moving averages as explained earlier.

Three striking features of Figure 1.10 deserve comment. First, low real rates and, in fact, negative real rates have been relatively common during modern financial history. Second, for the most part, returns to long-term and short-term safe assets have tracked each other very closely—with a premium of about 1% that has widened considerably since the well-documented decline of the mid-1980s (see, e.g., Holston, Laubach, and Williams, 2017). Third, a major stylized fact leaps out once we compare the safe rates of return in Figure 1.10 to the risky rates of return in Figure 1.7 above. Prior to WW2, real returns on housing, safe assets, and equities followed remarkably similar trajectories. After WW2 this was no longer the case.

Safe rates are far from stable in the medium-term. There is enormous time-series, as well as cross-country, variability. In fact, real safe rates appear to be as volatile as real risky rates (sometimes more volatile), a topic we return to in the next section. Considerable variation in the risk premium often comes from sharp changes in safe real returns, not from real returns on risky assets.

Two four-decade-long declines in real rates stand out: (1) from 1870 to WW1 (with a subsequent further collapse during the war); and (2) the well-documented decline that started in the mid-1980s. We could add to this list the briefer, albeit more dramatic decline that followed the Great Depression into WW2. Some observers have therefore interpreted the recent downward trend in safe rates as a sign of a new era of "secular stagnation" (see, e.g., Summers, 2014).

However, in contrast to 1870–1913 and the 1930s, the more recent decline is characterized by a much higher term premium—a feature with few precedents in our sample.[44] There are other periods in which real rates remained low, such as in the 1960s. They were pushed below zero, particularly for the longer tenor bonds, during the 1970s inflation spike, although here too term premiums remained relatively tight. Returns also dipped dramatically during both world wars. This is perhaps to be expected: demand for safe assets spikes during disasters although the dip may also reflect periods of financial repression and high inflation that usually emerge during times of conflict, and which often persist into peacetime. Thus, from a broad historical perspective, high rates of return on safe assets and high term premiums are more the exception than the rule.

Summing up, over more than 140 years from the late 19th to the 21st century, real returns on safe assets have been low—on average 1% for bills and 2.5% for bonds—relative to alternative investments. Although the return volatility—measured as annual standard deviation—is lower than that of housing and equities, these assets offered little protection during high-inflation eras and during the two world wars, both periods of low consumption growth.

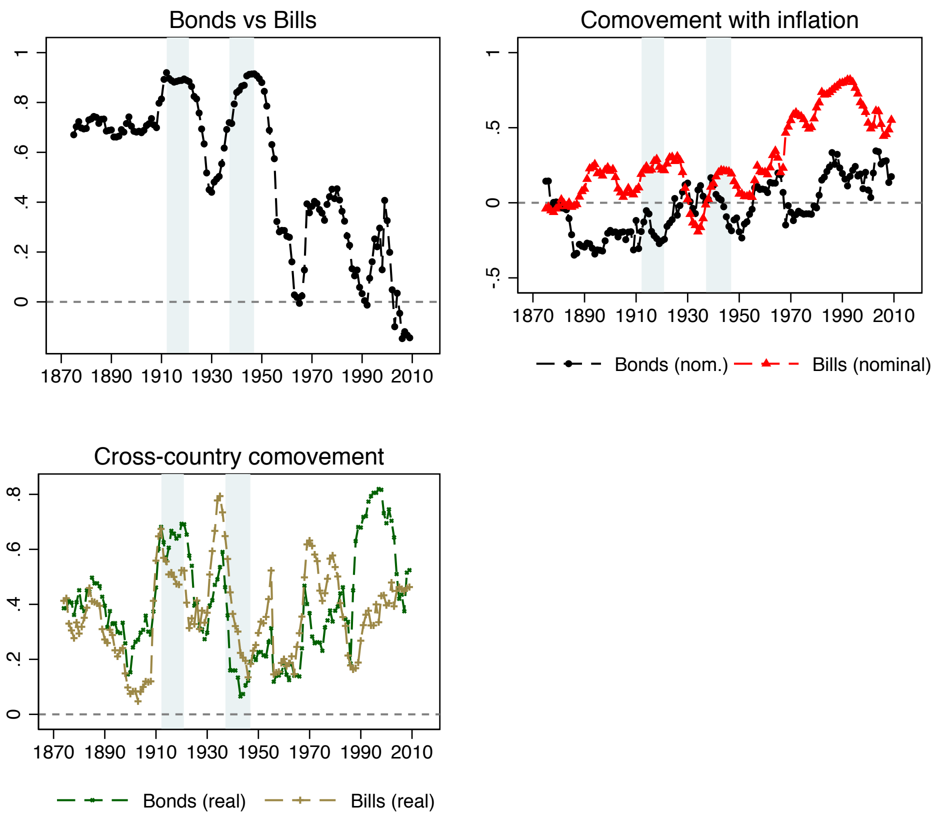
Again moving on to examine correlations, Figure 1.11 explores additional key moments of the data. The top-left panel plots the correlation between real bond and real bill returns, again using decadal rolling windows and computed as the cross-sectional average of correlations. In parallel to our discussion of the term premium, real returns on bonds and bills have been highly correlated for most of the sample up until the 1960s. From the 1970s onwards, the era of fiat money and higher average inflation, this correlation has become much weaker, and near zero at times, coinciding with a widening term premium.

The top right panel of Figure 1.11 displays the correlation between nominal safe asset returns and inflation, both for real bond and real bill returns. The figure shows that safe assets provided more of an inflation hedge starting in the 1970s, around the start of the era of modern central banking. However, as Figure 1.10 showed, both bonds and bills have experienced prolonged periods of negative real returns—both during wartime inflations, and in the high-inflation period of the late 1970s. Although safe asset rates usually comove positively with inflation, they do not always compensate the investor fully.

The bottom panel of Figure 1.11 displays the cross correlation of safe returns over rolling decadal windows, averaged for all country-pair combinations, to examine how much risk can be diversified with debt instruments. Cross-country real safe returns have exhibited positive comovement throughout history. The degree of comovement shows a few marked increases in WW1 and in the 1930s. The effect of

44. One important qualification is that this is the ex post, not ex ante term premium. It therefore includes any unexpected shocks that affect either the short rate or the long-run bond return series. Furthermore, because the long-run bond return measure includes capital gains, and the short-term rate measure is the yield only (since the security matures within one year), most of the post-1980 increase in the term premium is driven by higher capital gains on long-term government bonds.

Figure 1.11. Correlations across safe asset returns



Note: Rolling decadal correlations. The global correlation coefficient is the average of individual countries for the rolling window. Cross-country correlation coefficient is the average of all country pairs for a given asset class. Country coverage differs across time periods.

these major global shocks on individual countries seems to have resulted in a higher correlation of cross-country asset returns.

Turning to cross-sectional features, Table 1.10 shows country-specific safe asset returns for three samples: all years, post-1950, and post-1980. Here the experiences of a few countries stand out. In France, real bill returns have been negative when averaged over the full sample. In Portugal and Spain, they have been approximately zero. In Norway, the average return on bills has been negative for the post-1950 sample. However, most other countries have experienced reasonably similar returns on safe assets, in the ballpark of 1%–3%.

Aside from the investor perspective discussed above, safe rates of return have important implications for government finances, as they measure the cost of raising and servicing government debt. What matters for this is not the level of real return *per se*, but its comparison to real GDP growth, or $r^{safe} - g$. If the rate of return exceeds real GDP growth, $r^{safe} > g$, reducing the debt/GDP ratio requires continuous budget surpluses. When r^{safe} is less than g, however, a reduction in debt/GDP is possible even

Table 1.10. Real rates of return on bonds and bills

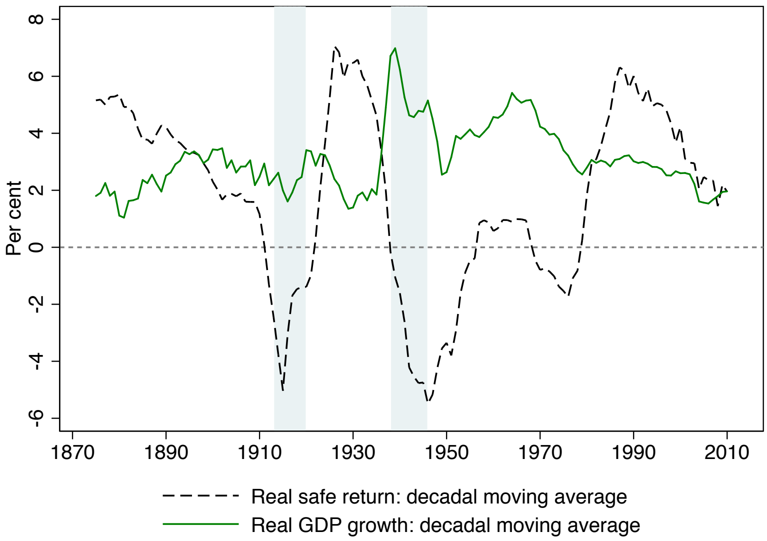
Country	Full Sample		Post 1950		Post 1980	
	Bills	Bonds	Bills	Bonds	Bills	Bonds
Australia	1.29	2.24	1.32	2.45	3.23	5.85
Belgium	1.21	3.01	1.61	3.86	2.51	6.24
Denmark	3.08	3.58	2.18	3.50	2.80	7.13
Finland	0.64	3.22	0.63	4.86	2.61	5.76
France	-0.47	1.54	0.96	2.97	2.24	6.96
Germany	1.51	3.15	1.86	3.70	1.97	4.23
Italy	1.20	2.53	1.30	2.83	2.42	5.85
Japan	0.68	2.54	1.36	2.83	1.48	4.53
Netherlands	1.37	2.71	1.04	2.14	2.08	5.59
Norway	1.10	2.55	-0.26	1.94	1.50	5.62
Portugal	-0.01	2.23	-0.65	1.59	0.65	6.25
Spain	-0.04	1.41	-0.32	1.21	2.20	5.72
Sweden	1.77	3.25	0.82	2.71	1.52	6.60
Switzerland	0.89	2.41	0.12	2.33	0.33	3.35
UK	1.16	2.29	1.14	2.63	2.70	6.67
USA	2.23	2.85	1.43	2.77	1.91	5.90
Average, unweighted	1.14	2.61	0.91	2.77	2.01	5.77
Average, weighted	1.34	2.51	1.23	2.70	1.98	5.64

Note: Average annual real returns. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both real bill and bond returns. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

with the government running modest deficits. Existing evidence points to $r^{safe} < g$ being the norm rather than the exception, both in recent years and broader historical data (Ball, Elmendorf, and Mankiw, 1998; Mehrotra, 2017).

Figure 1.12 plots the representative "safe rate of return" as the arithmetic average of bond and bill returns (dashed line) alongside real GDP growth (solid line), again as decadal moving averages. Starting in the late 19th century, safe rates were higher than GDP growth, meaning that any government wishing to reduce debt had to run persistent budget surpluses. Indeed, this was the strategy adopted by Britain to pay off the debt incurred during the Napoleonic War (Crafts, 2016). The two world wars saw low real returns, but nevertheless a large debt accumulation to finance the wartime effort. The aftermath of these two wars, however, offered vastly different experiences for public finances. After WW1, safe returns were high and growth low, requiring significant budgetary efforts to repay the war debts. This was particularly difficult for many countries given the large interlocking reparations imposed by the Treaty of Versailles, and the turbulent macroeconomic environment at the time. After WW2, on the contrary, high growth and inflation helped greatly reduce the value of national debt, creating $r^{safe} - g$ gaps as large as –10 percentage points.

Figure 1.12. Trends in real return on safe assets and GDP growth



Note: Mean returns and GDP growth for 16 countries, weighted by real GDP. Decadal moving averages. The safe rate of return is an arithmetic average of bonds and bills.

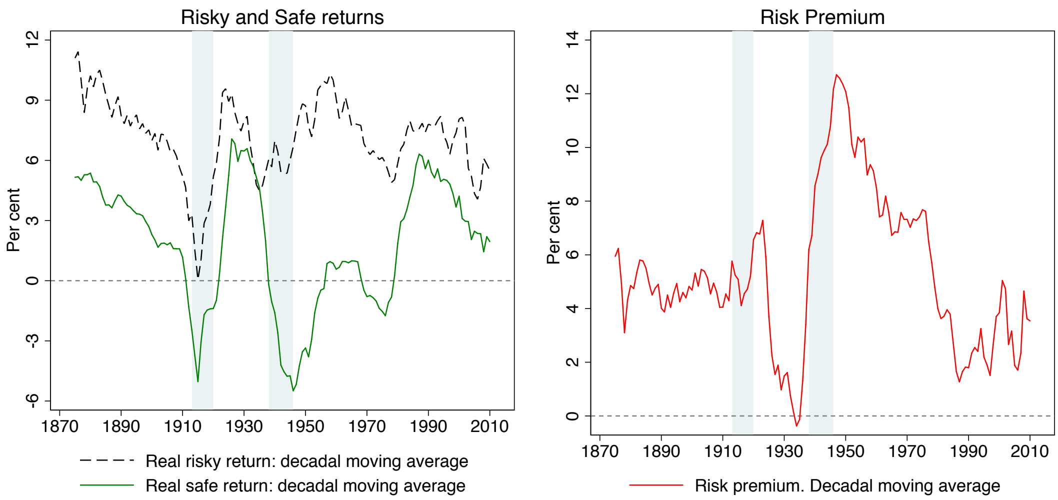
More recently, the Great Moderation saw a reduction in inflation rates and a corresponding increase in the debt financing burden, whereas the impact of $r^{safe} - g$ in the aftermath of the Global Financial Crisis remains broadly neutral, with the two rates roughly equal. On average throughout our sample, the real growth rate has been around 1 percentage point higher than the safe rate of return (3% growth versus 2% safe rate), meaning that governments could run small deficits without increasing the public debt burden.

In sum, real returns on safe assets have been quite low across the advanced countries over the last 150 years. In fact, for some countries, these returns have often been persistently negative. Periods of unexpected inflation, in war and peace, have often diluted returns, and flights to safety may have depressed safe returns even further in the more turbulent periods of global financial history. The low return for investors has, on the flipside, implied a low financing cost for governments, which was particularly important in reducing the debts incurred during WW2.

1.6 Risky versus safe returns

Having established the general trends in each risky and safe asset class, we now turn to examine broader patterns of returns across the different asset classes. We start by comparing returns on risky and safe assets. Figure 1.13 depicts the trends in global safe and risky asset returns, again using decadal moving averages of GDP-weighted global return series.

Figure 1.13. Global real risky vs. real safe return



Note: Mean returns for 16 countries, weighted by real GDP. Decadal moving averages. Within each country, the real risky return is a weighted average of equities and housing, and safe return - of bonds and bills. The within-country weights correspond to the shares of the respective asset in the country's wealth portfolio. Risk premium = risky return - safe return.

The risky return in each country is a weighted average of housing and equity returns, with weights corresponding to equity market capitalization and housing wealth in each respective country. The safe return is a simple unweighted average of bonds and bills.[45] The left panel of Figure 1.13 shows the risky and safe asset returns, and the right panel depicts the risk premium, calculated as the risky minus safe difference.

As in Sections 1.4 and 1.5, the data presented in this Section measure ex post returns and risk premiums, inclusive of capital gains. For some of the debates that we touch on, however, a forward-looking expected return measure would have been preferable. Realized returns are likely to fall below ex ante return expectations during periods of large negative shocks, such as the two world wars, and rise above them in times of high capital gains, such as that between 1980 and today. Long-run data on expected returns are, however, difficult to obtain. Our focus on long run trends, to an extent, allows us to look through some of the unexpected shocks that drive a wedge between ex ante and ex post returns. Nevertheless, the discussion in this section should be treated with a note of caution.

Both risky and safe returns were high during the 19th century but had been gradually declining in the run up to WW1, after which they declined sharply, as is to be expected. After the war, returns were recovering during the 1920s. From

45. For details on the construction of the weighted returns and the asset weights, see Section 1.2.3 and Appendix 1.A.15.3. Appendix 1.A.16 further compares the portfolio-weighted returns to equally-weighted returns, i.e., a simple average of housing and equity.

1930 onwards, the risky return stayed high and relatively stable, whereas the safe return dropped sharply and remained low until the late 1970s, before increasing and falling back again during the past three decades. These findings have implications for current debates around secular stagnation and the pricing, or mis-pricing, of risk.

Secular stagnation is associated with low rates of return, driven by an excess of savings or a general unwillingness to borrow and invest. These in turn reflect a variety of potential factors, including: (1) lower rates of productivity growth; (2) lower fertility and mortality rates; (3) a decline in the relative price of investment goods; (4) greater firm level market power; and (5) higher income inequality (Rachel and Smith, 2015; Thwaites, 2015; Eggertsson, Mehrotra, and Robbins, 2017).

Indeed, we can see that the safe return fell sharply during the 1930s, when Hansen (1939) originally proposed the secular stagnation hypothesis. That time also coincided with a demographic bust and was preceded by a big rise in income inequality in the run-up to the Great Depression. The safe return has been falling again since the mid-1980s as many have noted.[46] Understandably, this has led some observers to suggest that advanced economies are again in danger of entering secular stagnation, e.g., Summers (2014), and Eggertsson and Mehrotra (2014).

But the picture changes radically when we consider the trend in risky returns in addition to safe returns. Unlike safe returns, risky returns have remained high and broadly stable through the best part of the last 100 years, and show little sign of a secular decline. Turning back to the trend for safe assets, even though the safe return has declined recently, much as it did at the start of our sample, it remains close to its historical average. These two observations call into question whether secular stagnation is quite with us. The high and stable risky return coupled with falling safe rates could also be consistent with the notion of a "safety trap" brought about by the relative shortage of safe assets (Caballero and Farhi, 2017). However with risk premiums still not far off their historical averages, the evidence for a safety trap is thus far also not clear-cut.

We now turn to examine the long-run developments in the ex post risk premium, i.e., the spread between safe and risky returns (right panel of Figure 1.13). This spread was low and stable at around 5 percentage points before WW1. It rose slightly after WW1, before falling to an all-time low of near zero by around 1930. The decades following the onset of WW2 saw a dramatic widening in the risk premium, with the spread reaching its historical high of around 14 percentage points in the 1950s, before falling back to around its historical average.

Interestingly, the period of high risk premiums coincided with an era of few systemic banking crises. In fact, not a single such crisis occurred in our advanced-

46. Note that the safe interest rate—i.e. the component of the safe return that excludes capital gains, and is more relevant for the secular stagnation and safety trap debates, has also fallen sharply since 1980. However, like the bill rate in Figure 1.10, it remains close to its historical average.

Table 1.11. Real risky and safe asset returns across countries and time

Country	Full Sample		1950–1980		Post 1980	
	Risky return	Safe return	Risky return	Safe return	Risky return	Safe return
Australia	6.96	1.77	6.51	-1.34	7.71	4.54
Belgium	8.31	1.82	9.68	1.05	7.99	4.38
Denmark	8.02	3.05	8.57	0.49	6.84	4.97
Finland	10.87	2.16	13.47	1.28	13.06	4.18
France	6.54	0.54	12.33	-1.15	6.61	4.60
Germany	7.90	3.34	7.00	1.77	5.20	3.10
Italy	5.32	2.28	7.08	-0.83	5.21	4.14
Japan	6.79	1.29	10.86	0.05	4.81	3.00
Netherlands	7.30	1.31	10.26	-0.89	7.42	3.83
Norway	7.96	1.59	7.75	-2.34	10.65	3.56
Portugal	6.46	0.45	5.19	-3.30	7.41	3.45
Spain	5.39	0.68	7.27	-3.56	5.46	3.96
Sweden	8.52	2.35	8.67	-1.12	11.42	4.06
Switzerland	6.51	1.57	6.07	0.25	7.76	1.84
UK	6.35	1.51	8.33	-1.36	7.66	4.69
USA	7.12	1.92	6.44	-0.32	7.28	3.91
Average, unweighted	7.44	1.88	8.48	-0.81	7.65	3.89
Average, weighted	7.22	1.89	7.88	-0.56	6.66	3.81

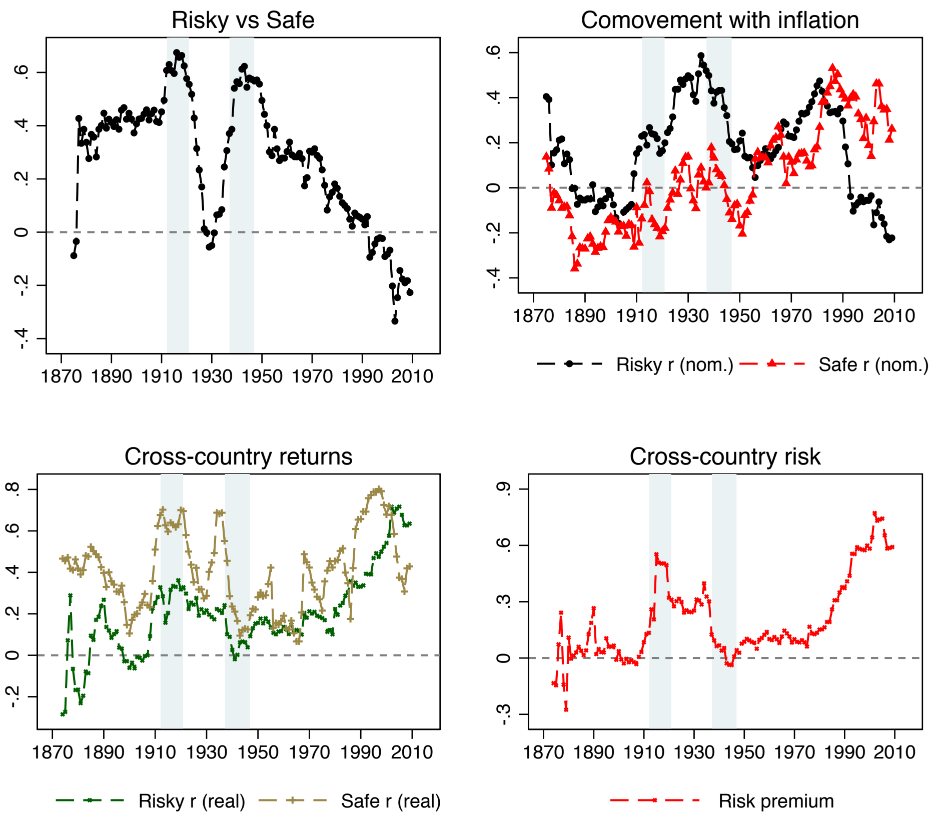
Note: Average annual real returns. Real risky return is a weighted average of equity and housing, and safe return - of bonds and bills. The weights correspond to the shares of the respective asset in the country's wealth portfolio. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both the risky and safe return. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

economy sample between 1946 and 1973. By contrast, banking crises appear to have been relatively more frequent when risk premiums were low. This finding speaks to the recent literature on the mispricing of risk around financial crises. Among others, Krishnamurthy and Muir (2017) argue that when risk is underpriced, i.e., risk premiums are excessively low, severe financial crises become more likely.

The long-run trends in risk premiums presented here seem to confirm this hypothesis. Appendix 1.A.6 further examines how these long-run movements in the risk premium, and the returns on the individual risky and safe asset classes, correspond to the changing monetary regimes, and finds, in accordance with Figure 1.13, that the risk premium during the Bretton-Woods fixed exchange rate era was unusually high by historical standards, driven largely by the low, even negative, safe asset returns, but also by reasonably high housing returns.

Table 1.11 zooms in to examine the evolution of safe and risky asset returns across different countries, as well as time periods. To enable a comparison with the aggregate trends in Figure 1.13, we split the post-WW2 period into two subperiods: 1950–1980, when global risk premiums were high and global safe returns low,

Figure 1.14. Correlations across risky and safe asset returns



Note: Rolling decadal correlations. The global correlation coefficient is the average of individual countries for the rolling window. Cross-country correlation coefficient is the average of all country pairs for a given asset class. Country coverage differs across time periods.

and post-1980, which saw an initial recovery, and subsequent decline in global safe returns.

The vast majority of countries in our sample follow similar patterns. The risky return is largely stable across time, even though it varies somewhat across countries: from just over 5% in Italy and Spain to 11% in Finland. Risk premiums were at or near their highest level in almost every country during the period 1950–1980, largely due to low returns on safe assets. The real safe rate of return was close to zero or negative for the majority of the countries in the sample, with the lowest level of –3.5% observed in Spain and Portugal, and only Belgium, Finland and Germany experiencing robustly positive real returns. Meanwhile, risky returns were also somewhat above their long-run level in a number of countries, but the differences are relatively smaller than those for safe rates. The post-1980 period saw a recovery in safe returns across the board, with the recent downward trend not yet apparent in these longer-run period averages. Risky returns, meanwhile, were close to their historical levels in most countries, with only Japan experiencing a strong decline following the bursting of its asset price bubble in the 1990s.

We now turn to examine the correlations between risky and safe returns, which are displayed in Figure 1.14. The top-left panel of this figure shows the rolling decadal correlation between the risky and safe returns, calculated as the average of rolling correlations in individual countries in a similar fashion to the calculations in Figure 1.8. Throughout most of the historical period under consideration, risky and safe returns had been positively correlated. In other words, safe assets have not generally provided a hedge against risk since safe returns were low when risky returns were low—in particular during both world wars—and vice versa. This positive correlation has weakened over the more recent decades, and turned negative from the 1990s onwards. This suggests that safe assets have acted as a better hedge for risk during both the Great Moderation and the recent Global Financial Crisis.

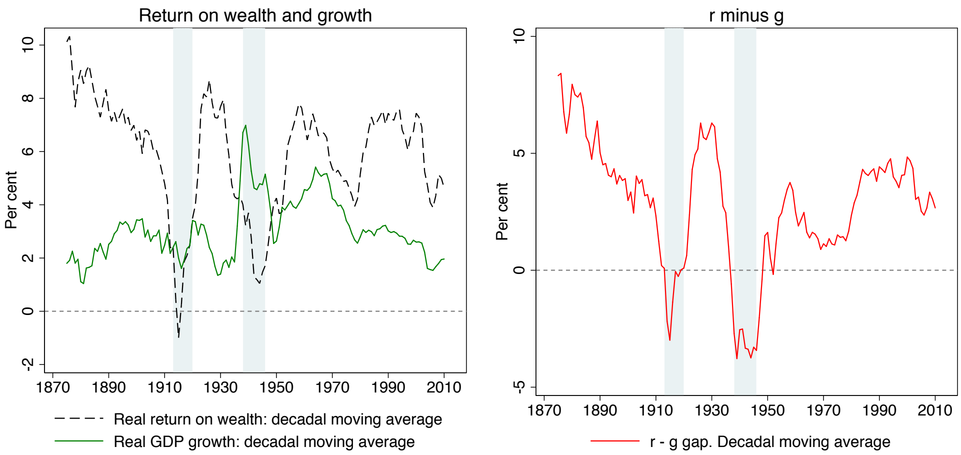
The top-right panel of Figure 1.14 shows the comovement of risky and safe nominal returns with inflation. Mirroring our findings presented in the preceding sections, safe returns have tended to comove more strongly with inflation, particularly during the post-WW2 period. Moving to cross-country correlations depicted in the bottom two panels of Figure 1.14, historically safe returns in different countries have been more correlated than risky returns. This has reversed over the past decades, however, as cross-country risky returns have become substantially more correlated. This seems to be mainly driven by a remarkable rise in the cross-country correlations in risk premiums, depicted in the bottom-right panel of Figure 1.14. This increase in global risk comovement may pose new challenges to the risk-bearing capacity of the global financial system, a trend consistent with other macro indicators of risk-sharing (Jordà, Schularick, and Taylor, 2017).

1.7 *r* versus *g*

Our analysis also provides insights into the debate on inequality. Piketty (2014) and Piketty and Zucman (2014) argue that inequality and wealth-to-income ratios in advanced economies have followed a U-shaped pattern over the past century and a half. They further hypothesize that wealth inequality may continue to rise in the future, along with a predicted decline in the rate of economic growth. The main theoretical argument for this comes about from a simple relation: $r > g$. In their approach, a higher spread between the real rate of return on wealth, denoted r, and the rate of real GDP growth, g, tends to magnify the steady-state level of wealth inequality. Benhabib and Bisin (2016) show that in a wide class of models featuring stochastic returns to wealth, a higher gap between r and g increases the Pareto index of the steady-state wealth distribution, making it more unequal.

Of course, this is not the only channel through which rates of return can impact the wealth distribution. Rate of return differentials between asset classes can affect the wealth distribution if there are systematic differences in the portfolio composi-

Figure 1.15. Real return on wealth and real GDP growth



Note: Mean returns and real GDP growth for 16 countries, weighted by real GDP. Decadal moving averages. Within each country, the real return on wealth is a weighted average of bonds, bills, equity and housing. The within-country weights correspond to the shares of the respective asset in each country's wealth portfolio.

tion between rich and poor households as Kuhn, Schularick, and Steins (2017) show, or if rates of returns vary with portfolio size as stressed by Piketty (2014). Studying administrative Swedish data, Bach, Calvet, and Sodini (2016) find that wealthy households earn higher returns on their portfolios, and Fagereng, Guiso, Malacrino, and Pistaferri (2016) use Norwegian tax data to document substantial heterogeneity in wealth returns. Rates of return on wealth are beginning to receive attention in the theoretical literature. For instance, Benhabib and Bisin (2016) point to return differences of assets as one potential channel to explain diverging trends between income and wealth inequality, and Garbinti, Goupille-Lebret, and Piketty (2017a) show that asset price effects played an important role in shaping the French wealth distribution over the past 200 years. Further, wealth inequality may depend not only on the magnitude of r in relation to g, but also on return volatility. Higher return volatility can increase the dispersion of wealth outcomes, and make the distribution of wealth more unequal.

To bring our data to bear on these debates, we construct a measure of the world's real return on wealth as a weighted average of real returns on bonds, equities, and housing—reflecting the typical portfolio of a private household end-investor. We then compare this measure to the rate of real GDP growth of economies over the long-run. Importantly, our approach differs from Piketty (2014) in that we rely on annual returns from observed market prices and yields for each individual asset class, rather than implicit returns derived from aggregate balance sheet data at selected benchmark dates. This, we think, is more robust and provides a vital cross check for the core argument.

Similarly to the risky returns in Section 1.6, we weight the individual returns by the size of the respective asset portfolio: stock market capitalization, housing wealth, and public debt (divided equally between bonds and bills).[47] Figure 1.15 displays the long-run trends in the global real rate of return on wealth (dashed line) and the global real GDP growth rate (solid line) since the late 19th century, again using decadal moving averages of GDP-weighted data.

Our data show that the trend long-run real rate of return on wealth has consistently been *much* higher than the real GDP growth rate. Over the past 150 years, the real return on wealth has substantially exceeded real GDP growth in 13 decades, and has only been below GDP growth in the two decades corresponding to the two world wars. That is, in peacetime, r has always exceeded g. The gap between r and g has been persistently large. Since 1870, the weighted average return on wealth (r) has been about 6.0%, compared to a weighted average real GDP growth rate (g) of 3.0%, with the average $r-g$ gap of 3.0 percentage points, which is about the same magnitude as the real GDP growth rate itself. The peacetime gap between r and g has been larger still, averaging around 3.8 percentage points.

Table 1.12 shows how the rate of return on wealth and the GDP growth rate have varied across different countries and time periods. Despite some variation, the positive gap between r and g is a persistent feature of the data: r is bigger than g in every country and every time period that we consider. The last few decades prior to the Global Financial Crisis saw a general widening of this gap, mirroring the aggregate pattern shown in Figure 1.15.

As previously discussed, returns on housing play an important part in this story—but with scant data until now, their exact role was unclear. The high level of housing returns that we have uncovered serves to push up the level of r, and thus, potentially, wealth inequality. But what is the counterfactual? We need to bear in mind that housing wealth is more equally distributed than, for instance, equities (see, e.g., Kuhn, Schularick, and Steins, 2017), and returns on housing are less volatile than those on shares—with both of these factors serving to flatten the distribution of wealth changes, making the overall impact of housing returns on wealth inequality less clear-cut and offering substantial scope for further research.

Rognlie (2015) notes that recent trends in wealth and income could be influenced primarily by what has happened in housing. Real house prices have experienced a dramatic increase in the past 40 years, coinciding with the rapid expansion of mortgage lending (Jordà, Schularick, and Taylor, 2015; Jordà, Schularick, and Taylor, 2016; Knoll, Schularick, and Steger, 2017). This is very much evident from Table 1.9. Measured as a ratio to GDP, rental income has been growing, as Rognlie

47. For details on the construction of the weighted returns and the asset weights, see Section 1.2.3 and Appendix 1.A.15.3. Appendix 1.A.16 further compares the portfolio-weighted returns to equally-weighted returns, with the equally-weighted return on wealth a simple average of equity, housing, and bonds.

Table 1.12. Return on wealth and GDP growth across countries and time

Country	Full Sample		Post 1950		Post 1980	
	Return on wealth	GDP growth	Return on wealth	GDP growth	Return on wealth	GDP growth
Australia	5.91	3.51	7.39	3.73	7.53	3.19
Belgium	6.38	2.32	7.29	2.68	6.90	2.17
Denmark	7.37	2.70	7.21	2.51	6.62	1.60
Finland	9.76	3.49	11.92	3.16	11.81	2.16
France	4.92	2.55	7.76	3.17	6.29	1.92
Germany	7.07	2.81	5.26	2.80	4.72	2.40
Italy	5.08	3.82	5.07	3.30	5.01	1.37
Japan	5.59	4.18	6.35	4.20	4.23	2.09
Netherlands	5.33	3.16	6.67	3.21	6.71	2.29
Norway	6.86	3.06	7.67	3.45	9.35	2.80
Portugal	5.87	3.39	5.65	3.48	6.99	2.13
Spain	4.58	3.21	5.50	4.03	5.34	2.56
Sweden	7.41	2.89	8.69	2.86	9.87	2.36
Switzerland	5.63	2.33	5.98	2.69	7.03	1.95
UK	4.75	2.09	5.90	2.49	7.23	2.45
USA	6.03	3.38	5.91	3.33	6.58	2.82
Average, unweighted	6.30	2.86	6.92	3.23	7.01	2.26
Average, weighted	5.98	3.04	6.09	3.33	6.08	2.48

Note: Average annual real returns. Real return on wealth is a weighted average of bonds, bills, equity and housing. The weights correspond to the shares of the respective asset in each country's wealth portfolio. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both the real return on wealth and the real GDP growth rate. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

(2015) argues. However, the rental yield has declined slightly—given the substantial increase in house prices—so that total returns on housing have remained pretty stable, as we have discussed. In this sense, recent housing trends have diverged little.

Our data allow us to more formally examine whether movements in the $r-g$ gap are more closely related to return fluctuations or movements in the real GDP growth rate. Appendix 1.A.17 and Table 1.A.26 document the correlations between $r-g$ and g, and $r-g$ and r across different time horizons, for the full sample and the period after 1950. Overall, the correlation between $r-g$ and g is negative, and somewhat stronger at longer horizons, with the correlation coefficients ranging between –0.2 and –0.6 depending on the historical subperiod and time window. At the same time, the $r-g$ gap is even more robustly related to changes in the return on wealth r, with a positive correlation between the two and a correlation coefficient of around 0.9, both over short and long run. This suggests that both falling GDP growth and higher returns would tend to increase the $r-g$ gap, although historically much of the changes in $r-g$ have come about from movements in the return on wealth. During peacetime r has been quite stable, and so has been the $r-g$ gap.

Since the 1970s, the stable and high levels of the rate of return on wealth have coincided with high and rising wealth-to-income ratios (see Piketty and Zucman, 2014, and Appendix Figure 1.A.7). Together, these two facts have meant that the capital share of GDP has increased across advanced economies (Karabarbounis and Neiman, 2014). A large part of these high returns, and of the increase in wealth ratios, can be attributed to high capital gains on risky assets, namely housing and equity. Rognlie (2015) argues that house prices have played an important role in the evolution of wealth-to-income ratios in the US. Kuvshinov and Zimmermann (2018) show that most of the recent increase in the value of listed firms in our cross-country sample is accounted for by higher equity valuations.

These high capital gains in recent decades have allowed the stock of measured wealth to rise without running into diminishing returns. Understanding the drivers behind these long-run trends in returns and valuations seems key to disentangling the underlying causes behind the recent upsurge in wealth, inequality, and the capital share of income.

1.8 Conclusion

In this paper we provide an investigation of the long history of advanced economy asset returns for all the major categories of the investible wealth portfolio. Our work brings new stylized facts to light and rigorously documents many broad patterns that have stimulated so much research in core areas of economics and finance over the last two centuries.

The returns to risky assets and risk premiums have been high and stable over the past 150 years. Substantial diversification opportunities exist between risky asset classes, and across countries. Arguably the most surprising result of our study is that long run returns on housing and equity look remarkably similar. Yet while returns are comparable, residential real estate is less volatile on a national level, opening up new and interesting risk premium puzzles.

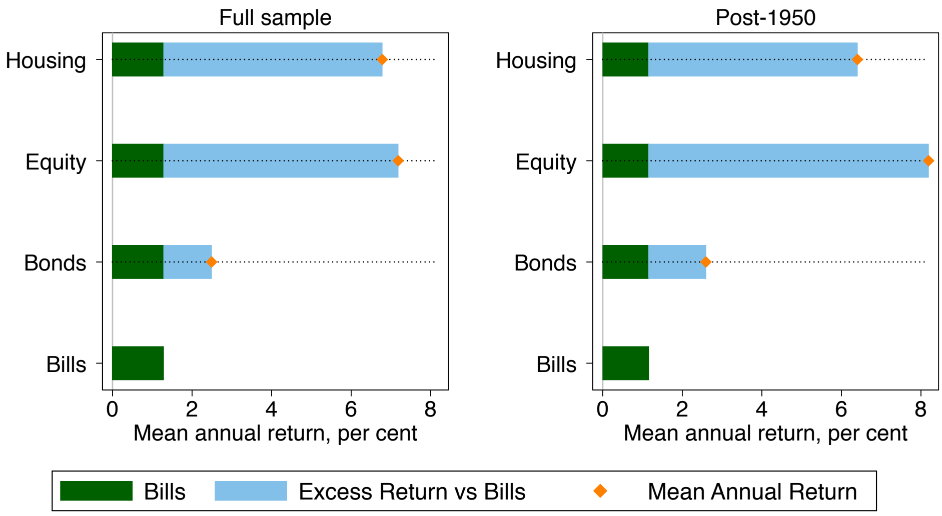
Our research speaks directly to the relationship between r, the rate of return on wealth, and g, the growth rate of the economy, that figures prominently in the current debate on inequality. One robust finding in this paper is that $r \gg g$: globally, and across most countries, the weighted rate of return on capital was twice as high as the growth rate in the past 150 years.

These and other discoveries can provide a rich agenda for future research, by us and by others. Many issues remain to be studied, among them determining the particular fundamentals that drive the returns on each of the asset classes in typical economies. For now, we hope our introduction of this new compilation of asset return data can provide the evidentiary basis for new lines of exploration in years to come.

Appendix 1.A Aggregate rates of return: Robustness checks

1.A.1 The effect of GDP weighting

Figure 1.A.1. GDP-weighted returns

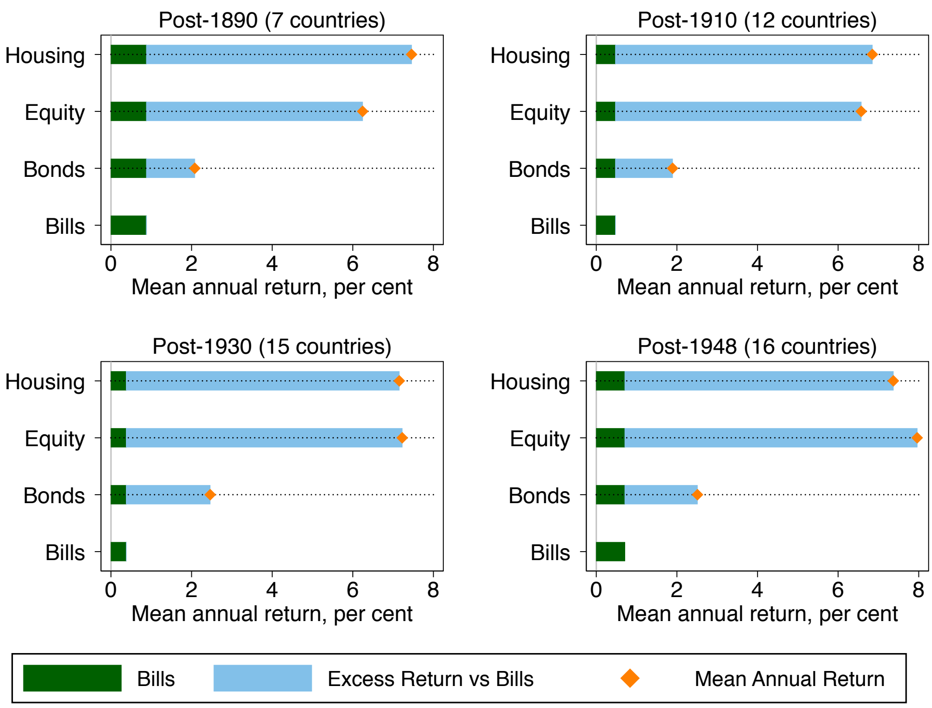


Notes: Arithmetic average real returns p.a., weighted by real GDP. Consistent coverage within each country: each country-year observation used to compute the average has data for all four asset returns.

This chart shows global average returns for the four asset classes weighted by country GDP, effectively giving greater weight to the largest economies in our sample, namely the U.S., Japan, and Germany. The overall effects are relatively minor. For the full sample, returns on equity and housing are similar at around 7% in real terms. For the post-1950 period, equities outperform housing by about 2pps. on average. The post-1990 housing bust in Japan and the underperformance of the German housing market contribute to this result.

1.A.2 More on sample consistency

Figure 1.A.2. Consistent samples



Note: Average real returns p.a. (unweighted). Consistent coverage across and within countries.

Throughout the paper, we always use a sample that is consistent within each table and graph, that is, for any table that shows returns on bills, bonds, equity, and housing, each yearly observation has data for all four asset returns. For tables showing bonds versus bills only, each yearly observation has data on both bonds and bills, but may be missing data for equities or housing. At the same time, returns for different countries generally cover different time periods.

Here we investigate whether adjusting for sample consistency affects our results. First, Figure 1.A.2 plots returns for samples that are consistent both within and across countries, starting at benchmark years. The later the benchmark year, the more countries we can include. The resulting return patterns confirm that the basic stylized facts reported earlier continue to hold even under these more stringent sampling restrictions, and regardless of the time period under consideration.

Next, we consider whether going to a fully "inconsistent" sample —that is, taking the longest time period available for each asset, without within-country consistency— would change the results. Table 1.A.1 thus shows returns for the maximum possible sample for each asset. Table 1.A.2, on the contrary, shows returns for a sample that is consistent within each country, across all four asset classes. The results in this table can be compared to Table 1.2 in the main text. On balance, the choice of the sample makes almost no difference to our headline results.

Table 1.A.1. Returns using longest possible sample for each asset

Country	Bills	Bonds	Equity	Housing
Australia	2.02	2.17	8.39	6.37
Belgium	1.68	3.01	5.89	7.89
Denmark	2.98	3.59	7.54	8.22
Finland	0.64	3.22	9.42	9.58
France	-0.47	0.84	3.21	6.39
Germany	1.50	3.12	8.83	7.82
Italy	1.20	2.11	6.09	4.77
Japan	0.63	2.54	9.64	6.54
Netherlands	1.37	2.71	6.96	7.22
Norway	1.10	2.55	5.67	8.33
Portugal	-0.01	2.76	4.05	6.31
Spain	0.70	1.34	5.77	5.21
Sweden	1.77	3.25	8.00	8.30
Switzerland	1.64	2.41	6.50	5.63
UK	1.16	2.29	6.86	5.44
USA	2.23	2.85	8.40	6.10
Average, unweighted	1.18	2.61	7.02	7.18
Average, weighted	1.34	2.48	7.40	6.69

Note: Average annual real returns. Longest possible sample used for each asset class, i.e. returns are not consistent across assets or within countries. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

Table 1.A.2. Returns using the full within-country-consistent sample

Country	Bills	Bonds	Equity	Housing
Australia	1.29	2.26	7.72	6.54
Belgium	0.77	2.87	6.78	8.64
Denmark	2.99	3.50	7.39	8.29
Finland	0.08	4.25	10.03	9.58
France	-0.47	1.54	3.99	7.14
Germany	2.65	4.03	7.11	7.82
Italy	1.37	3.19	7.25	4.77
Japan	0.39	2.18	6.00	6.54
Netherlands	0.78	1.85	6.96	7.28
Norway	0.90	2.29	5.67	8.03
Portugal	-0.48	1.37	4.51	6.31
Spain	-0.03	1.39	6.32	5.09
Sweden	1.56	3.14	8.02	8.30
Switzerland	0.81	2.33	6.69	5.77
UK	1.15	1.88	6.83	5.44
USA	1.52	2.33	8.46	6.10
Average, unweighted	1.21	2.62	6.70	7.33
Average, weighted	1.29	2.49	7.18	6.78

Note: Average annual real returns. Returns consistent within countries, i.e. each yearly observation for a country has data on each of the four asset classes. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

1.A.3 Returns during world wars

Table 1.A.3. Real returns on risky assets during world wars

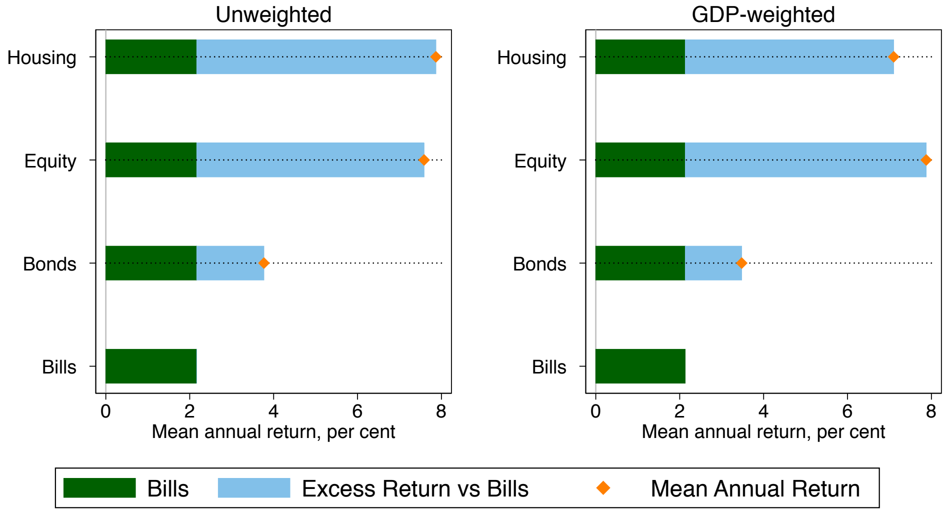
Country	World War 1		World War 2	
	Equity	Housing	Equity	Housing
Australia	0.20	1.22	4.86	4.12
Belgium	-3.75	-5.84	3.12	8.69
Denmark	6.70	4.35	2.85	11.75
Finland	4.68		0.55	-9.79
France	-12.48	-9.37	-4.05	-1.51
Germany	-12.37	-26.53	3.82	
Italy	-6.11			
Japan	15.88			
Netherlands	-0.20	5.07	5.71	9.10
Norway	-6.00	-1.38	0.62	2.54
Portugal	-3.99		3.96	
Spain	-5.77	-0.71	-0.73	-4.56
Sweden	-15.72	-3.93	5.56	7.89
Switzerland	-11.19	-4.46	1.32	3.08
UK	-6.67	-0.72	4.56	1.60
USA	0.96	0.06	4.90	8.47
Average, unweighted	-3.70	-1.84	2.65	3.85
Average, weighted	-3.69	-2.02	5.39	6.77

Note: Average annual real returns. We include one year from the immediate aftermath of the war, such that World war 1 covers years 1914—1919, and World War 2 – 1939—1946. Period coverage differs across and within countries. We exclude World War 2 periods for Italy and Japan because of hyperinflation. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

The performance of different assets during the major wars is an important issue for asset pricing models that argue that high risk premiums on equities reflect the risk of economy-wide disasters. This argument rests on the work of Barro (2006), developed further in collaboration with Emi Nakamura, John Steinsson and Jose Ursua (Barro and Ursua, 2008; Nakamura, Steinsson, Barro, and Ursúa, 2013). Table 1.A.3 shows the returns of housing and equity markets during WW1 and WW2. The data confirm large negative returns in different countries, especially during WW1. In both wars, housing markets tended to outperform equity, making it potentially more difficult to explain the large housing risk premium that we find. This being said, the positive returns in various countries during WW2 are in some cases influenced by price controls affecting our CPI measure and direct government interventions into asset markets that aimed at keeping prices up (see Le Bris, 2012, for the case of France). Further, as we do not adjust our return series for changes in the housing stock, the series here underestimate the negative impact of wartime destruction on housing investments. As a result, the war time returns shown here likely mark an upper bound, and wars can still be seen as periods with typically low returns on risky assets.

1.A.4 Returns excluding world wars

Figure 1.A.3. Returns excluding world wars, full sample



Note: Average real returns p.a., excluding world wars. Consistent coverage within each country.

In Figure 1.A.3 we exclude WW1 and WW2 from the calculation of aggregate returns, but maintain the within country consistency of the sample, as before. As expected, excluding the wars pushes up aggregate returns somewhat, but overall risk premiums and the relative performance of the different assets classes remain comparable.

Table 1.A.4. Real returns on bonds and bills, including and excluding world wars

Country	Full Sample		Excluding wars	
	Bills	Bonds	Bills	Bonds
Australia	1.29	2.24	1.73	2.65
Belgium	1.21	3.01	1.83	3.65
Denmark	3.08	3.58	3.80	4.39
Finland	0.64	3.22	2.17	5.34
France	-0.47	1.54	0.89	3.12
Germany	1.51	3.15	2.46	4.06
Italy	1.20	2.53	2.63	4.23
Japan	0.68	2.54	1.85	3.80
Netherlands	1.37	2.71	2.22	3.70
Norway	1.10	2.55	1.91	3.56
Portugal	-0.01	2.23	0.94	3.30
Spain	-0.04	1.41	1.17	2.73
Sweden	1.77	3.25	2.59	4.39
Switzerland	0.89	2.41	1.67	3.47
UK	1.16	2.29	2.03	3.22
USA	2.23	2.85	3.00	3.60
Average, unweighted	1.14	2.61	2.19	3.84
Average, weighted	1.34	2.51	2.27	3.52

Note: Average annual real returns. Returns excluding wars omit periods 1914—1919 and 1939—1947. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both real bill and bond returns. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

Table 1.A.4 displays country returns for bills and bonds including and excluding war periods. The effect on returns on bonds and bills, both weighted and unweighted, is substantial. The rate of return on bills almost doubles in real terms when the two war windows are excluded, and returns on bonds jump by about 1 percentage point.

Table 1.A.5. Real returns on equity and housing, including and excluding world wars

Country	Full Sample		Excluding wars	
	Equity	Housing	Equity	Housing
Australia	7.79	6.37	8.47	6.95
Belgium	6.23	7.89	7.47	8.73
Denmark	7.49	8.22	7.87	8.08
Finland	10.03	9.58	11.73	11.31
France	3.21	6.39	4.75	7.76
Germany	7.11	7.82	7.28	8.13
Italy	7.25	4.77	6.60	4.51
Japan	6.00	6.54	6.75	6.79
Netherlands	6.96	7.28	7.39	7.22
Norway	5.67	8.03	6.56	8.85
Portugal	4.51	6.31	4.51	6.31
Spain	5.83	5.21	6.92	6.41
Sweden	8.02	8.30	9.51	8.98
Switzerland	6.51	5.63	8.01	6.44
UK	6.83	5.44	7.82	5.69
USA	8.46	6.10	9.28	6.22
Average, unweighted	6.67	7.26	7.57	7.88
Average, weighted	7.12	6.72	7.86	7.10

Note: Average annual real returns. Returns excluding wars omit periods 1914—1919 and 1939—1947. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both real housing and equity returns. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

In Table 1.A.5 we look at the performance of risky assets for the full sample and excluding war periods. The effects are visible, but less strong than in the case of bonds and bills before. Excluding war years pushes up returns on equity and housing by 50 to 80 basis points. These effects are largely independent of the GDP-weighting.

Table 1.A.6. Real risky and safe asset returns, including and excluding world wars

Country	Full Sample		Excluding wars	
	Risky return	Safe return	Risky return	Safe return
Australia	6.96	1.77	7.46	2.20
Belgium	8.31	1.82	8.53	2.61
Denmark	8.02	3.05	7.88	3.82
Finland	10.87	2.16	12.68	3.55
France	6.54	0.54	7.39	2.01
Germany	7.90	3.34	8.19	3.36
Italy	5.32	2.28	5.01	2.94
Japan	6.79	1.29	7.11	2.08
Netherlands	7.30	1.31	7.36	2.39
Norway	7.96	1.59	8.83	2.55
Portugal	6.46	0.45	6.46	0.45
Spain	5.39	0.68	6.26	1.96
Sweden	8.52	2.35	9.51	3.41
Switzerland	6.51	1.57	7.37	2.50
UK	6.35	1.51	6.91	2.39
USA	7.12	1.92	7.45	2.73
Average, unweighted	7.44	1.88	8.09	2.93
Average, weighted	7.22	1.89	7.67	2.80

Note: Average annual real returns. Returns excluding wars omit periods 1914—1919 and 1939—1947. Real risky return is a weighted average of equity and housing, and safe return - of bonds and bills. The weights correspond to the shares of the respective asset in the country's wealth portfolio. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both the risky and safe return. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

Table 1.A.6 underlines the outperformance of risky assets once we exclude the wars. Average safe returns are about 1 percentage point lower in the full sample, relative to the sample that exclude war years. By contrast, risky returns only rise by between 40 and 60 basis points when we exclude wars. As discussed above the measurement of returns in wars is problematic and we are inclined not to read too much into the relative outperformance of risky assets in war times.

Table 1.A.7. Return on capital and GDP growth, including and excluding world wars

Country	Full Sample		Excluding wars	
	Return on wealth	GDP growth	Return on wealth	GDP growth
Australia	5.91	3.51	6.48	3.64
Belgium	6.38	2.32	6.77	2.51
Denmark	7.37	2.70	7.32	2.75
Finland	9.76	3.49	11.63	3.63
France	4.92	2.55	6.03	2.75
Germany	7.07	2.81	7.30	2.98
Italy	5.08	3.82	4.94	3.23
Japan	5.59	4.18	6.31	4.31
Netherlands	5.33	3.16	5.86	3.16
Norway	6.86	3.06	7.71	3.13
Portugal	5.87	3.39	5.87	3.39
Spain	4.58	3.21	5.69	3.44
Sweden	7.41	2.89	8.45	2.97
Switzerland	5.63	2.33	6.56	2.55
UK	4.75	2.09	5.50	2.23
USA	6.03	3.38	6.63	3.19
Average, unweighted	6.30	2.86	7.12	2.93
Average, weighted	5.98	3.04	6.69	2.97

Note: Average annual real returns. Returns excluding wars omit periods 1914—1919 and 1939—1947. Real return on wealth is a weighted average of bonds, bills, equity and housing. The weights correspond to the shares of the respective asset in each country's wealth portfolio. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both the real return on wealth and the real GDP growth rate. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

Table 1.A.7 looks at the effects of war periods on the aggregate return on capital and GDP growth on a country level and for the global sample. The aggregate return on capital is about 75 basis points higher outside world wars, while GDP growth rates are barely affected as the war effort boosted GDP in many countries in the short term.

1.A.5 Returns before the Global Financial Crisis

Table 1.A.8. Asset returns before the Global Financial Crisis

Country	Bills	Bonds	Equity	Housing
Australia	1.30	1.95	8.20	6.49
Belgium	1.35	2.86	6.07	8.22
Denmark	3.31	3.56	7.18	8.73
Finland	0.76	3.10	10.65	9.96
France	-0.46	1.17	3.20	6.69
Germany	1.64	3.13	7.21	7.80
Italy	1.30	2.24	8.16	5.32
Japan	0.74	2.51	6.25	6.88
Netherlands	1.48	2.50	7.19	7.77
Norway	1.14	2.41	5.63	8.14
Portugal	-0.00	1.64	5.50	7.19
Spain	0.01	0.95	6.21	5.89
Sweden	1.86	3.09	7.81	8.32
Switzerland	0.99	2.17	6.59	5.40
UK	1.32	2.16	7.11	5.74
USA	2.43	2.71	8.55	6.29
Average, unweighted	1.24	2.42	6.79	7.50
Average, weighted	1.46	2.36	7.23	6.93

Note: Average annual real returns excluding the Global Financial Crisis (i.e. sample ends in 2007). Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for returns on all four asset classes. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

This Table cuts the sample off in 2007, i.e., before the Global Financial Crisis. Comparing this table to Tables 4 and 5 in the main text shows that the effects are relatively minor. The crisis only shaves off about 10-20 basis points from equity and housing returns, and adds about 10 basis points to bills and bonds.

Table 1.A.9. Risky and safe returns, including and excluding the GFC

Country	Full Sample		Excluding the GFC	
	Risky return	Safe return	Risky return	Safe return
Australia	6.96	1.77	7.16	1.63
Belgium	8.31	1.82	8.58	1.80
Denmark	8.02	3.05	8.28	3.15
Finland	10.87	2.16	11.38	2.19
France	6.54	0.54	6.82	0.36
Germany	7.90	3.34	7.90	3.49
Italy	5.32	2.28	5.93	2.18
Japan	6.79	1.29	7.03	1.28
Netherlands	7.30	1.31	7.70	1.19
Norway	7.96	1.59	8.04	1.52
Portugal	6.46	0.45	7.36	-0.26
Spain	5.39	0.68	6.06	0.47
Sweden	8.52	2.35	8.44	2.30
Switzerland	6.51	1.57	6.44	1.49
UK	6.35	1.51	6.66	1.52
USA	7.12	1.92	7.22	1.91
Average, unweighted	7.44	1.88	7.65	1.84
Average, weighted	7.22	1.89	7.38	1.87

Note: Average annual real returns excluding the Global Financial Crisis (i.e. sample ends in 2007). Real risky return is a weighted average of equity and housing, and safe return - of bonds and bills. The weights correspond to the shares of the respective asset in the country's wealth portfolio. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both the real risky and the real safe return. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

This Table recalculates risky and safe returns including and excluding the Global Financial Crisis on a country level and for the global average. As noted before, the effects are quantitatively small. Excluding the crisis boosts risky returns by 10-20 basis, and lower safe returns by no more than 5 basis points. In light of the long time horizon of nearly 150 years, asset performance in the recent crisis plays a minor role for the returns presented here.

1.A.6 Returns across different monetary regimes

Table 1.A.10. Returns across different monetary regimes

	Bills	Bonds	Equity	Housing	Risk premium
Gold standard (1870–1913)					
Mean return p.a.	3.06	2.81	5.96	7.49	4.25
Standard deviation	3.22	5.00	8.65	8.87	7.29
Geometric mean	3.01	2.69	5.60	7.14	4.01
Observations	305	305	305	305	301
Amended gold standard (1919–1938)					
Mean return p.a.	4.25	7.11	7.63	7.86	2.39
Standard deviation	7.77	13.58	21.58	12.04	13.53
Geometric mean	3.96	6.29	5.52	7.25	1.51
Observations	264	264	264	264	264
Bretton-Woods (1946–1973)					
Mean return p.a.	-0.82	-1.01	6.43	8.98	9.70
Standard deviation	5.59	8.85	21.00	11.17	10.00
Geometric mean	-1.00	-1.44	4.44	8.45	9.29
Observations	403	403	403	403	403
Floating exchange rates (1974–2015)					
Mean return p.a.	1.38	4.34	8.56	6.30	4.42
Standard deviation	3.39	10.82	26.29	7.61	11.49
Geometric mean	1.32	3.79	5.27	6.03	3.84
Observations	670	670	670	670	670

Note: Annual global returns and risk premiums in 16 countries, equally weighted. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for all four asset returns. Risk premium is the risky return (weighted average of equities and housing) minus the safe return (weighted average of bonds and bills).

Table 1.A.10 examines how the returns on the four asset classes in this paper, and the risk premium, vary across monetary regimes. We roughly divide the time period in this study into the fixed exchange rates under the pre WW1 gold standard, the amended interwar gold standard, the Bretton-Woods era, and the post-Bretton-Woods floating exchange rates. Consistent with Figures 1.7 and 1.13, the returns on the risky assets have been stable across all four regimes. The volatility of equity returns has generally increased overtime, starting at levels comparable to housing under the Gold Standard, before reaching a peak of 26% under floating exchange rates. Housing returns were highest during the Bretton-Woods era (average 9% p.a.), whereas equity returns—during the recent floating exchange rate period (average

8.6% p.a.). Real bond and bill returns have, on the contrary, been much more variable across these different regimes, mirroring the pattern discussed in Section 1.6: they were high during the gold standard, both pre WW1 and interwar, and low—in fact, negative—during Bretton-Woods. The risk premium has also varied across these four regimes, with the swings largely driven by those in the safe rate. As discussed in Section 1.6, the risk premium was highest during the Bretton-Woods era, which also saw remarkably few financial crises.

1.A.7 Returns during disasters

Table 1.A.11. Returns during consumption disasters

Country	Consumption			Real returns			
	Peak	Trough	Drop	Equity	Housing	Bonds	Bills
Australia	1913	1918	23.8	3.69	12.43	-4.62	-5.92
	1927	1932	23.4	14.40	-0.28	18.83	35.88
	1938	1944	30.1	5.70	35.89	0.49	-8.45
Belgium	1913	1917	44.5	-54.74	-33.05		
	1937	1942	53.0	-15.22	-35.88	-61.72	-66.44
Denmark	1919	1921	24.1	-37.32	-5.60	-37.63	-19.08
	1939	1941	26.1	-13.93	-3.64	-16.06	-14.06
	1946	1948	14.4	9.08	51.83	1.96	4.93
Finland	1913	1918	36.0	-7.43		-72.72	-60.10
	1928	1932	19.9	-6.10	17.14	49.87	49.00
	1938	1944	25.4	12.13	-45.62	-32.83	-38.39
	1989	1993	14.0	-57.92	-17.40	32.92	30.18
France	1912	1915	21.5	-7.11	16.14	-19.25	7.93
	1938	1943	58.0	110.36	-11.38	-19.55	-45.70
Germany	1912	1918	42.5	-39.37		-52.74	-50.38
	1928	1932	12.1	-54.91	15.25	12.24	45.08
	1939	1945	41.2	79.03			1.91
Netherlands	1889	1893	9.8		21.90	21.46	12.20
	1912	1918	44.0	10.55	27.55	-25.95	-15.59
	1939	1944	54.5	48.15	87.58	-19.97	-26.12
Norway	1916	1918	16.9	9.76	-1.40	-26.97	-26.54
	1919	1921	16.1	-43.18	4.65	-28.47	-11.32
	1939	1944	10.0	27.47	-2.47	-16.72	-25.84
Portugal	1913	1919	21.5	-21.90		-46.06	-53.88
	1934	1936	12.1	33.56		35.30	8.91
	1939	1942	10.4	5.76		-4.54	0.99
	1974	1976	9.8	-73.88	-32.62	-53.52	-28.47
Spain	1913	1915	12.8	-4.97	8.23	-0.18	8.35
	1929	1930	10.1	-0.24	6.10	-1.51	0.68
	1935	1937	46.1	42.88	-5.45	3.65	
	1940	1945	14.5	-23.44	-16.64		-42.18
	1946	1949	13.1	-9.90	-21.44	-32.36	-28.93
Sweden	1913	1917	11.5	-8.83	5.83	-22.59	-9.90
	1920	1921	13.2	-20.13	14.01	-12.25	3.46
	1939	1945	18.2	16.00	30.53	-34.64	-18.16

Continued overleaf

Returns during consumption disasters, continued

Country	Consumption			Real returns			
	Peak	Trough	Drop	Equity	Housing	Bonds	Bills
Switzerland	1912	1918	10.8	-34.04	-28.62	-28.50	-27.71
	1939	1945	17.3	-15.15	-9.19	-20.37	-23.49
UK	1915	1918	16.7	-34.12	8.93	-48.94	-31.30
	1938	1943	16.9	-12.27		-10.14	-27.42
USA	1917	1921	16.4	-49.07	-11.88	-44.32	-36.24
	1929	1933	20.8	-51.64	-5.58	50.71	46.35
Averages:							
All disasters			23.3	-6.71	2.23	-14.94	-12.46
Consistent sample			22.0	-8.94	3.59	-14.74	-9.30
No WW2			21.2	-19.68	0.74	-12.05	-5.35

Note: Consumption and cumulative real total returns on each asset class during consumptions disasters. Disaster dates from Barro and Ursúa (2008). Cumulative consumption drop from peak to trough year, and cumulative real returns from one year before consumption peak to one year before the trough. Negative return means an asset return drop during disaster. Disasters with missing or poor quality asset return data excluded. All-disaster average uses all disasters where we have data for the particular asset class. Consistent sample average uses only those disasters with data for all four assets. No WW2 average excludes World War 2.

Table 1.A.11 summarises returns on the four asset classes during consumption disasters. Disaster dates and consumption data are taken from Barro and Ursúa (2008). A disaster is defined as a real consumption drop of 10% or more. Asset returns are from this paper, and include both the capital gain and yield components, net of inflation. As in Barro and Ursúa (2008), the return figures are from the year before consumption peak to the year before the trough, and do not typically correspond to peak-to-trough asset price declines. We exclude those disasters where asset data are missing, or unreliable—for example, due to a hyperinflation—from the comparison. Despite these, the data for remaining disasters are imperfect, largely due to various potential biases discussed in Sections 1.3.2 and 1.3.3. There is likely to be some upward bias to returns because equity returns do not fully capture the effect of delistings, and housing returns do not fully capture the effect of wartime housing stock destruction. Some returns may also be subject to market intervention and price controls, especially during WW2. Still, we have strived to improve the data quality and coverage of asset return data during disasters, and hope that the above comparison provides a best-practice assessment of asset market performance during these time periods.

The upper part of Table 1.A.11 lists the cumulative returns on each of the four asset classes during individual disaster episodes. Most disasters occur during the two world wars and the Great Depression. The several additional country-specific events include the Portuguese Carnation revolution (1974) and the Spanish Civil War (1936–1940). For most disasters, returns on all assets are low and typically negative. The disaster that had the most pronounced impact on asset markets is the Portuguese Carnation Revolution of 1974. In the aftermath of the revolution, equity returns fell by three-quarters, housing returns by one-third, and bill and bond returns by between one-third and one-half.

Some disasters saw postive returns, in particular for the risky asset classes. This is especially true for a number of countries during the WW2 German occupation. Equity returns in France, Netherlands and Germany during WW2 rose by between 50 and 100 per cent in real terms. Much of this is likely due to financial repression by the occupying government, aimed at directing funds to the stock market in order to finance the war effort (see Le Bris, 2012, for the case of France). As discussed in Section 1.3.2, timing also played a role: for a while, the government was able to intervene in markets and sustain high returns, but asset prices usually underwent a correction at some point. Taking the French case again, the stock market fell by a factor of 18 (95%) between years 1943 and 1950, more than reversing the gains made during 1938–1943. Some of the high housing returns observed during WW2 could be driven by price and rent controls, and the fact that our data do not account for wartime destruction. To this purpose, when comparing averages at the bottom of Table 1.A.11, we have included a comparison for disasters outside of the Second World War.

The disasters with the worst asset market performance were not necessarily those with the highest consumption drops. For example, the Portuguese revolution only saw a 10% consumption drop, and is excluded from some measures of consumption disastes such as that in Nakamura, Steinsson, Barro, and Ursúa (2013). The years leading up to the Spanish Civil War, on the contrary, saw a 46% consumption drop while the stock market boomed.

On average, consumption disasters are associated with low and negative asset returns (Table 1.A.11, bottom panel). Safe assets do the worst, largely because they provide a poor hedge for inflation risk. Housing returns, on average, seem more robust to disasters, recording a low but positive real return (+2.2% cumulative), even excluding the period of WW2 that saw widespread housing market controls and war destruction (+0.7% cumulative real return excluding WW2). Equities do much worse in disasters outside of WW2, with an average cumulative real return of -20%, compared to -6.7% return in all disasters. This is consistent with the notion that stock markets were artificially supported through regulations and financial repression in some countries during the Second World War.

One type of disaster event that we exclude from our main sample are hyperinflations. We do this because of very large measurement error during these time periods: not only are price changes and capital gains on assets measured imprecisely, but timing matters a lot: for example, if the price change is measured from June one year to June the next, but asset prices are measured in December each year, and the inflation rate is increasing sharply, real returns would be vastly overstated. Nevertheless, our data do allow us to make some tentative estimates on the size of returns

Table 1.A.12. Returns during hyperinflations

Episode	Real returns			Nominal returns			Inflation
	Bills	Bonds	Equity	Bills	Bonds	Equity	
Finland, 1917–1918	-69.54	-60.67	-60.76	4.00	34.29	33.96	241.41
Germany, 1920–1924	-100.00	-100.00	-29.03	36.95	90.37	2.25e+13	3.17e+13
Italy, 1943–1944	-76.35	-70.08	-43.40	5.10	32.96	151.53	344.39
Japan, 1944–1947	-97.66	-97.59		8.30	11.58		4522.09

Note: Cumulative real and nominal returns during hyperinflations. No housing returns data are available for these episodes. Because of potential timing differences and uncertainties about inflation rates, the returns are potentially subject to a large measurement error.

during these time periods. These are presented in Table 1.A.12. We do not have any housing return data for hyperinflation periods, but we have equity, bond and bill return data for three episodes (Finland post WW1, Germany 1920s and Italy WW2), and bond and bill return data for the post WW2 hyperinflation in Japan.

One can see that the nominally fixed assets—bonds and bills—fare worst during these time periods. Domestic government debt holders were more or less wiped out during the German hyperinflation, and lost most of their capital during the Japanese post-WW2 episode. Equity holders fare somewhat better, but still suffer losses. During the German hyperinflation, for example, equities generally kept pace with the explosive growth in consumer prices. Including hyperinflations in our sample would have a very small effect on the aggregate average equity returns because of the infrequency of these episodes, and the fact that equity holders are not generally wiped out during hyperinflations. Cross-country arithmetic average returns on bonds and bills would also change little, but the geometric mean return on these two assets would fall to near zero, reflecting the highly negative real returns on these assets during the German hyperinflation.

1.A.8 Returns without interpolation

Table 1.A.13. Returns with and without interpolation

	Baseline		No interpolation	
	Equity	Housing	Equity	Housing
Mean return p.a.	6.73	6.93	6.86	6.99
Standard deviation	21.91	10.31	21.80	10.15
Observations	1790	1790	1783	1782

Note: Equity and housing returns with (baseline) and without interpolation. Interpolation only used to cover the following disaster periods. Equity: Spanish Civil War, Portuguese Carnation Revolution. Housing: Belgium WW1, Sweden WW1, Spanish Civil War. We only interpolate either house prices or rents, never both. 16 countries, unweighted.

For a very small number of observations, we interpolate our equity and housing return series to obtain a proxy of the annual return during disasters, for example when the stock exchange was closed or when only either annual rental or house price data exist, but not both. These cover selected periods in WW1, the Spanish Civil War and the Portuguese Carnation Revolution, with a total of 7 observations for equities and 8 for housing. Table 1.A.13 compares our baseline interpolated returns with those without interpolation. The returns and their volatility (annual standard deviation) are almost identical, albeit interpolated returns are slightly lower because they better proxy the returns during disaster periods.

1.A.9 US Dollar returns

Table 1.A.14. Global real returns for a US-Dollar investor

	Real returns				Nominal Returns			
	Bills	Bonds	Equity	Housing	Bills	Bonds	Equity	Housing
Full sample:								
Mean return p.a.	1.97	3.52	7.89	8.17	4.41	5.93	10.45	10.84
Standard deviation	12.05	15.49	24.98	15.72	11.63	14.81	25.21	16.09
Geometric mean	1.18	2.33	5.02	6.99	3.69	4.86	7.62	9.64
Mean excess return p.a.	0.21	1.76	6.13	6.41				
Standard deviation	11.23	14.63	24.63	15.73				
Geometric mean	-0.47	0.71	3.32	5.23				
Observations	1767	1767	1767	1767	1767	1767	1767	1767
Post-1950:								
Mean return p.a.	2.26	4.12	9.62	9.02	5.72	7.60	13.22	12.70
Standard deviation	10.59	13.75	26.32	14.89	10.95	13.88	26.81	15.34
Geometric mean	1.72	3.23	6.50	8.04	5.17	6.73	10.11	11.70
Mean excess return p.a.	0.80	2.66	8.16	7.56				
Standard deviation	10.58	13.83	26.11	15.02				
Geometric mean	0.25	1.75	5.06	6.54				
Observations	1022	1022	1022	1022	1022	1022	1022	1022

Note: Global average US-Dollar returns, equally weighted. Real returns subtract US inflation. Excess returns are over US Treasury bills. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for all four asset returns.

Table 1.A.14 shows nominal and real returns from the perspective of a U.S. dollar (USD) investor. The Table can be directly compared to Table 1.2 in the paper. Overall, calculating returns in dollars increases their volatility, since returns now also fluctuate with nominal exchange rate movements. It also adds up to 1 percentage point to the local currency returns reported in Table 1.2. The higher average return is, for the most part, driven by the higher volatility—exchange rate movements amplify both positive and negative returns, but because returns are on average positive, the average return increases. The effects are stronger after WW2, going hand-in-hand with the greater exchange rate volatility after the collapse of the Bretton Woods system.

Table 1.A.15. USD returns by country

Country	Bills	Bonds	Equity	Housing
Australia	1.77	2.59	8.53	7.28
Belgium	0.96	3.27	7.36	8.92
Denmark	3.63	4.12	7.94	8.93
Finland	1.93	6.48	12.08	12.00
France	1.05	3.10	5.16	8.87
Germany	4.33	5.81	8.75	9.69
Italy	2.85	4.81	8.73	6.37
Japan	2.37	4.15	8.00	8.74
Netherlands	1.87	2.93	7.91	8.67
Norway	1.64	3.04	6.72	8.88
Portugal	0.23	2.11	6.09	7.10
Spain	0.93	2.36	7.47	6.38
Sweden	2.08	3.64	8.63	8.87
Switzerland	2.04	3.62	7.62	7.14
UK	1.89	2.66	7.66	6.25
USA	1.52	2.33	8.46	6.10
Average, unweighted	2.08	3.55	7.67	8.36
Average, weighted	2.03	3.28	7.95	7.63

Note: Average annual real US-Dollar returns. Calculated as nominal US-Dollar return minus US inflation. Period coverage differs across countries. Consistent coverage within countries. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

In Table 1.A.15 we display Dollar returns for individual asset classes and individual countries for the full sample. For US-Dollar based fixed income investors, Germany and Finland offered the highest returns. In housing markets, Germany and Finland again stand out, and high returns are seen in Belgium, France, Netherlands and the Scandinavian countries. In equity markets, Finland, Italy, and Sweden were the best performing markets.

1.A.10 Risky returns ranked by country

Table 1.A.16. Risky returns ranked by country

Country	Full sample	Post-1950	Post-1980
Finland	10.87	13.10	13.06
Sweden	8.52	10.23	11.42
Belgium	7.60	8.72	7.99
Denmark	8.01	7.85	6.84
Norway	7.96	9.32	10.65
Germany	7.90	5.83	5.20
Average, unweighted	7.44	8.10	7.65
Netherlands	7.30	8.77	7.42
USA	7.12	7.03	7.28
Australia	6.96	8.43	7.71
Japan	6.79	7.05	4.81
France	6.54	9.09	6.61
Switzerland	6.51	7.04	7.76
Portugal	6.46	6.21	7.41
UK	6.35	7.85	7.66
Spain	5.39	6.17	5.46
Italy	5.32	5.85	5.21

Note: Average annual real risky returns. Real risky return is a weighted average of equity and housing. The weights correspond to the shares of the respective asset in the country's wealth portfolio. Period coverage differs across countries. The figure is the unweighted arithmetic average of individual country returns.

In Table 1.A.16 we rank risky returns in the different countries. We calculate risky returns as a combination of equity and housing weighted by the share of each asset in the country's total wealth portfolio. North-western Europe—essentially the Scandinavian countries plus Germany and Belgium—stands out as the region with the highest aggregate returns on risky assets. The U.S. returns are about average, while the southern European countries have comparatively low long-run returns.

1.A.11 Risky returns over longer time horizons

Table 1.A.17. Equity and housing return moments over longer time horizons

	1-year (baseline)		5-year		10-year		20-year	
	Equity	Housing	Equity	Housing	Equity	Housing	Equity	Housing
Mean return p.a.	6.88	7.06	7.11	7.00	7.11	7.16	7.24	7.10
Standard deviation	21.79	9.93	9.12	5.39	7.42	4.02	6.41	2.75
Observations	1767	1767	337	337	152	152	58	58

Note: Average global real returns in 16 countries averaged over 1-year, 5-year, 10-year and 20-year non-overlapping windows.

Table 1.A.17 computes average equity and housing returns over longer time horizons. Because of high transaction costs, housing returns are likely to exhibit much more sluggish responses to shocks than equity returns, and display larger volatility or variability at medium to long horizons. Columns 1 and 2 present the baseline 1-year returns compared to average returns over 5, 10 and 20-year windows. The volatility of both equity and housing returns falls, the longer the smoothing window. But the two fall at the same rate, and the standard deviation of housing returns is always around half of equity returns, regardless of the time horizon.

1.A.12 Equity return comparison with existing sources

Table 1.A.18. Real equity returns in our dataset compared to existing literature

	Full sample	Post-1900	Dimson et al.	Barro & Ursua
Australia	8.39	7.81	8.93	10.27
Belgium	5.89	5.65	5.32	
Denmark	7.54	7.49	7.10	7.50
Finland	9.42	9.72	10.23	12.68
France	3.21	2.27	5.83	5.43
Germany	8.83	9.11	9.18	7.58
Italy	6.09	5.68	6.35	5.10
Japan	9.64	8.60	9.50	9.28
Netherlands	6.96	6.96	6.92	9.01
Norway	5.67	5.73	7.20	7.16
Portugal	4.05	3.36	8.66	
Spain	5.77	5.77	5.77	6.10
Sweden	8.00	7.94	7.87	9.23
Switzerland	6.50	6.50	6.19	7.26
UK	6.86	6.74	7.06	6.41
USA	8.40	8.43	8.59	8.27
Average, unweighted	7.02	6.69	7.50	7.95
Average, weighted	7.40	7.18	7.71	7.83

Note: Average annual real equity returns. Period coverage differs across countries. Consistent coverage within countries. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns. Full sample estimates cover our entire dataset, post-1900 estimates are our data restricted to years 1900 and after. Dimson et al. data are from the Dimson-Marsh-Staunton dataset, which covers the period from 1900 onwards. Barro & Ursua are data from Barro and Ursúa (2008), some series starting in 1870 and others in later years. Barro & Ursua weighted average is weighted by 2000 real GDP.

Table 1.A.18 compares our equity return estimates with those in the existing literature, using the updated dataset in Dimson, Marsh, and Staunton (2009) (DMS), and the data in Barro and Ursúa (2008) (BU), with the latter sourced from *Global Financial Data*. Our equity return estimates are similar to those of DMS and BU, but generally slightly lower, with the largest differences in the returns for France and Portugal. The lower returns for France come from a higher-quality value weighted index in Le Bris and Hautcoeur (2010), which eliminates a number of accuracy problems in older indices, mainly related to selection bias and weighting, and hence results in lower returns than the indices used in DMS and BU.[48] The low returns for Portugal are partly attributable to better coverage of the market crash after the Carnation Revolution of 1974 (see Section 1.3.2 and Table 1.3). The volatility of equity returns is similar across the different sources. The standard deviation of equity returns in

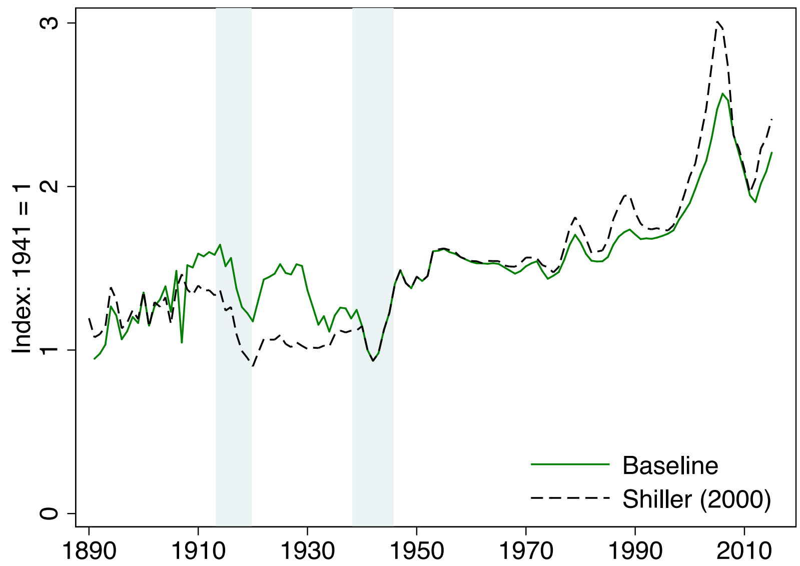
48. See Le Bris and Hautcoeur (2010) for a further discussion of these issues.

our dataset is 21.8 percentage points p.a., compared to 24.4 in DMS and 23.1 in BU.[49]

49. The BU standard deviation figure is an average of country-specific standard deviations; whereas our and DMS standard deviation is computed across the whole cross-country portfolio.

1.A.13 Housing return comparison to Shiller

Figure 1.A.4. Real house price index in this paper compared to Shiller



Note: Real house price indices, normalised to equal 1 in 1941. Our house price index is sourced from Knoll, Schularick, and Steger (2017). Shiller house price index from Shiller (2000), updated.

Table 1.A.19. Comparison of our housing return estimates to the Shiller series

	Shiller (2000)	Our baseline		
	Real capital gain	Real capital gain	Rental yield	Total return
Mean return p.a.	0.76	0.90	5.33	6.10
Standard deviation	6.35	7.84	0.75	8.12
Geometric mean	0.56	0.61	5.32	5.80
Observations	125	125	125	125

Note: Real housing returns, capital gains and rental yield for the US. Shiller data use the house price index from Shiller (2000).

The house price index in this paper is sourced from Knoll, Schularick, and Steger (2017). Figure 1.A.4 compares this house price index with that in Shiller (2000). The main difference between the indices in Knoll, Schularick, and Steger (2017) and Shiller (2000) is that Knoll, Schularick, and Steger (2017) use the index in Fishback and Kollmann (2015) for the period 1929–1940, which is a hedonic index based on 106 cities, and has wider geographic coverage than the Shiller (2000) index of 5 cities during this period. We normalize the indices to equal 1 in 1941 the point before which we use a different index to Shiller. The Fishback and Kollmann

(2015) index registers larger house price falls in the Great Depression, which means that extrapolated back, the house prices in the early 20th century are somewhat higher. Knoll, Schularick, and Steger (2017) also use the FHFA OFHEO index for the period 1975–2012, which reports similar growth to Shiller's index. Across the full sample, the differences between the two indices are small. For further detail on the differences between the two series and the underlying methodologies used, see Online Appendix of Knoll, Schularick, and Steger (2017). The first two columns of Table 1.A.19 summarize the key moments of the two indices. The real capital gains and their standard deviation are almost identical for the two series. Table 1.A.19 adds the rental yield estimates to arrive at a total housing return figure for the US. As discussed in Section 1.4.3, most of the real return in our series comes from the rental yield, both for the US and other countries—although the rental yield share in returns is higher in the US than in most countries, and the real capital gain—lower.[50] Therefore, for our index as for that in Shiller (2000), real capital gains are low but total housing returns are high. For equities too—especially in geometric mean terms—most of the real return comes about from the dividend yield (see Section 1.4.3 and Appendix 1.A.14).

50. As in Section 1.4.3, we subtract the inflation rate from the capital gain rather than yield.

1.A.14 Further detail on housing and equity return decomposition

Tables 1.A.20–1.A.22 present some alternative decompositions of risky returns into capital gain and dividend or rental income. Table 1.A.20 shows the decomposition of *nominal* returns, and does not take a stance on subtracting inflation from capital gains or dividend/rental income. The nominal return is split roughly 50/50 between capital gains and dividend or rental income, with, as in Table 1.8, a somewhat higher share of the equity returns accoutned for by capital gains. After 1950, roughly three-quarters of the nominal equity return comes about from capital gains. Table 1.A.21 shows the nominal return decomposition by country. Capital gains account for most of the nominal return in every country, whereas rental yield accounts for most of the housing return in all but two countries, with the remaining two countries split roughly equally between dividends and capital gains. Table 1.A.22 goes back to the real return decomposition (aggregates show in Table 1.8), but now shows the by-country splits in geometric mean terms. The importance of dividend and rental income becomes rather ubiquitous. The geometric real capital gain for housing in the US is low at 0.6, echoing the findings of Shiller (2000) (see also Table 1.A.19).Even though real capital gains n the US are the lowest in our sample, their level is by no means unusual—for example, both Spain and Sweden display similarly low capital gains, and several countries record lower geometric total returns than the US. Portugal stands out as the only country with a negative cumulative real geometric return on a risky asset (equities), largely driven by the damage done by the Carnation Revolution, and generally low dividend income.[51]

51. The Portuguese equity returns in Table 1.A.22 are from year 1948 onwards, to maintain consistency with housing data. For the period from 1870 to today, the geometric mean real total return on Portuguese equities is above zero.

Table 1.A.20. Total nominal return components for equity and housing

	Equity			Housing		
	Capital gain	Dividend income	Total return	Capital gain	Rental income	Total return
Full sample:						
Mean return p.a.	6.59	4.17	10.77	5.55	5.50	11.05
Standard deviation	21.89	1.74	22.36	10.34	2.05	10.69
Geometric mean	4.48	4.16	8.64	5.08	5.48	10.57
Observations	1707	1707	1707	1707	1707	1707
Post-1950:						
Mean return p.a.	9.25	3.80	13.06	7.02	5.22	12.24
Standard deviation	24.45	1.81	25.08	9.79	1.93	10.22
Geometric mean	6.59	3.79	10.34	6.62	5.21	11.82
Observations	995	995	995	995	995	995

Note: Average annual nominal capital gain, dividend or rental income, and total return across 16 countries, unweighted. Standard deviation in parentheses. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both equity and housing returns, capital gains and yields. Dividend and rental income expressed in percent of previous year's asset price.

Table 1.A.21. Total nominal return components for equity and housing by country

	Equity				Housing			
	Capital gain	Dividend income	Total return	Capital gain share	Capital gain	Rental income	Total return	Capital gain share
Australia	6.92	4.90	11.82	0.59	6.57	3.99	10.57	0.45
	(16.53)	(1.08)	(17.18)		(13.60)	(0.92)	(13.69)	
Belgium	6.84	3.83	10.67	0.64	5.78	6.15	11.93	0.32
	(23.73)	(1.64)	(24.35)		(10.09)	(1.46)	(9.94)	
Denmark	5.75	4.95	10.69	0.54	4.24	7.13	11.37	0.23
	(17.39)	(2.09)	(17.91)		(7.77)	(2.42)	(7.82)	
Finland	10.43	5.08	15.50	0.67	8.44	7.14	15.58	0.37
	(30.75)	(1.95)	(31.35)		(14.54)	(2.86)	(15.59)	
France	4.67	3.73	8.40	0.56	7.06	5.09	12.15	0.41
	(20.89)	(1.33)	(21.31)		(9.20)	(1.14)	(9.75)	
Germany	4.29	4.08	8.72	0.51	3.50	6.03	9.52	0.22
	(21.16)	(1.58)	(21.83)		(10.20)	(2.61)	(10.85)	
Italy	9.14	3.61	12.75	0.72	7.29	3.49	10.77	0.51
	(30.40)	(1.34)	(30.69)		(14.74)	(1.59)	(15.03)	
Japan	6.78	2.65	9.81	0.72	5.89	4.70	10.60	0.39
	(18.50)	(1.77)	(18.85)		(9.60)	(1.24)	(9.97)	
Netherlands	6.82	4.87	11.71	0.58	5.25	5.96	11.21	0.31
	(19.28)	(1.57)	(19.72)		(8.59)	(1.68)	(9.14)	
Norway	4.71	4.21	8.88	0.53	4.62	6.72	11.34	0.26
	(19.92)	(1.60)	(20.34)		(8.08)	(1.19)	(8.31)	
Portugal	9.86	2.28	12.16	0.81	8.54	4.47	13.01	0.49
	(37.95)	(1.22)	(38.24)		(11.44)	(1.98)	(12.51)	
Spain	7.40	4.53	11.65	0.62	7.20	4.16	11.36	0.46
	(20.62)	(2.30)	(21.21)		(12.95)	(1.60)	(13.28)	
Sweden	7.12	4.12	11.17	0.63	4.34	7.12	11.47	0.23
	(19.94)	(1.05)	(20.40)		(7.48)	(1.61)	(7.83)	
Switzerland	5.76	3.20	8.93	0.64	3.57	4.54	8.11	0.28
	(19.78)	(1.46)	(19.81)		(6.10)	(0.62)	(6.19)	
UK	6.20	4.53	10.73	0.58	5.39	3.94	9.33	0.41
	(20.89)	(1.39)	(21.70)		(9.81)	(0.86)	(9.97)	
USA	6.70	4.38	11.08	0.60	3.54	5.33	8.87	0.25
	(18.22)	(1.57)	(18.45)		(8.24)	(0.75)	(8.40)	

Note: Arithmetic average of annual nominal capital gain, dividend or rental income, and total return, full sample. Standard deviation in parentheses. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both equity and housing returns, capital gains and yields. Dividend and rental income expressed as percentage of previous year's asset price. Capital gain share is the proportion of nominal total return attributable to nominal capital gains.

Table 1.A.22. Real geometric return components for equity and housing by country

	Equity				Housing			
	Real capital gain	Dividend income	Real total return	Capital gain share	Real capital gain	Rental income	Real total return	Capital gain share
Australia	1.69	4.90	6.37	0.26	2.05	3.99	5.90	0.21
	(16.87)	(1.03)	(16.75)		(9.04)	(0.89)	(8.76)	
Belgium	-0.05	3.82	3.57	0.01	0.87	6.14	6.79	0.07
	(23.04)	(1.58)	(23.08)		(14.78)	(1.38)	(14.47)	
Denmark	1.44	4.93	6.23	0.23	1.01	7.10	7.95	0.07
	(15.96)	(1.99)	(15.65)		(7.00)	(2.27)	(7.15)	
Finland	1.03	5.06	5.78	0.17	1.85	7.10	8.52	0.12
	(28.56)	(1.86)	(28.33)		(13.69)	(2.67)	(13.99)	
France	-2.62	3.73	0.90	0.41	1.10	5.08	5.90	0.10
	(21.59)	(1.28)	(21.58)		(9.60)	(1.08)	(9.77)	
Germany	0.69	4.06	5.08	0.14	1.45	5.99	7.36	0.11
	(20.43)	(1.51)	(19.67)		(9.10)	(2.43)	(9.28)	
Italy	0.24	3.60	3.66	0.06	1.10	3.47	4.39	0.14
	(26.69)	(1.29)	(26.53)		(8.10)	(1.54)	(8.23)	
Japan	1.46	2.63	4.34	0.36	1.70	4.70	6.22	0.16
	(18.13)	(1.72)	(17.93)		(7.73)	(1.19)	(7.80)	
Netherlands	1.55	4.86	6.27	0.24	1.43	5.94	7.16	0.11
	(19.39)	(1.49)	(18.99)		(8.07)	(1.58)	(8.16)	
Norway	-0.24	4.19	3.79	0.05	1.16	6.71	7.68	0.08
	(19.53)	(1.53)	(19.31)		(8.17)	(1.12)	(8.11)	
Portugal	-2.76	2.27	-0.67	0.55	0.69	4.45	4.79	0.08
	(34.92)	(1.19)	(35.02)		(9.54)	(1.87)	(9.16)	
Spain	-0.18	4.50	3.81	0.04	0.63	4.15	4.56	0.07
	(19.90)	(2.18)	(19.71)		(11.16)	(1.53)	(11.10)	
Sweden	2.17	4.11	6.08	0.35	1.01	7.11	7.92	0.07
	(19.68)	(1.01)	(19.51)		(8.81)	(1.50)	(8.57)	
Switzerland	1.13	3.19	4.26	0.26	0.61	4.54	5.03	0.06
	(20.27)	(1.41)	(19.83)		(6.46)	(0.59)	(6.43)	
UK	0.68	4.52	5.02	0.13	1.25	3.94	5.05	0.14
	(19.34)	(1.32)	(18.83)		(8.77)	(0.83)	(8.70)	
USA	2.35	4.37	6.66	0.35	0.61	5.32	5.80	0.05
	(19.47)	(1.51)	(18.81)		(7.63)	(0.71)	(7.46)	

Note: Geometric average of annual real capital gain, dividend or rental income, and total return, full sample. Standard deviation in parentheses. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both equity and housing returns, capital gains and yields. Dividend and rental income expressed as percentage of previous year's asset price. Capital gain share is the proportion of real total return attributable to real capital gains.

1.A.15 Further detail on the composition of wealth

1.A.15.1 Investible assets and capital stock

Table 1.A.23. Composition of investible assets by country

Country	Housing	Equity	Bonds	Bills	Deposits	Other financial	Other non-financial
France	23.2	28.0	5.1	1.5	10.4	11.9	19.8
Germany	22.2	24.2	5.6	0.2	14.0	17.3	16.4
Japan	10.9	13.4	13.1	1.5	18.9	12.9	29.4
UK	27.5	24.8	6.1	0.2	10.7	12.6	18.1
USA	15.8	38.5	9.7	0.9	7.4	6.0	21.7
Average share	19.9	25.8	7.9	0.9	12.3	12.1	21.1

Note: Ratios to total investable assets, percentage points. End-2015.

Table 1.A.24. Composition of capital stock by country

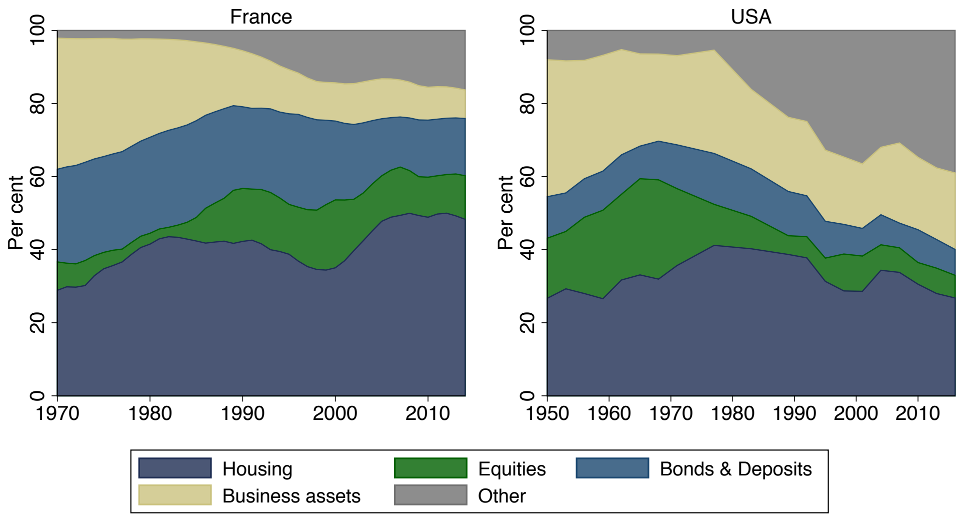
Country	Housing	Other buildings	Machinery and equipment	Other non-financial
France	53.9	25.5	4.4	16.2
Germany	57.6	38.0	1.2	3.1
Japan	27.0	47.1	6.7	19.2
UK	60.2	22.5	9.5	7.8
USA	42.2	37.0	11.4	9.3
Average share	48.2	34.0	6.6	11.1

Note: Ratios to total capital stock, percentage points. End-2015.

Tables 1.A.23 and 1.A.24 show the individual-country investible asset and capital stock shares of each individual class. The average shares correspond to those in Section 1.2.1 Figure 1.1. Housing, equity, bonds and bills account for the majority of investible assets, and housing—for around half of capital stock in most countries. Japan stands out as having a relatively large stock of government bonds and non-residential real estate, and the US for relatively high equity wealth. In relative terms, UK has the highest share of housing in investible assets and capital stock.

1.A.15.2 Household balance sheet composition

Figure 1.A.5. Household balance sheet composition in the US and France



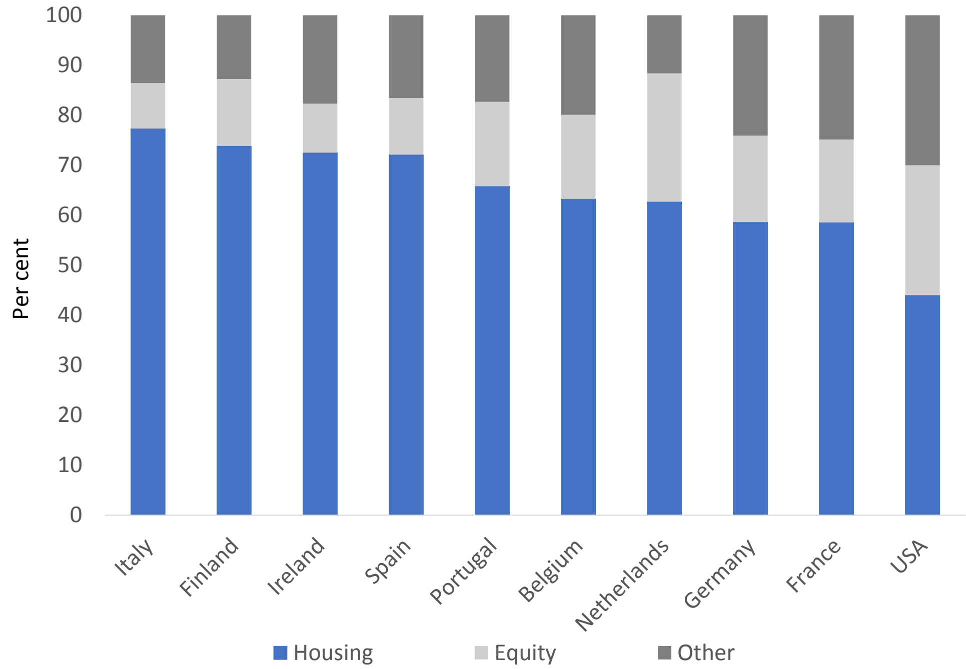
Note: Ratio of each asset class to total household assets, in percent. Housing assets are net of mortgage debt. Other assets include agricultural land, pension, insurance and investment fund claims. Sources: Survey of Consumer Finances and French Distributional National Accounts.

This paper measures returns on the major asset classes from the perspective of a representative household. To gauge the relative importance of these assets on household balance sheets, Figure 1.A.5 plots the historical composition of assets held by households in France and the US, from the mid twentieth century to today. The data for France come from the Distributional National Accounts in Garbinti, Goupille-Lebret, and Piketty (2017b), and US data are from the Historical Survey of Consumer Finances in Kuhn, Schularick, and Steins (2017). The figures present the share of each asset class in total household assets, with housing measured net of mortgage debt. Residential real estate (dark blue area) is by far the largest asset on household balance sheets. Taken together, housing and equity assets account for 40–50% of total household assets in both countries.

Bonds and deposits represent between 10 and 20 percent of assets. Aside from corporate bonds, the return on these assets is well proxied by our government bond and bill series.[52] Business assets—the light yellow bar—account for between 10 and 30 percent of wealth, depending on the country and time period. The equity returns in our sample can be used as a proxy for the levered return on business assets. Other assets include agricultural land–the remaining non-financial asset—and other

52. Data on the total returns on corporate bonds are generally sparse, but the small amount of data that we have suggest that corporate bond returns fall in-between those of government bonds and equities, and somewhat closer to government bonds in terms of level and volatility, with a roughly 1 percentage point spread between long-run government and corporate bond returns.

Figure 1.A.6. Household balance sheet composition throughout advanced economies, 2013-2014



Note: Ratio of each asset class to total household assets, in percent. Housing assets are net of mortgage debt. Other assets include business wealth, agricultural land, bonds, deposits, pension, insurance and investment fund claims. Source: ECB Household Finance and Consumption Survey and Survey of Consumer Finances.

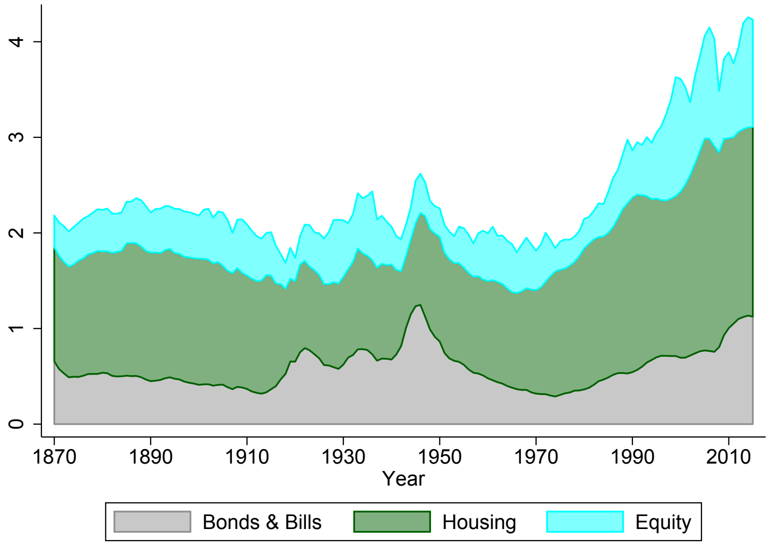
financial assets, which mainly consist of pension fund and insurance claims. Returns on these claims should, in turn, be closely related to the equity and bond series in our sample. The share of other financial assets has been growing over the past several decades, largely reflecting the fact that, especially in the US, an increasing proportion of equities is held indirectly through financial intermediaries.

Figure 1.A.6 shows the balance sheet composition for a larger selection of countries in 2013-2014, using data from the ECB Household Finance and Consumption Survey and the Survey of Consumer Finances. In other European countries, the role of housing and equities is even more important than in France and the US. In Italy, housing accounts for close to 80% of household wealth, and housing and equities together for nearly 90%. These data suggest that for our cross-country sample, the share of housing and equity assets in household balance sheets is likely to be somewhat larger than that pictured for the US and France in Figure 1.A.5.

1.A.15.3 The historical global asset portfolio

For the historical dataset in this paper, we measure the size of each asset class in country portfolios in order to calculate weighted returns, such as the return on total wealth in Section 1.2.3. Due to the lack of consistent historical data on household

Figure 1.A.7. Assets considered in this study as a share of GDP



Note: Average of asset-to-GDP shares in individual countries, weighted by real GDP. Equity is the total stock market capitalization. Housing is the stock of housing wealth. Bonds and bills are the stock of public debt.

balance sheets of the type shown in Figure 1.A.6, we focus solely on the assets in our study: equity, housing, bonds and bills. As outlined in Section 1.2.3, we weight the individual asset returns within each country according to the market-capitalization shares of the respective asset types in the country's investible wealth portfolio, to arrive at the composite return measures. Thus, by this choice of method, significant non-market asset weights are not included, notably non-traded equity wealth.

Figure 1.A.7 shows the evolution of marketable wealth in advanced economies from the late 19th century to today. In line with the more recent data on household balance sheets (Figures 1.A.5 and 1.A.6), housing has been the dominant asset in the countries' portfolios throughout the sample. Public debt has tended to increase in size after wars, and most recently after the Global Financial Crisis. The stock market has tended to be small relative to housing, but has increased in size during the last several decades. The last four decades have also seen a marked increase in the aggregate stock of assets pictured in Figure 1.A.7, in line with the findings of Piketty and Zucman (2014), who cover a broader selection of assets, but have fewer countries and observations in their sample. Taken together, the assets in our study add up to over four times the national income. For the modern period, this constitutes around 70% of total national wealth as measured by Piketty and Zucman (2014).

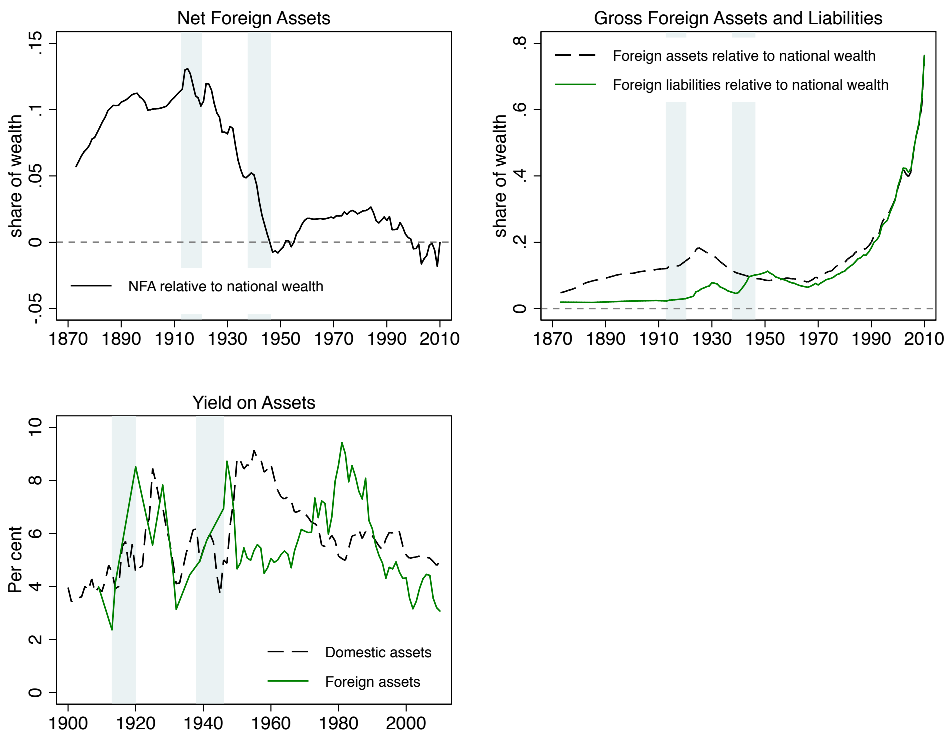
The sources for the asset portfolio data are as follows. The data on equity wealth come from Kuvshinov and Zimmermann (2018), and measure the total stock market capitalization of the specific country. These series cover the total size of the domestic stock market, excluding foreign-owned companies, and aggregating across multiple stock exchanges within the country, excluding cross listings, at each year in the historical sample. Due to data limitations we have had to rely on data for individual markets for a number of countries and historical periods (e.g., only counting the Lisbon listings, but not the Porto listings for Portugal), and rely on interpolation to construct some of the early annual estimates. The stock market capitalization data are sourced from a wide variety of publications in academic journals, historical statistical publications, and disaggregated data on stock listings and company reports of listed firms.

To measure the value of housing wealth for each country, we went back to the historical national wealth data to trace the value of buildings and the underlying land over the past 150 years. We heavily relied on the national wealth estimates by Goldsmith (Garland and Goldsmith, 1959; Goldsmith, 1962; Goldsmith, 1985a) as well as the on the collection of national wealth estimates from Piketty and Zucman (2014) for the pre-WW2 period. We also drew upon the work of economic and financial historians, using the national wealth estimates of Stapledon (2007) for Australia, Abildgren (2016) for Denmark, Artola Blanco, Bauluz, and Martínez-Toledano (2017) for Spain, Waldenström (2017) for Sweden, and Saez and Zucman (2016) for the US. For the postwar decades, we turned to published and unpublished data from national statistical offices such as the U.K. Office of National Statistics or Statistics Netherlands (1959). Particularly for the earlier periods, many of the sources provided estimates for benchmark years rather than consistent time series of housing wealth. In these cases, we had to use interpolation to arrive at annual estimates.

We use total public debt from the latest vintage of the long-run macrohistory database (Jordà, Schularick, and Taylor, 2017) as a proxy for the stock of bonds and bills, and divide public debt equally between these two financial instruments.

Even though the data come from different sources than the household balance sheets in Figures 1.A.5 and 1.A.6, the relative shares of the three asset classes are similar to the shares of similar asset holdings in the household portfolios. Under both definitions, residential real estate is the largest asset in the portfolio. The share of equity in marketable wealth (Figure 1.A.7) is somewhat larger than in household balance sheets, but since some equity is held indirectly, and equity to some extent proxies the levered return on business wealth, the relative weight in our data is broadly reflected of equity returns' contribution to the overall return on household wealth. Finally, the public debt share is similar to bonds and deposits in French household wealth data (Figure 1.A.5, left-hand graph), and to bonds, deposits and part of indirectly held assets in the US balance sheets (Figure 1.A.5, right-hand graph).

Figure 1.A.8. Size and yield on foreign assets



Note: Unweighted average of four countries: France, Germany, UK and US. Data from Piketty and Zucman (2014). Shaded areas indicate world wars. Yield on wealth calculated as domestic or foreign capital income received in proportion to (domestic or foreign) wealth.

1.A.15.4 Foreign assets

The returns in our dataset are calculated from the perspective of an investor in the domestic market, and therefore exclude the return on foreign investments. But households also hold foreign assets, and the return on these assets may impact the return on the overall household portfolio. Including these assets in our portfolio could lower the return for some countries and periods, if they held large foreign investments in foreign asset markets that were subject to subsequent disasters, such as the investments in Russia before the 1917 revolution. On the other hand, returns for some periods and countries may be higher if we include the profitable colonial asset holdings, or those in fast-growing emerging markets.

Long run data on returns on foreign assets are generally sparse. To an extent, the returns from a global investor perspective in Table 1.A.14 try to abstract from the national dimension and get around the issue of foreign holdings. Still, these returns only cover the countries in our dataset, and do not inform us about country-specific portfolios of foreign investments.

To further analyze the potential impact of foreign asset holdings on our findings, Figure 1.A.8 plots the size of foreign assets and a rough proxy of the asset return, using data for four countries—France, Germany, UK, and US—collected by Piketty and Zucman (2014). The top-left panel of Figure 1.A.8 shows the share of net foreign assets in national wealth. This has generally been small, with a high of around 10% at the turn of the 20th century (equivalent to 60–80% of national income), and near-zero net positions since the end of WW2. These zero net positions, however, mark a sharp increase in gross foreign assets and liabilities (Figure 1.A.8, right-hand panel), which have been edging towards 80% of national wealth in late 2000s. Despite these large gross positions, the net income on foreign assets and liabilities remains small, and therefore unlikely to materially affect returns on household wealth.

We can use the asset income reported in national accounts to obtain a proxy of the yield on foreign assets and compare it to that on domestic assets, with the two depicted in the bottom panel of Figure 1.A.8. The balance sheet yield on foreign assets, calculated as foreign capital income relative to foreign assets (Figure 1.A.8, bottom panel, solid line) is similar to that on domestic assets (Figure 1.A.8, bottom panel, dashed line), both in terms of size and time trend. These yield measures exclude capital gains, and are likely to understate negative returns due to foreign asset market disasters. Recent research by Meyer, Reinhart, and Trebesch (2015), however, shows that the average return on risky foreign assets, even during asset market disasters, is surprisingly high. Meyer, Reinhart, and Trebesch (2015) find that the total return on defaulted foreign government bonds is generally positive and higher than inflation, using historical data that covers all foreign bonds listed on major exchanges back to the early 1800s—largely due to high pre-default interest payments, and low haircuts in the event of default.

1.A.16 Equally-weighted portfolio returns

Table 1.A.25. Equally-weighted portfolio returns

Country	Portfolio weights		Equal weights	
	Risky return	Return on wealth	Risky return	Return on wealth
Australia	6.96	5.91	7.13	5.51
Belgium	8.31	6.38	7.71	6.10
Denmark	8.02	7.37	7.74	6.27
Finland	10.87	9.76	9.81	7.95
France	6.54	4.92	5.56	4.22
Germany	7.90	7.07	7.47	6.32
Italy	5.32	5.08	6.01	5.07
Japan	6.79	5.59	6.27	4.91
Netherlands	7.30	5.33	7.12	5.36
Norway	7.96	6.86	6.85	5.33
Portugal	6.46	5.87	5.41	4.06
Spain	5.39	4.58	5.70	4.27
Sweden	8.52	7.41	8.16	6.49
Switzerland	6.51	5.63	6.23	4.93
UK	6.35	4.75	6.13	4.72
USA	7.12	6.03	7.28	5.63
Average, unweighted	7.44	6.30	7.01	5.54
Average, weighted	7.22	5.98	6.98	5.49

Note: Average annual real returns for the full sample. The portfolio-weighted averages use country-specific stocks of housing, equity, bonds and bills as weights for the individual asset returns. Portfolio-weighted risky return is a weighted average of housing and equity, using stock market capitalization and hosuing wealth as weights. Portfolio-weighted real return on wealth is a weighted average of equity, housing, bonds and bills, using stock market capitalization, housing wealth and public debt stock as weights. Equally-weighted risky return is an unweighted average of housing an equity. Equally-weighted return on wealth is an unweighted average of housing, equity and bonds. Period coverage differs across countries. Consistent coverage within countries: each country-year observation used to compute the statistics in this table has data for both the real risky return, and the return on overall wealth. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country returns.

Table 1.A.25 assesses the impact of portfolio weighting on our return estimates. The weighting has a relatively small impact on the risky rates, because returns on housing and equity are generally similar. It raises the return on capital by around one percentage point, because the outstanding stock of public debt is substantially smaller than that of risky assets. The basic stylized facts of $r \gg g$, and high long-run risky returns continue to hold regardless of the weighting, both on average and across the individual countries in our sample.

1.A.17 Correlations between *r* and *g*

Table 1.A.26. Correlations between return on wealth and real GDP growth across different time horizons

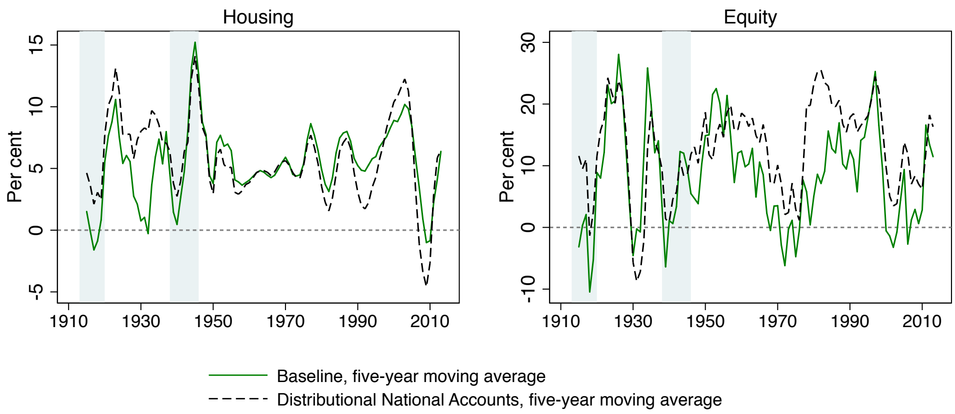
	1-year	5-year	10-year	20-year
Full sample:				
r-g, g	-0.35	-0.34	-0.23	-0.40
r, g	0.14	0.16	0.17	-0.02
r-g, r	0.88	0.87	0.92	0.93
Observations	1763	336	152	58
Post-1950:				
r-g, g	-0.08	-0.06	-0.31	-0.59
r, g	0.26	0.35	0.19	-0.21
r-g, r	0.94	0.91	0.88	0.91
Observations	1008	200	91	30

Note: Pairwise correlations of data averaged over 1-year, 5-year, 10-year and 20-year non-overlapping moving windows. 16 countries, equally weighted. r is the real return on capital, g is real GDP growth, and $r - g$ is the gap between the two.

Table 1.A.26 presents the pairwise correlations between the real GDP growth, g, and the $r-g$ gap, as well as between the gap and r, and also between r and g. The $r-g$ gap and g show a negative correlation, although the size of the correlation coefficient is not overly large. The correlation generally becomes stronger and more robust at longer time horizons, in particular when looking at 20-year periods. Even though the correlation between r and g has become close to zero at short horizons after 1950, the longer-run correlation has remained negative. The negative correlation between the $r-g$ gap and g is explained by the fact that the rate of return r is relatively uncorrelated with the GDP growth rate g, or at least shows no robustly negative correlation. The $r-g$ gap is mainly driven by changes in the rate of return r, as the two show strongly positive comovement at short and long horizons, before and after 1950.

1.A.18 Comparison to balance sheet returns on housing and equity wealth

Figure 1.A.9. Our return estimates compared to Distributional National Accounts



Note: Five-year moving average returns for the US. Baseline refers to estimates in this paper. Distributional National Accounts returns are from Piketty, Saez, and Zucman (2018).

The return on an asset from a national accounts perspective, or the "balance sheet approach" to returns, r_t^{BS} is calculated as follows:

$$\text{Balance sheet return:} \quad r_t^{BS} = \underbrace{\frac{\text{Capital Income}_t}{\text{Wealth}_t}}_{\text{Yield}} + \underbrace{\Delta\text{Wealth}_t - \text{Investment}_t}_{\text{Capital Gain}} \tag{1.A.1}$$

Similarly to the market return in equation (1.1), the balance sheet return can be divided into the yield—or capital income relative to the stock of wealth—and capital gain, which is the residual accumulated wealth, net of investment. These returns can be computed for aggregate wealth and national income, and for individual asset classes. The balance sheet returns on housing should be comparable to market returns, whereas corporate wealth returns should be somewhat higher than market data because they are computed gross of corporate taxes.

Figure 1.A.9 compares the five-year moving average of the housing and equity returns in this paper (solid line) to those computed using the balance sheet approach for the US in Piketty, Saez, and Zucman (2018) (dashed line; further: PSZ). In terms of equation (1.A.1), the yield on housing is net rental income, and the yield on equities is total profits, gross of corporate taxes and net of depreciation. PSZ primarily rely on the Shiller (2000) index, rather than investment and wealth data, to estimate the capital gains on housing. Our housing returns are very similar to PSZ, both in level and trend (Figure 1.A.9, left-hand side). The slight differences in the 1930s and 1990s arise because of differences between the Shiller (2000) house price

index used by PSZ and the housing index in Knoll, Schularick, and Steger (2017) used in this paper.[53]

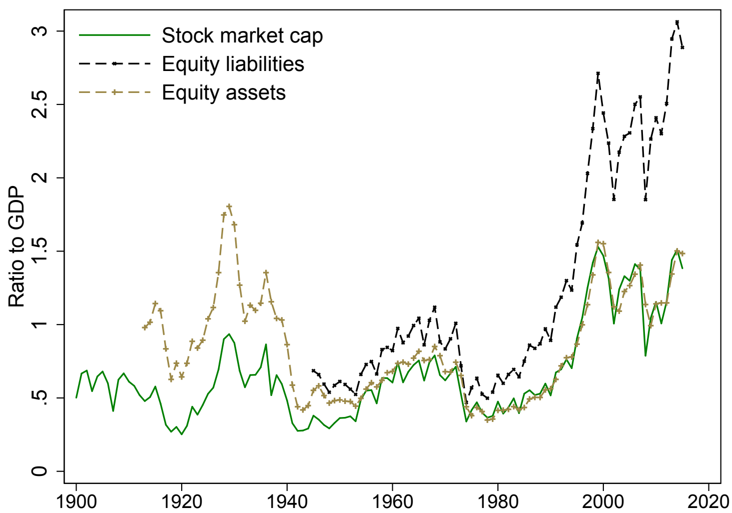
Equity returns follow a similar trend, but the level of balance sheet return is higher than the market return in our data. Part of the difference is due to the fact that balance sheet returns are gross of corporate taxes (see Section 1.A.19 and Table 1.A.27). Adding back corporate taxes, however, still leaves our equity returns somewhat below those in Piketty, Saez, and Zucman (2018). The remaining difference could either be explained by higher returns of unlisted relative to listed equities, or measurement error in one of the series. Even though it is difficult to precisely attribute the remaining difference, it is unlikely that the higher returns on unlisted equities play an important role: Moskowitz and Vissing-Jørgensen (2002) find that unlisted equities face a similar or worse risk-return tradeoff relative to listed equity, and the national accounts estimates of return on unlisted equity are themselves derived from listed equity data. It is also unlikely that there is much measurement error in the listed equity return data for the US, since these returns are directly observable, recorded in historical stock listings, and have been widely studied. It is somewhat more likely that there is some measurement error in the national accounts data—i.e., the figures either overstate the profits of corporations, understate equity wealth, or overstate the capital gains (i.e., overstate the growth in equity wealth or understate investment).

There is some tentative evidence that equity wealth in the US National Accounts may be somewhat underestimated during the mid-20th century. Figure 1.A.10 shows that during the mid 20th century, US equity wealth was similar to the market capitalization of listed US firms, and equity liabilities were only slightly above the listed stock market cap. In principle, equity liabilities and assets should include the equity of unlisted as well as listed firms, and should be substantially higher than the market capitalization.[54] The closeness of the three series during the mid 20th century therefore suggests that the size of the US equity wealth in national accounts may be somewhat understated, which could account for some of the remaining return differential between our data and balance sheet approach estimates in Piketty, Saez, and Zucman (2018). Alternatively, if we take the estimates in Figure 1.A.10 at face value, they suggest that the share of unlisted firm equity wealth is small relative to listed wealth, meaning that return differentials between listed and unlisted companies, even if large, cannot explain the remaining return differential.

53. For further details on the differences in house price indices of Knoll, Schularick, and Steger (2017) and Shiller (2000), see Appendix 1.A.13.

54. Equity liabilities are more comparable to stock market capitalization, since much of the US equity liabilities are held by foreigners and do not appear on the asset side of the (consolidated) household balance sheet.

Figure 1.A.10. US corporate equity assets and liabilities compared to capitalization of listed US firms



Note: Equity wealth and equity liabilities include both listed and unlisted firms (data from Piketty, Saez, and Zucman, 2018). Stock market capitalization includes listed US firms on all US stock exchanges (data from Kuvshinov and Zimmermann, 2018).

1.A.19 Adjusting for corporate taxes and leverage

Throughout this paper, we calculate the return on equity from a household perspective, as the pre-tax cashflows and capital gains received by the equity holder, consistent with the literature on stock market returns (see, for example Shiller, 1981). The market return, however, differs from the return on the underlying asset—i.e., business wealth—in two ways. First, it is net of corporate profit taxes, which are paid out before distribution. Second, it is a levered return, in the sense that corporate equity is a levered claim on business wealth. Adjusting for corporate taxes and leverage would allow us to proxy the return on business wealth. Our housing returns already provide a proxy for the return on housing wealth (see also Section 1.A.18.

We gross up the equity returns as follows. First, from Section 1.2.3 note that the real equity return $r_{eq,t}$ is the nominal capital gain CG_t plus dividend income Y_t, deflated by the inflation rate π_t:

$$r_{eq,t} \;=\; (CG_t + Y_t)/(1 + \pi_t) \tag{1.A.2}$$

To gross up for corporate taxes, we scale up Y_t to add back all profit taxes (both distributed dividends and retained earnings) attributable to equity wealth:

$$Y_t^{gross} \;=\; Y_t + Y_t * \frac{\text{Profit}_t}{\text{Dividends}_t} * \frac{\tau}{1-\tau_t} * \frac{\text{Equity Wealth}_t}{\text{Total non-housing wealth}_t}, \tag{1.A.3}$$

where τ_t is the effective corporate tax rate. The second term on the right hand side is our proxy for the corporate tax payments of listed firms. We start with the dividends paid by listed firms Y_t, and scale them up to estimate the total post-corporate-tax profits of listed firms, $Y_t * \frac{\text{Profit}_t}{\text{Dividends}_t}$.[55] We then scale the post-tax profits up by $1/(1-\tau_t)$ to proxy the pre-tax profits. The difference between pre- and post- tax profit—i.e., the corporate tax—is then $Y_t * \frac{\text{Profit}_t}{\text{Dividends}_t} * \frac{\tau}{1-\tau_t}$.[56] In principle, corporate tax falls on both corporate equity and business wealth. We follow Piketty, Saez, and Zucman (2018) and assume that the share of corporate taxes attributable to equities is equal to the ratio of equity wealth to total non-housing wealth.[57] The real total return gross of corporate tax is then computed as follows:

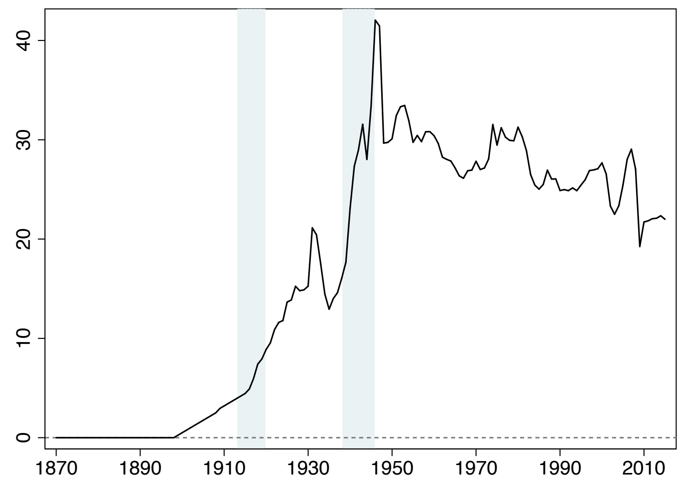
$$r_{eq,t}^{gross} \;=\; (CG_t + Y_t^{gross})/(1 + \pi_t) \tag{1.A.4}$$

55. This assumes that the payout ratios for listed firms are similar to those of unlisted firms.

56. The corporate tax paid is pre-tax profits $Y_t * \frac{\text{Profit}_t}{\text{Dividends}_t} * \frac{1}{1-\tau_t}$ less net profits $Y_t * \frac{\text{Profit}_t}{\text{Dividends}_t}$, equals $Y_t * \frac{\text{Profit}_t}{\text{Dividends}_t} * \frac{\tau}{1-\tau_t}$.

57. This ratio of around one-third is consistent, or slightly above the estimates in Graham, Leary, and Roberts (2015), that is computed from firm balance sheet data.

Figure 1.A.11. Effective corporate tax rate, average of 5 countries



Note: Average effective tax rate in Australia, France, Germany, Japan and US, equally weighted. Japanese tax rate interpolated between 1900 and 1930. Effective tax rate is total taxes paid / net corporate profits. Where effective data are not available, we extrapolate the series using statutory (top marginal) tax rates.

Finally, we also calculate the delevered pre-tax return on business wealth $r_{eq,t}^{delev,gross}$ as:

$$r_{eq,t}^{delev,gross} = ([CG_t + Y_t^{gross}] * (1 - \alpha_t) + \alpha_t * r_{debt,t})/(1 + \pi_t), \qquad (1.A.5)$$

where α_t is corporate leverage, and $r_{debt,t}$ is the interest paid on corporate debt liabilities.

As an example, assume that in a given year the total return on equity is 10%, of which 5% are capital gains and 5% are dividends, and inflation is 3%, so the real return is 7%. Assume the corporate tax rate is 25%, half the profits are paid out as dividends, and the ratio of equities to total non-housing wealth is 1/3. The pre-tax nominal yield Y_t^{gross} is then $5\% + 5\% * 2 * 0.33 * 0.33 = 6.1\%$, and the pre-tax real total return $r_{eq,t}^{gross}$ is 8.1%. Now further assume that corporate leverage is 1/3, and the corporate financing rate is 5%. The delevered real pre-tax return is then roughly $11.1 * 0.67 + 5 * 0.33 - 3$, or 6.7%.

We use the following data to calculate the pre-tax and delevered return proxies. For τ_t, we use the effective corporate tax rate, or corporate taxes paid in proportion to gross (post-depreciation) corporate profits. We have collected annual corporate tax data for a subsample of five countries, four from 1870 (Australia, France, Germany, and US), and one from 1930 (Japan). For countries and years where effective tax data were not available, we have approximated the effective rate using changes in statutory rates. Figure 1.A.11 shows the evolution of the effective corporate tax rate from 1870 to today. Taxes were near zero before the 1920s, before increasing and peaking at around 40% of profits after WW2, and declining slightly thereafter.

Table 1.A.27. Adjusting equity returns for corporate taxes and corporate leverage

Country	Baseline	Gross of corporate tax	Gross of corporate tax, delevered
Australia	8.39	9.41	8.07
France	3.21	3.69	3.00
Germany	8.83	9.02	7.77
Japan	9.49	12.18	10.12
United States	8.40	10.20	8.74
Average	7.52	8.66	7.35

Note: Annual total real equity returns. Baseline returns are the raw data in this paper. Gross of corporate tax returns add back the share of corporate profit taxes attributable to business equity. Gross of corporate tax, delevered adds back corporate taxes and delevers the pre-tax equity return. Coverage differs across countries. Consistent coverage within countries. Average is the unweighted mean of all observations for which we have tax data, and is thus not exactly equal the average of the country-specific means.

The ratio of profits to dividends largely comes from national accounts and various historical statistics, and covers a slightly smaller sample of countries and years, with gaps filled using sample averages. The share of equity wealth to total non-financial assets, and corporate leverage data are only available for the US, from Piketty, Saez, and Zucman (2018) and Graham, Leary, and Roberts (2015) respectively. US corporate leverage is measured as total debt relative to capital. We assume that the equity share in total wealth, and corporate leverage in other countries follow the US trend. We are grateful to Mark Leary for sharing the corporate leverage data with us. For the interest rate on corporate debt, we use the government bond yield plus a 1 percentage point spread (broadly indicative of corporate bond and business lending spreads).

Table 1.A.27 shows the effect of adjusting equity returns for corporate taxes and leverage. Column 1 is the baseline average real total equity return equivalent to Table 1.2, shown for the corporate tax data sample, separately for each country and in aggregate. Column 2 adds back corporate taxes using equations (1.A.3) and (1.A.4). Column 3 delevers the returns using equation (1.A.5). Adding back corporate taxes increases the average real equity return by +1.1 pps per year, with the country-specific impact ranging from +0.5 pps in France to +2.7 pps in Japan. Delevering reduces this pre-tax return by 1.3 pps per year, with the delevered pre-tax business equity return (Table 1.A.27 column 3) close to our baseline average.

Appendix 1.B Data appendix

1.B.1 Data overview

Table 1.B.1. Overview of bill and bond data

Country	Bills		Bonds	
	Period	Type of rate	Period	Type of bond
Australia	1870–1928	Deposit rate	1900–1968	Long maturity, central gov't
	1929–1944	Money market rate	1969–2015	Approx. 10y, central gov't
	1948–2015	Government bill rate		
Belgium	1870–1899	Central bank discount rate	1870–1913	Perpetual
	1900–1964	Deposit rate	1914–1940	Long maturity, central gov't
	1965–2015	Government bill rate	1941–1953	Perpetual
			1954–2015	Approx. 10y, central gov't
Denmark	1875–2015	Money market rate	1870–1923	Perpetual
			1924–1979	Long maturity, central gov't
			1980–2015	Approx. 10y, central gov't
Finland	1870–1977	Money market rate	1870–1925	Long maturity, central gov't
	1978–2015	Interbank rate	1926–1991	Approx. 5y, central gov't
			1992–2015	Approx. 10y, central gov't
France	1870–1998	Money market rate	1870–1969	Perpetual
	1999–2015	Government bill rate	1970–2015	Long maturity, central gov't
Germany	1870–1922	Money market rate	1870–1878	Long maturity, local gov't
	1924–1944	Interbank rate	1879–1943	Long maturity, central gov't
	1950–2015	Money market rate	1948–1955	Mortgage bond
			1956–2015	Long maturity, central gov't
Italy	1870–1977	Money market rate	1870–1913	Perpetual
	1978–2015	Government bill rate	1914–1954	Long maturity, central gov't
			1955–2015	Approx. 10y, central gov't
Japan	1876–1956	Deposit rate	1881–1970	Long maturity, central gov't
	1957–2015	Money market rate	1971–2015	Approx. 10y, central government
Netherlands	1870–1957	Money market rate	1870–1899	Perpetual
	1958–1964	Central bank discount rate	1900–1987	Long maturity, central gov't
	1965–2015	Money market rate	1988–2003	$\geq$ 8y, central government
			2004–2015	Approx. 10y, central government
Norway	1870–2015	Deposit rate	1870–1919	Long maturity, central gov't
			1920–2015	Approx. 10y, central gov't
Portugal	1880–1914	Money market rate	1870–1974	Long maturity, central gov't
	1915–1946	Central bank discount rate	1975–2015	Approx. 10y, central gov't
	1947–1977	Deposit rate		
	1978–2015	Money market rate		
Spain	1870–1921	Money market rate	1900–1994	Long maturity, central gov't
	1922–1974	Deposit rate	1995–2015	7–8y, central government
	1975–2015	Money market rate		
Sweden	1870–1998	Deposit rate	1874–1918	Long maturity, central gov't
	1999–2015	Government bill rate	1919–1949	Perpetual
			1950–2015	Approx. 10y, central gov't
Switzerland	1870–1968	Deposit rate	1900–2006	Long maturity, central gov't
	1969–2015	Money market rate	2007–2015	Approx. 10y, central gov't
United Kingdom	1870–2015	Money market rate	1870–1901	Perpetual
			1902–1989	Approx. 20y, central gov't
			1990–2015	Approx. 15y, central gov't
United States	1870–2013	Deposit rate	1870–1926	Approx. 10y, central gov't
	2014–2015	Money market rate	1927–2015	Long maturity, central gov't

Table 1.B.2. Overview of equity and housing data

Country	Equity			Housing	
	Period	Coverage	Weighting	Period	Coverage
Australia	1870–1881	Listed abroad	Market cap	1901–2015	Urban
	1882–2015	Broad	Market cap		
Belgium	1870–2015	All share	Market cap	1890–1950	Urban
				1951–1961	Mixed
				1977–2015	Nationwide
Denmark	1873–1899	All share	Market cap	1876–1964	Mixed
	1900–1999	Broad	Market cap	1965–2015	Nationwide
	2000–2001	Blue chip	Market cap		
	2002–2015	All share	Market cap		
Finland	1896–1911	Broad	Book cap	1920–1964	Urban
	1912–1969	All share	Market cap	1965–1969	Mixed
	1970–1990	Broad	Market cap	1970–2015	Nationwide
	1991–2015	All share	Market cap		
France	1870–2015	Blue chip	Market cap	1871–1935	Urban
				1936–1948	Mixed
				1949–2015	Nationwide
Germany	1870–1889	Blue chip	Market cap	1871–1912	Mixed
	1890–1913	All share	Market cap	1913–1938	Urban
	1914–1959	Blue chip	Market cap	1939–1947	Mixed
	1960–2015	Broad	Market cap	1948–1970	Nationwide
				1971–2015	Mixed
Italy	1870–1887	Selected stocks	Book cap	1928–1998	Urban
	1888–2015	Broad	Market cap	1999–2015	Mixed
Japan	1882–1975	Broad	Transaction volume	1931–1946	Urban
	1976–2015	All share	Mix of equal and market cap	1947–2015	Mixed
Netherlands	1900–2003	Broad	Mostly market cap	1871–1969	Mixed
	2004–2015	All share	Market cap	1970–2015	Nationwide
Norway	1881–1920	All share	Market cap	1871–2015	Urban
	1921–1955	All share	Mix of equal and book cap		
	1956–2000	All share	Mix of book cap & company turnover		
	2001–2015	Most traded shares	Market cap		
Portugal	1871–2015	All share	Market cap	1948–2015	Mixed
Spain	1900–2015	All share	Market cap	1901–1957	Mixed
				1958–2015	Nationwide
Sweden	1871–2001	Broad	Market cap	1883-1959	Urban
	2002–2015	All share	Market cap	1960–2015	Mixed
Switzerland	1900–1925	All share	Market cap	1902–1930	Urban
	1926–1959	Broad	Equally weighted	1931–1940	Mixed
	1960–2015	Broad	Market cap	1941–2015	Nationwide
United Kingdom	1870–1928	All share	Market cap	1895–1899	Urban
	1929–1963	Blue chip	Market cap	1900–1913	Mixed
	1964–2015	All share	Market cap	1914–1929	Urban
				1930–1946	Mixed
				1947–2015	Nationwide
United States	1872–2015	Broad	Market cap	1891–1952	Urban
				1953–2015	Mixed

1.B.2 Housing returns

The total return on housing is a combination of the capital gain and rental income. The capital gain series are computed from the house price data in Knoll, Schularick, and Steger (2017). The rental indices are drawn from the unpublished PhD thesis of Knoll (2017), which also extended the dataset in Knoll, Schularick, and Steger (2017) to cover three additional countries (Italy, Portugal and Spain). The sources for these new series are reproduced in Appendix 1.B.4. This section starts by describing the construction of the rental yield series. Appendix 1.B.3 then describes the general methodology for constructing the rental indices, and Appendix 1.B.4 details the sources of the rent data.

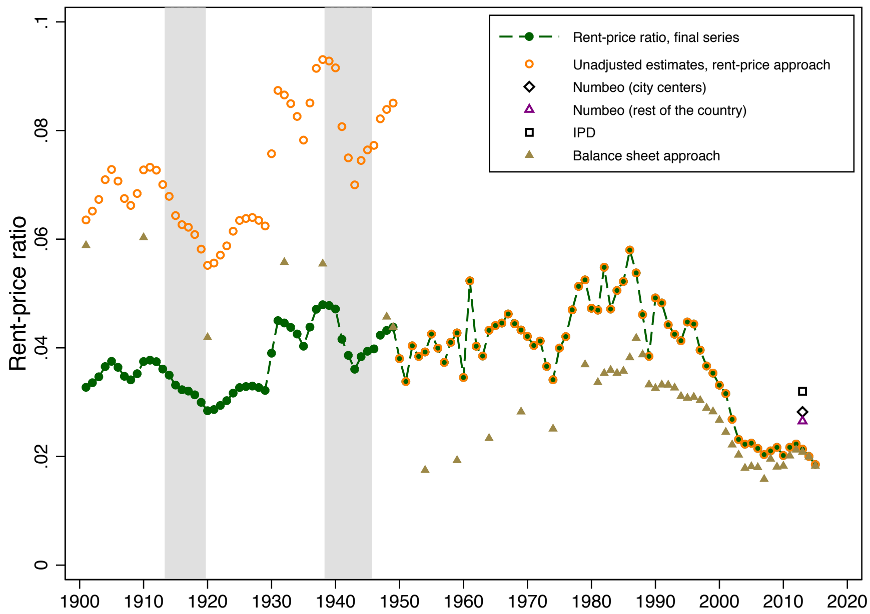
As described in Section 1.2.3, the baseline housing return series is constructed using the rent-price approach. To do this, we take a benchmark net rent-price ratio—adjusted down for maintenance and other costs—in the year 2012, 2013 or 2014, and extrapolate it back using growth in the house price and rent indices. We further check the rent-price approach estimates against various alternative historical benchmarks. These include the balance sheet approach constructed from National Accounts data (see Section 1.3.3 for more detail on this method), and independent estimates from books, journal articles and historical newspapers.

If the rent-price approach estimate differs substantially from those in the alternative sources, we adjust it so that the estimates are in line with each other. We do not adjust the series when these differences are small, or we have good reasons to doubt the quality of the alternative estimates. When we do adjust, we either benchmark our series to historical net rent-price ratios from alternative sources, or adjust the growth in the rental index by a multiplicative factor, such that the different estimates of historical rent-price ratios are broadly in line with each other.

In each of the Appendix Figures 1.B.1 to 1.B.16, the series that we use in the paper are the "Rent-price ratio, final series" estimates denoted as green circles. These incorporate any adjustments made to bring the data into line with historical sources. Alongside these, we also present the raw unadjusted rent-price approach series—orange circles—and the alternative historical estimates themselves. We also show alternative benchmark estimates for the present day to help assess the reliability of our baseline IPD rent-price ratio. These are generally sourced from data on rental expenditure and property values on Numbeo.com, for one- and three-bedroom apartments i). within city-centres and ii). in the rest of the country, and are adjusted down by us to proxy the impact of running costs and depreciation. For cases where data on running costs and depreciation were not available, we estimate these to be about one-third of gross rent, in line with the recent and historical experience in most countries (see Figure 1.3). For Australia and US, we additionally make use of benchmark rent-price ratio estimates based on detailed transaction-level data. In two countries—Australia and Belgium—we judge one of these alternative modern-day benchmarks to be more reliable than the IPD ratio, and use it to construct our final baseline net rent-price ratio series.

Australia

Figure 1.B.1. Australia: plausibility of rent-price ratio



For 2014, Fox and Tulip (2014) report a gross rental yield of 4.2 per cent, running costs excluding taxes and utilities of 1.1 per cent, and depreciation rate of 1.1 per cent, using data covering almost all properties advertized for rent in major Australian cities. This gives us a benchmark net rent-price ratio of 0.02. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.1. We make one adjustment to these series to correct for possible mismeasurement of rental growth when lifting the wartime price controls in 1949/50 (see below for details). This gives us the adjusted final rent-price ratio series—the green-circled line in Figure 1.B.1—used in this paper.

We obtain several scattered independent estimates of rent-price ratios in Australia. First, the IPD database (MSCI, 2016) reports a net rent-price ratio of 0.032 for the Australian residential real estate in 2013 (black square in Figure 1.B.1). Balance sheet approach estimates (brown triangles) are obtained using a variety of sources. OECD (2016b), Stapledon (2007), Australian Bureau of Statistics (2014) and Butlin (1985) provide estimates of gross rental expenditure and various maintenance and running costs, as well as depreciation, for present-day and historical periods. As with the benchmark yield calculation, we subtract all non-tax and non-utilities related running costs, plus depreciation, to calculate total net rental expenditure. We then combine it with the housing wealth data from Stapledon (2007) and Piketty and Zucman (2014) to calculate the net rental yield.

The historical balance-sheet approach estimates are broadly in line with the unadjusted rent-price approach series (orange circles) over recent decades, but below it for the earlier years. Note that the long-run rent-price ratio shows a structural break in 1949/1950 stemming from a surge in house prices after the lifting of wartime price controls in 1949 (price controls for houses and land were introduced in 1942). While the abandonment of price controls undoubtedly had an effect on house prices, it is unclear whether it also resulted in a single sudden shift in the relationship between house prices and rents. To guard against measurement uncertainty, we benchmark our historical rent-price ratio to the balance sheet approach estimate in 1949. Figure 1.B.1 shows that the adjusted long-run rent price ratio—the green circle line—generally concords with the balance-sheet approach estimates, being on average slightly lower during 1900–1940, and higher during 1950–1980.

Finally, modern-day gross rental yield estimates are available from Numbeo.com for one- and three-bedroom apartments i). within city-centres and ii). in the rest of the country. We adjust these down using the cost estimates from Fox and Tulip (2014) to obtain a proxy of net yield. The resulting estimates fall in-between those of the MSCI (2016), and the other approaches.

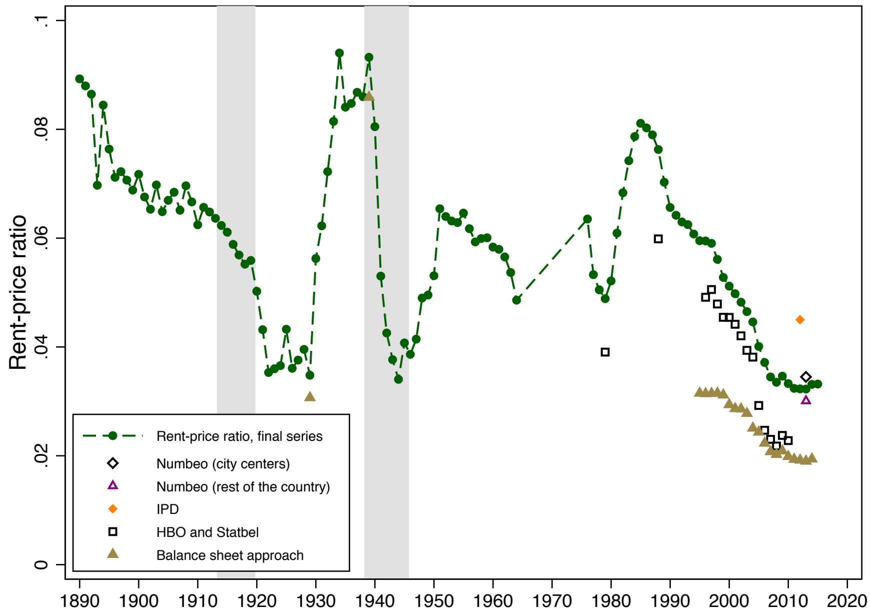
Belgium

We construct the benchmark rent-price ratio using the rental yield data from Numbeo.com, taking the average of in- and out-of-city-centre apartments, and adjusting down one-third to account for running costs and depreciation. This gives us a benchmark net rent-price ratio of 0.033 for 2012. Applying the rent-price approach gives us the long-run net rent-price ratio series depicted as green circles in Figure 1.B.2, which are the estimates used in this paper. Please note that the benchmark rent-price ratio from the IPD (MSCI, 2016)—0.045 for 2012—is substantially higher than the alternative approaches, which is why we rely on estimates from Numbeo.com instead.

We construct four independent estimates of rent-price ratios. First, for 1978–2010, Statistics Belgium publish estimates of average rental expenditure and house prices (Statistics Belgium, 2013b; Statistics Belgium, 2015). Assuming around one-third of gross rent is spent on maintenance, running costs and depreciation, this gives us a series of net rent-price ratios, depicted as square dots in Figure 1.B.2. The resulting series are consistent with both the level and the time trend in our baseline series constructed using the rent-price approach.

Second, we construct estimates of gross rent-price ratios using the balance-sheet approach, based on data on rental expenditure and housing wealth, and scale these down one-third to obtain the net yield proxy. For the modern period, Poullet (2013)

Figure 1.B.2. Belgium: plausibility of rent-price ratio



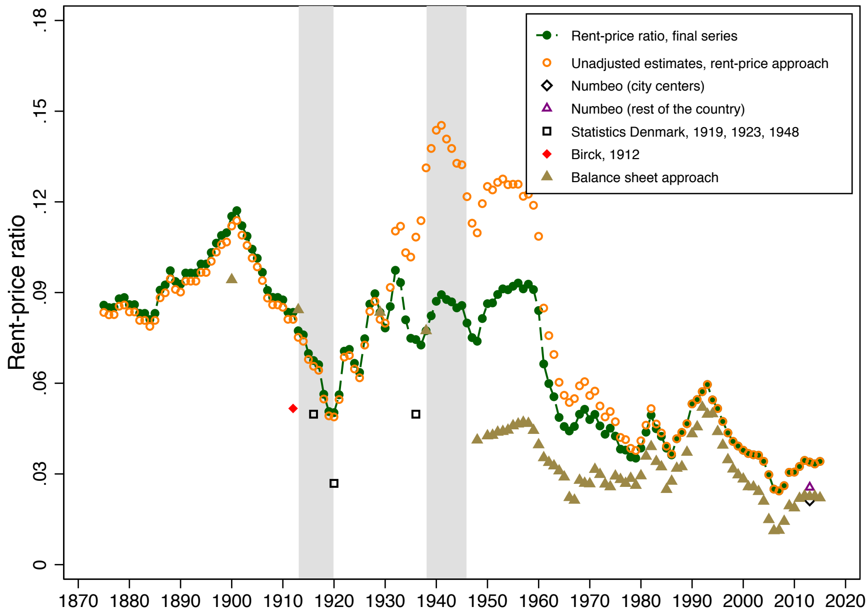
provides estimates of housing wealth, and Statistics Belgium (2013a) and OECD (2016b) of rental expenditure. For historical series, Peeters, Goossens, and Buyst (2005) reports estimates of total gross and net rents on all dwellings, which we scale down to obtain an estimate of net rental expenditure on residential real estate. Goldsmith and Frijdal (1975) report estimates of housing wealth for 1948–1971, which we extend back to 1929 using data in Goldsmith (1985a), and assuming a constant share of land to residential property value. The resulting net rental yield estimates are somewhat below our baseline rent-price ratio for the modern period, and broadly in line with its historical levels, falling within a reasonable margin of error given the substantial uncertainty in the Belgian housing wealth estimates.

We would like to thank Stijn Van Nieuwerburgh for sharing historical rent and house price data for Belgium.

Denmark

We obtain several additional estimates of rent-price ratios in Denmark throughout the past century and a half. First, we construct estimates using the balance sheet approach using data on total rental expenditure (Hansen, 1976; OECD, 2016b; Statistics Denmark, 2017b) and housing wealth (Abildgren, 2016). We estimate housing running costs and depreciation as fixed proportions of dwelling intermediate consumption, and depreciation of all buildings (Statistics Denmark, 2017a), and subtract these from gross rental expenditure to produce net rental yield estimates. The balance sheet approach yields are similar to the rent-price approach for the recent

Figure 1.B.3. Denmark: plausibility of rent-price ratio



decades and in the early 20th century, but diverge somewhat in the 1940s and 50s. Both estimates are subject to measurement error, but the large difference suggests that some of the high levels of the rent-price approach ratio may be a result of the rental index underestimating the rent growth during this period. To guard against accumulation of errors in the rent-price approach, we benchmark the historical yield to the balance sheet approach estimates in 1938 and 1929, and adjust the rent-price ratio growth for the in-between years, with the final series (green circles) being somewhere in-between the balance-sheet and rent-price approaches. For earlier the historical period, the rent-price and balance-sheet approaches display similar levels and time trend.

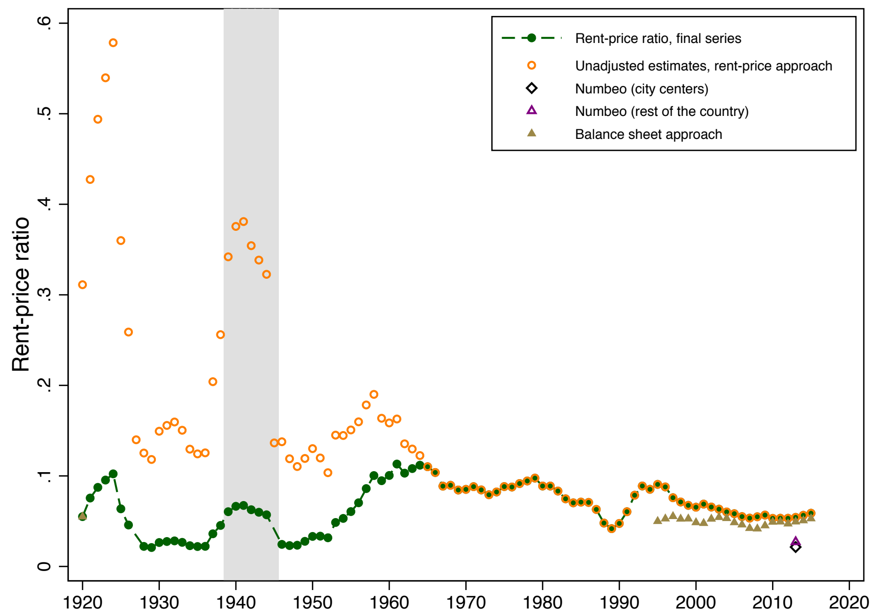
Our baseline rent-price ratio estimates are also in line with two further historical sources. First, according to Birck (1912), at the time of his writing, housing values in Copenhagen typically amounted to 13 times the annual rental income. Second, in line with this estimate, Statistics Denmark (1919) reports that housing values in urban areas in 1916 were about 13.5 times the annual rental income (note that housing values reported in Statistics Denmark (1919), Statistics Denmark (1923), Statistics Denmark (1948), and Statistics Denmark (1954) relate to valuation for tax purposes). These data imply a gross rent-price ratio of about 0.06–0.07, and a net rent-price ratio of around 0.04–0.05. For 1920, Statistics Denmark (1923) states that housing values in urban areas were about 25 times the annual rental income implying a gross rent-price ratio of roughly 0.04 (roughly 0.03 net). In 1936, rent-price ratios in urban areas had returned to pre-WW1 levels (Statistics Denmark, 1948).

Finally, estimates of net rent-price ratios based on data from www.Numbeo.com are similar to the modern-day values for the balance-sheet and rent-price approaches.

For 2013, the MSCI (2016) reports the rent-price ratio for Danish residential real estate of 0.034. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.3. We make one adjustment to these series to correct for possible mismeasurement of rental growth around WW2 (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.3—used in this paper.

Finland

Figure 1.B.4. Finland: plausibility of rent-price ratio



For 2013, the MSCI (2016) reports the rent-price ratio for Finnish residential real estate of 0.054. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.4. We make one adjustment to these series to correct for possible mismeasurement of rental growth during the rent controls imposed in the early-to-mid 20th century (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.4—used in this paper.

We obtain two alternative estimates of the net rent-price ratio for the modern period. First, we construct proxies of gross rental expenditure, running costs and depreciation, and total housing wealth back to 1995 using data from Statistics Finland and OECD. These are roughly the same as our benchmark rent-price ratio for the benchmark year, but are slightly lower in the late 1990s. Note, however, that data from Statistics Finland imply a housing depreciation rate of 3.5%, and running and maintenance costs of around 2%, which corresponds to an expected duration of the structure of less than 20 years. Therefore, the cost estimates are almost certainly too high, and adjusting these to more reasonable levels would leave the rent-price ratios on par, or above our baseline values. For 2013, we also obtain estimates of rent-price ratios for one- and three-bedroom apartments i) within city-centers and ii) in the rest of the country from www.Numbeo.com. Once adjusted for costs, these are somewhat lower than both the estimates using the rent-price and balance sheet approach.

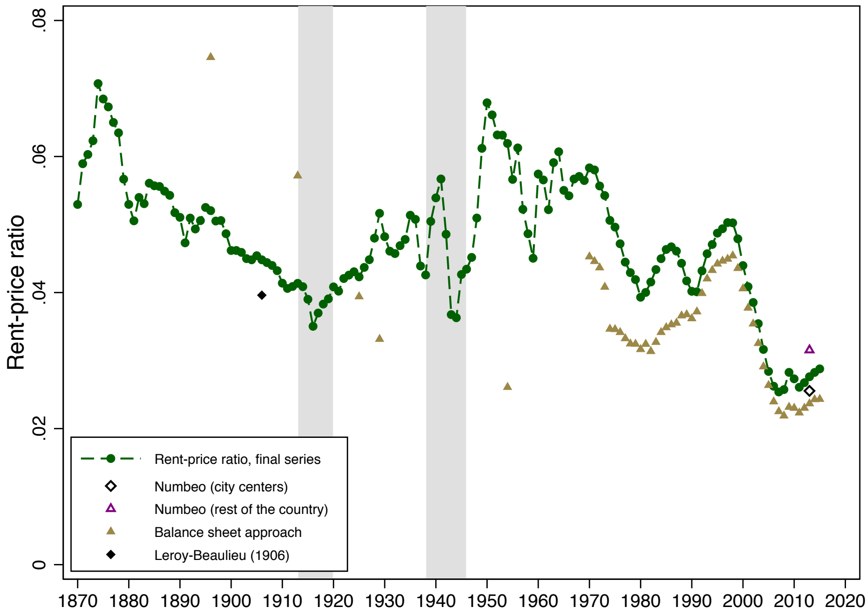
We also construct an independent estimate of the rent-price ratio in Finland in 1920 using data on total housing value (Statistics Finland, 1920) and total expenditure on rents (Hjerppe, 1989), adjusted down by one-third to account for running costs and depreciation. Figure 1.B.4 shows that this estimate is significantly below the long-run rent price ratio in 1920. Similarly to the case of Spain, the discrepancy between the rent-price approach and alternative estimates may reflect difficulties of the Finnish statistical office to construct a rent index after the introduction of wartime rent controls. Rent controls were introduced during WW2 and were only abolished under the *Tenancy Act* of 1961 (Whitehead, 2012). While this period of deregulation was rather short-lived—rent regulation was re-introduced in 1968 and parts of the private rental market were subject to rent regulation until the mid-1990s—the downward trend of the long-run rent-price ratio appears particularly remarkable. In other words, the data suggest that rents during the period of deregulation increased significantly less than house prices. To the best of our knowledge, no quantitative or qualitative evidence exists supporting such a pronounced fall in the rent-price ratio during the first half of the 1960s. We therefore conjecture that the rent index suffers from a downward bias during the period of wartime rent regulation and immediately thereafter. To mitigate this bias, we adjust the gross growth rate in rents between WW2 and 1965 up by a constant factor calibrated so that the adjusted long-run rent-price ratio concords with the independent estimate in 1920, which is a factor of 1.1. Figure 1.B.4 displays the resulting adjusted long-run rent-price ratio.

France

For 2013, the MSCI (2016) reports the rent-price ratio for French residential real estate of 0.028. Applying the rent-price approach to this benchmark gives us the long-run net rent-price ratio series depicted as green circles in in Figure 1.B.5, which are the estimates used in this paper.

We obtain several scattered independent estimates of rent-price ratios in France since 1870. First, we calculate rent-price ratios using the balance-sheet approach, based on the data on total housing value (Piketty and Zucman, 2014) and total ex-

Figure 1.B.5. France: plausibility of rent-price ratio



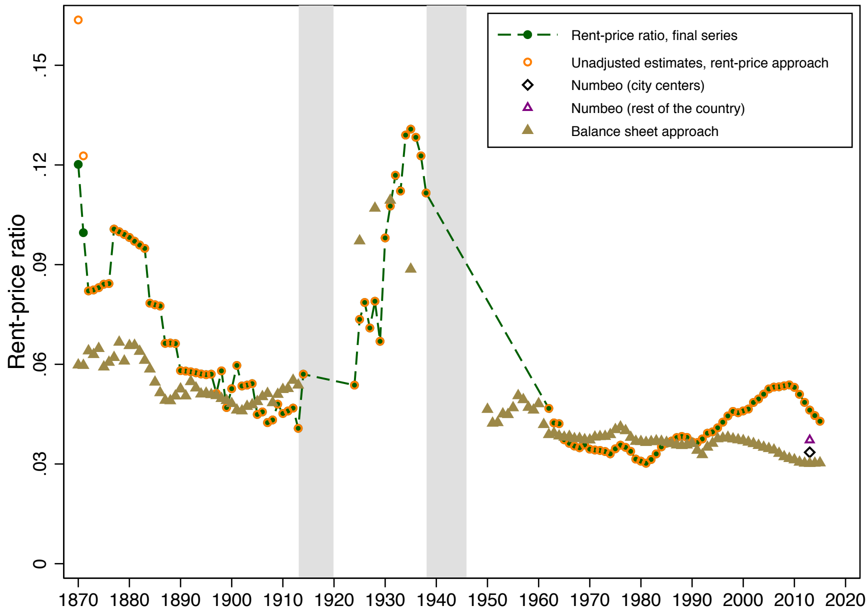
penditure on rents (Villa, 1994; Statistics France, 2016b) net of running costs and depreciation (Piketty and Zucman, 2014; Statistics France, 2016a; Statistics France, 2016b). These estimates are in line with those using the rent-price approach, even though the balance-sheet approach rental yield estimates for 1900–1920 are somewhat higher, and for 1920–1960 somewhat lower. Second, Numbeo.com estimates of modern-day rent-price ratios are in line with the IPD benchmark.

A few additional scattered estimates on housing returns for the pre-WW2 period are available. For 1903, Haynie (1903) reports an average gross rental yield for Paris of about 4 percent. For 1906, Leroy-Beaulieu (1906) estimates a gross rental yield for Paris of 6.36 percent, ranging from 5.13 percent in the 16th arrondissement to 7.76 percent in the 20th arrondissement. Simonnet, Gallais-Hamonno, and Arbulu (1998) state that the gross rent of residential properties purchased by the property investment fund *La Fourmi Immobiliere* amounted to about 6 to 7 percent of property value between 1899 and 1913. These estimates are generally comparable with an average annual net rental yield of about 5 percent for 1914–1938 for the final series used in this paper.

Germany

For 2013, the MSCI (2016) reports the rent-price ratio for German residential real estate of 0.047. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.6. We make one adjustment to these series to correct for possible mismea-

Figure 1.B.6. Germany: plausibility of rent-price ratio



surement of rental growth in the early 1870s (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.6—used in this paper.

We obtain three independent estimates of historical rent-price ratios in Germany. First, Numbeo.com estimates of modern-day rent-price ratios are broadly in line with the rent-price approach. Second, we calculate the balance sheet approach estimates for benchmark years based on data on total housing value and total expenditure on rents. The housing wealth series combines the data in Piketty and Zucman (2014), and various issues of *Statistik der Einheitswerte*. For the pre-WW1 period, we scale up the value of structures reported in Piketty and Zucman (2014) to obtain a proxy for total housing wealth. The rental expenditure data are from OECD (2016b) and Statistics Germany (2013) for the modern period, and (Hoffmann, 1965) for the period before WW2. Throughout we assume around one-third of gross rent is spent on costs and depreciation to obtain a proxy for net rental expenditure.

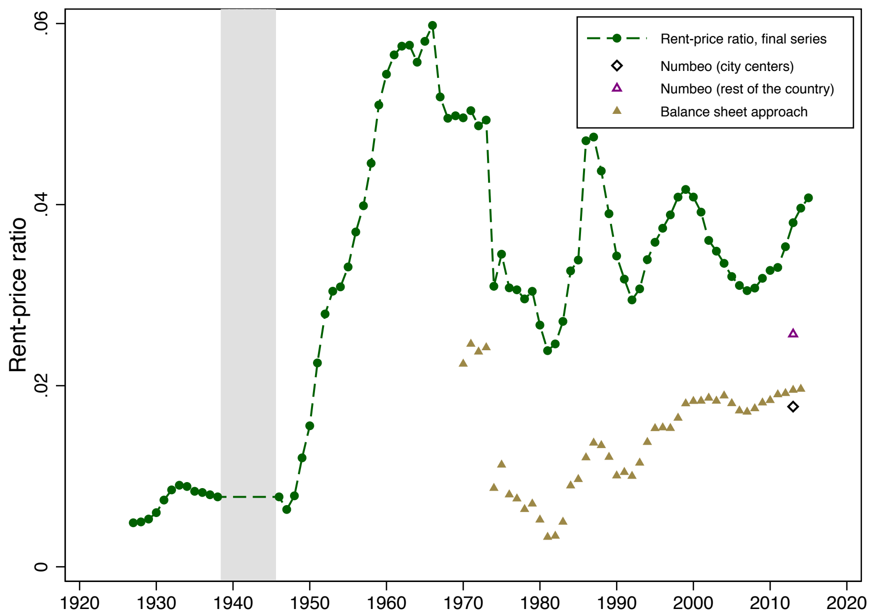
Figure 1.B.6 shows that the balance sheet approach estimates confirm the general level and historical time trend of the rent-price ratio: rents were high in the interwar period, and comparatively lower before WW1 and after WW2. The modern-day balance sheet approach estimates are somewhat below those in our final series, but within a reasonable margin of error, given the uncertainty in estimating housing wealth, imputed rents, running costs and depreciation. For the years 1870–1871, however, the balance sheet approach estimates of rental yield are relatively stable, whereas those using the rent-price approach are markedly high. It is likely that the rental index underestimated the rental growth during years 1870–1871,

when house prices grew sharply. However, the balance sheet approach net yield estimate is in itself highly uncertain, as housing wealth data may have been smoothed over time, and there is little data on the value of land underlying dwellings. We therefore adjust the rental yield down to the average of the rent-price figures, and an alternative rental yield series that extrapolates the growth of rents back using the balance sheet approach. This results in the green dots, our final series for 1870–1871, that suggests that rental yields fell during those years, but probably by less than suggested by the raw unadjusted series.

Finally, one additional series on housing returns is available for the pre-WW2 period. For 1870–1913, Tilly (1986) reports housing returns for Germany and Berlin. Average annual real net returns according to Tilly (1986) amount to about 8 percent—a figure similar to the circa 10 percent p.a. average annual real return calculated using the adjusted rent and house price data.

Italy

Figure 1.B.7. Italy: plausibility of rent-price ratio

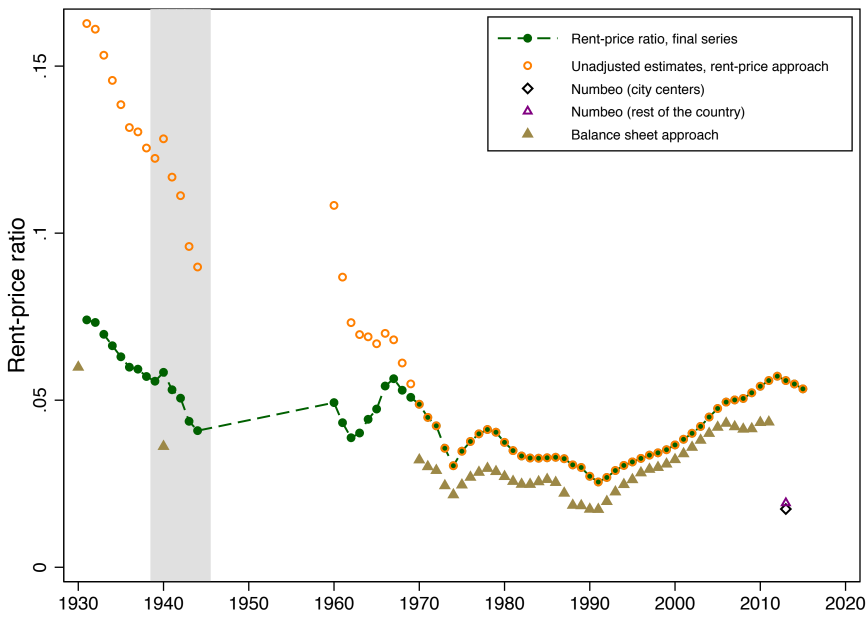


For 2013, the MSCI (2016) reports the rent-price ratio for Italian residential real estate of 0.038. Applying the rent-price approach to this benchmark gives us the long-run net rent-price ratio series depicted as green circles in in Figure 1.B.7, which are the estimates used in this paper.

To gauge the plausibility of historical rent-price ratios, we construct the balance-sheet approach rental yields as total rental expenditure net or running costs and depreciation, in proportion to total housing wealth (Piketty and Zucman, 2014; Istat, 2016). These are somewhat lower than the rent-price approach estimate, but confirm the general trend in the rent-price ratio from the 1970s onwards. Finally, Numbeo.com estimates of modern-day rent-price ratios are similar to the rent-price and balance sheet approach.

Japan

Figure 1.B.8. Japan: plausibility of rent-price ratio



For 2013, the MSCI (2016) reports the rent-price ratio for Japanese residential real estate of 0.056. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.8. We make one adjustment to these series to correct for possible mismeasurement of rental growth in the 1960s (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.8—used in this paper.

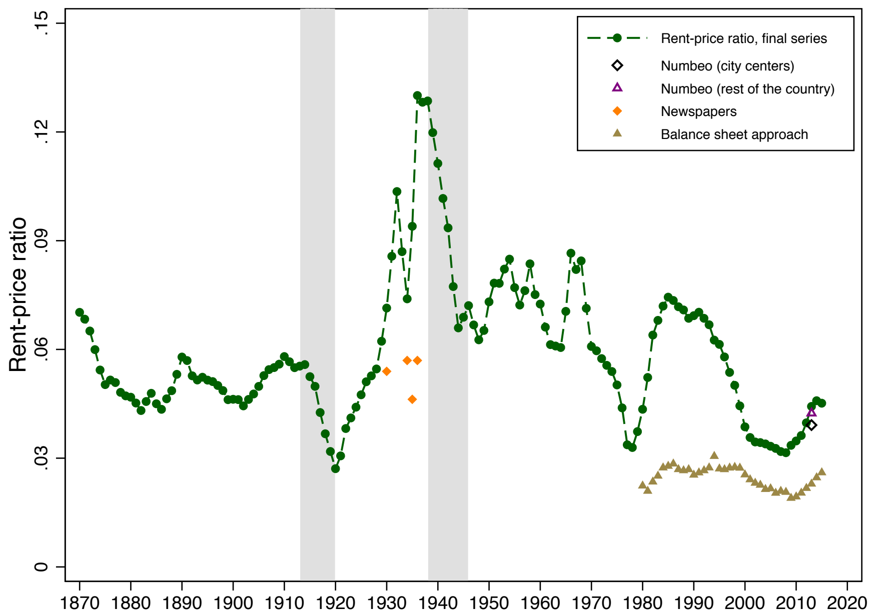
We obtain two independent estimates for rent-price ratios in Japan. First, we calculate rent-price ratios for benchmark years (1930, 1940, 1970–2011) based on data on total housing value (Goldsmith, 1985a; Piketty and Zucman, 2014) and total expenditure on rents (Shinohara, 1967; Cabinet Office. Government of Japan, 2012). To proxy the net rent-price ratio, we assume around one-third of gross rent is spent on running costs and depreciation. The resulting estimates are consistent with the long-run rent-price ratio for the period 1970–2011 (Figure 1.B.8). Yet, for 1930 and 1940 the estimates are much lower than those using the rent-price approach. This suggests that the rent index may have underestimated rent growth between

1940 and 1970, thus inflating the historical rental yield estimates. Indeed, the unadjusted series imply that the rent-price ratio fell dramatically during the 1970s, a trend not mirrored in any subsequent period, or in the balance-sheet approach data. To this end, we conjecture that the rental index understated the growth in rents by a factor of two during the 1960s. The resulting adjusted rent-price ratio (green circles) is then consistent with the historical estimates using the balance sheet approach.

Second, estimates of modern-day rent-price ratios from Numbeo.com are are somewhat below both the rent-price approach and balance-sheet approach estimates for the 2010s.

Netherlands

Figure 1.B.9. Netherlands: plausibility of rent-price ratio



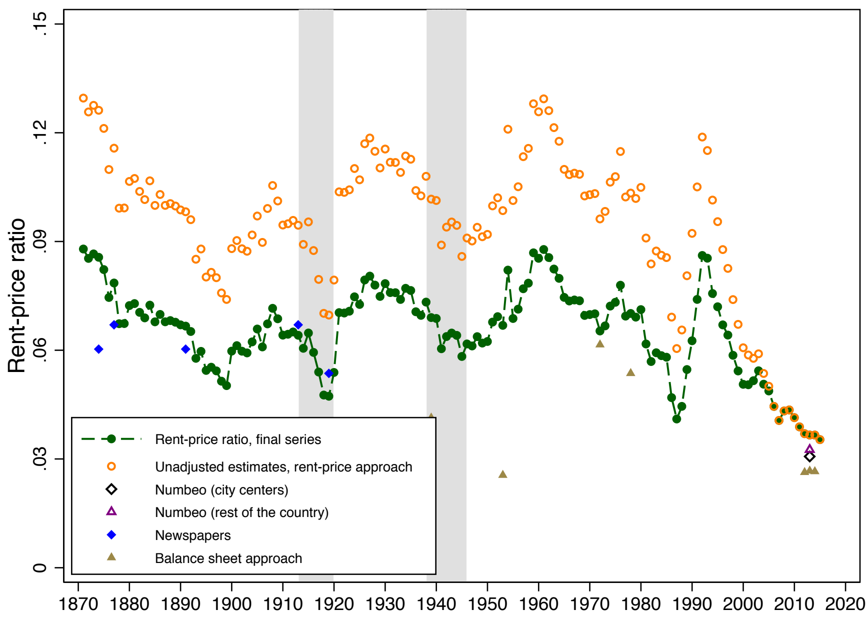
For 2013, the MSCI (2016) reports the rent-price ratio for Dutch residential real estate of 0.044. Applying the rent-price approach to this benchmark gives us the long-run net rent-price ratio series depicted as green circles in in Figure 1.B.9, which are the estimates used in this paper.

We obtain two independent estimates for rent-price ratios in the Netherlands. First, we calculate the rent-price ratio using the balance sheet approach, based on estimates of rental expenditure from OECD (2016b), and housing wealth estimated from non-financial balance sheet data in OECD (2016c) and Groote, Albers, and De Jong (1996) (brown trianges in Figure 1.B.9). We assume one-third of gross rental is spent on running costs and depreciation. The yields confirm the general trend in our benchmark series, although their levels are somewhat lower. It is worth noting that the estimates of housing wealth and running costs for the Netherlands are highly uncertain, hence we do not put too much weight on the level of the balance-sheet approach yields.

Second, a number of newspaper advertisements and articles in the mid-1930s report rent-price ratio levels of 0.07–0.09, which we conjecture are around 0.05 - 0.06 in net terms, once running costs and depreciation are taken out (Nieuwe Tilburgsche Courant, 1934; Limburgsch Dagblaad, 1935; Nieuwe Tilburgsche Courant, 1936). These are somewhat lower than our baseline series, but similar to the levels observed in the early 1930s, with the remaining margin of error easily attributed to location specificity (the advertisements are for city-center properties, with the correspondingly lower yiedls). More generally, residential real estate was perceived as a highly profitable investment throughout the decade (De Telegraaf, 1939). Finally, estimates of the rent-price ratio based on data from Numbeo.com are almost identical to our baseline IPD benchmark (MSCI, 2016).

Norway

Figure 1.B.10. Norway: plausibility of rent-price ratio



For 2013, the MSCI (2016) reports the rent-price ratio for Norwegian residential real estate of 0.037. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.10. We make one adjustment to these series to bring the estimates in line with alternative historical sources (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.10—used in this paper.

We obtain several scattered independent estimates of rent-price ratios in Norway since 1871. First, we calculate rent-price ratios for benchmark years using the balance-sheet approach, based on data on total housing value (Goldsmith, 1985a; OECD, 2016c) and total expenditure on rents (Statistics Norway, 1954; Statistics Norway, 2014; OECD, 2016b), and assuming one-third of gross rent is consumed by running costs and depreciation expenses to estimate the net rental yield. Note that for the historical expenditure series, we estimate rents as 80% of total housing expenditure, a proportion consistent with modern-day Norwegian data, and historical data for the US. We also collect scattered data from advertisements for Oslo residential real estate in *Aftenposten*, one of Norway's largest newspapers, with the gross advertised yield again adjusted down by one-third to proxy the net figure.

Both these sets of estimates confirm the general long-run trend in the rent-price ratio. The long-run rent-price ratio was essentially stable up until the early 2000s, with increases in early 20th century and late 1960s reversed by falls in WW1 and the 1980s, and is currently at a historical low. However the long-run level of the ratio is generally lower than the estimates using the rent-price approach (orange diamonds): around 6%–8% rather than 8%–12%, and this divergence is already apparent in the late 1970s. Based on this, we stipulate that the rental index during late 1990s and early 2000s—a period when house prices increased substantially—understated the growth of rents relative to prices, leading the rent-price approach to overstate the historical rental yields. To correct for this presumed bias, we adjust the growth in rents up by a factor of 1.5 for the years 1990 to 2005. The resulting adjusted rent-price ratio (green circles) is in line with the historical estimates both in terms of levels and trend.

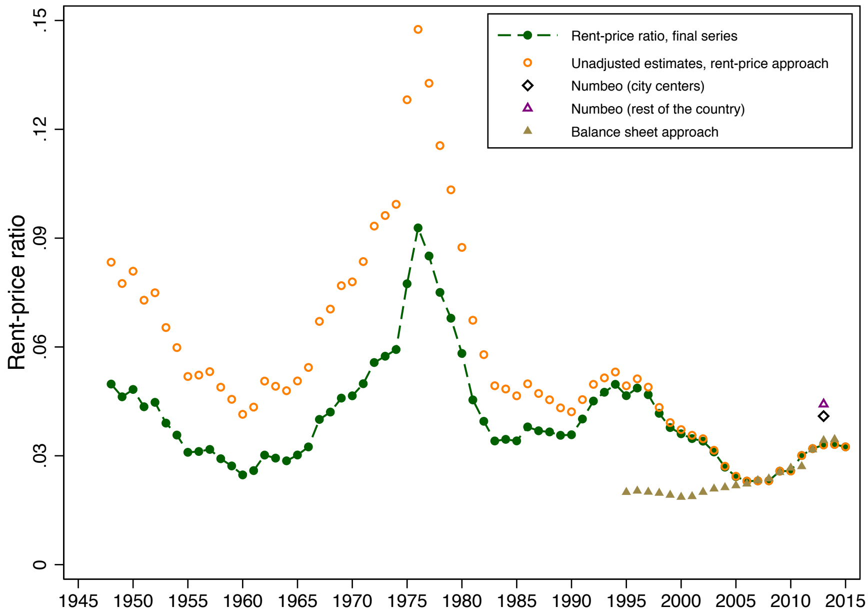
Lastly, estimates of the rent-price ratio based on data from www.Numbeo.com are in line with our baseline IPD benchmark (MSCI, 2016).

Portugal

For 2013, the MSCI (2016) reports the rent-price ratio for Portuguese residential real estate of 0.033. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.11. We make one adjustment to these series to correct for potential biases arising from rent mismeasurement during the prolonged period of rent controls in the last quarter of the 20th century (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.11—used in this paper.

We obtain several scattered independent estimates of rent-price ratios in Portugal. First, estimates of the rent-price ratio based on data from www.Numbeo.com are slightly above, but broadly in line with our baseline IPD benchmark (MSCI, 2016).

Figure 1.B.11. Portugal: plausibility of rent-price ratio

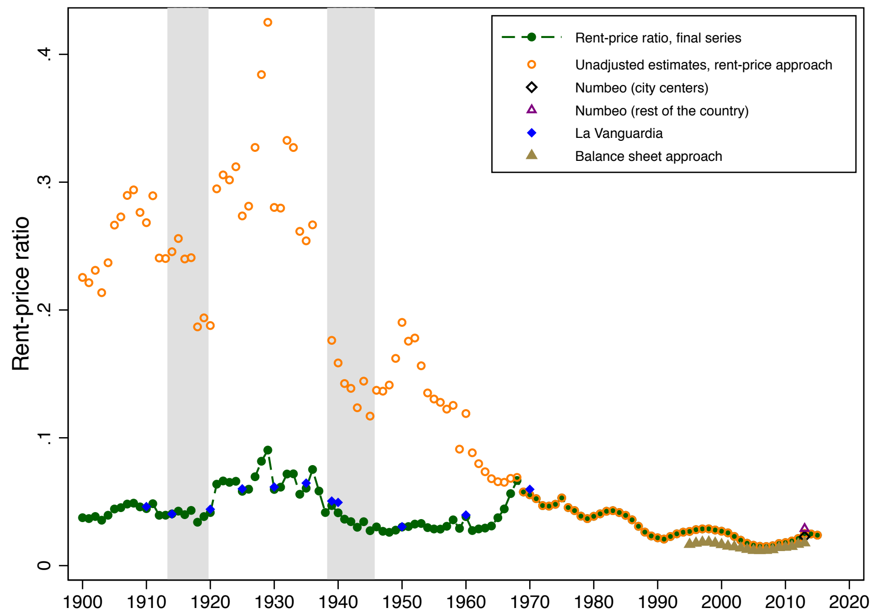


Second, we compute the rental yield using the balance-sheet approach, based on data on total rental expenditure (OECD, 2016b) and total housing wealth (Cardoso, Farinha, and Lameira, 2008), scaled down one-third to adjust for running costs and depreciation. These are almost identical to the rent-price approach for the recent years, but diverge somewhat in the late 1990s. More generally, the historical growth in rents relative to house prices in Portugal may have been understated due to the imposition of rent controls in 1974, which remained in place in various forms until well into the 2000s. This seems likely given the high levels of the unadjusted rent-price approach yields in the 1970s and early 1980s (orange circles in Figure 1.B.11). Unfortunately, no alternative historical estimates of the rent-price ratio before 1995 are available for Portugal. Instead, we stipulate that the rent-price ratio in the 1940s and 50s, before the reported high rent inflation of the 1960s (Cardoso, 1983) and the subsequent rent controls, was at levels similar to the 1980s and 1990s. To achieve that, we adjust rental growth up by a factor of 1.2 for years 1974–2005; the period for which rent controls were in place.

The resulting adjusted long-run rent-price ratio (green circles in Figure 1.B.11) concords with the narrative evidence on house prices and rent developments in Portugal. Real house prices in Portugal rose after the end of WW2 until the Carnation Revolution in 1974. After a brief but substantial house price recession after the revolution, real house prices embarked on a steep incline (Azevedo, 2016). By contrast, real rents remained broadly stable between 1948 and the mid-1960s as well as after 1990 but exhibit a pronounced boom and bust pattern between the mid-1960s and the mid-1980s. According to Cardoso (1983), the rapid growth of inflation-adjusted rents between the mid-1960s and the mid-1970s was the result of both rising construction costs and high inflation expectations. In 1974, new rent legislation provided for a rent freeze on existing contracts. Rent increases were also regulated between tenancies but unregulated for new construction. These regulations resulted in lower rent growth rates and rents considerably lagging behind inflation (Cardoso, 1983), and a consequent fall in the rent-price ratio.

Spain

Figure 1.B.12. Spain: plausibility of rent-price ratio



For 2013, the MSCI (2016) reports the rent-price ratio for Spanish residential real estate of 0.025. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in in Figure 1.B.12. We make one adjustment to these series to correct for possible mismeasurement of rental growth during the rent controls imposed in the early-to-mid 20th century (see below for details). This gives us the final adjusted rent-price ratio series—the green-circled line in Figure 1.B.12—used in this paper.

We obtain several scattered independent estimates of rent-price ratios in Spain. First, estimates of the rent-price ratio based on data from www.Numbeo.com are almost identical to our baseline IPD benchmark (MSCI, 2016). Second, we construct net rent-price ratios using the balance sheet approach, as total rental expenditure (OECD, 2016b) less running costs and depreciation (assumed to be one-third of gross rent), in relation to housing wealth (Artola Blanco, Bauluz, and Martínez-

Toledano, 2017). These are slightly below but broadly in line with the rent-price approach for the overlapping years.

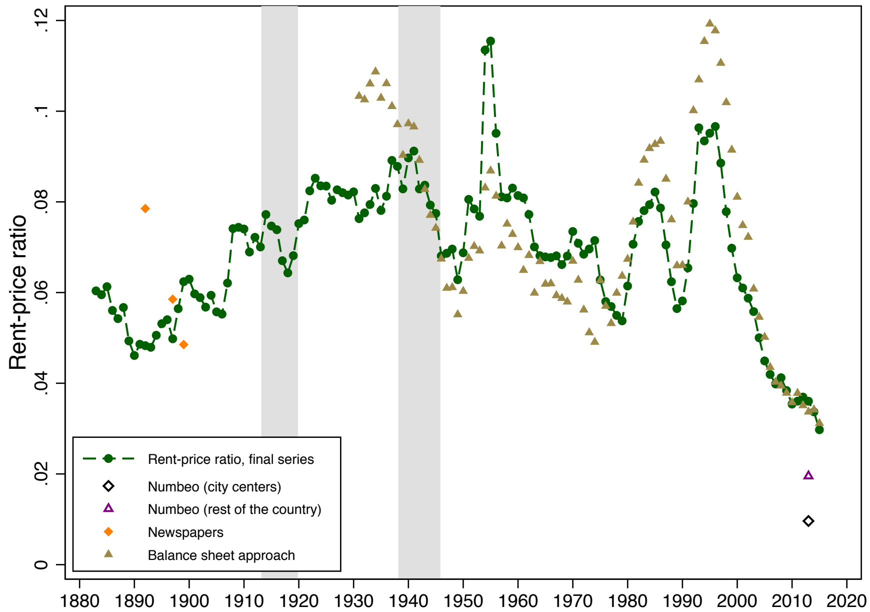
Finally, we collected scattered data on rent-price ratios from advertisements for Barcelona residential real estate in *La Vanguardia* for benchmark years (1910, 1914, 1920, 1925, 1930, 1935, 1940, 1950, 1960, 1970). For each of the benchmark years, we construct an average rent-price ratio based on between 25 and 46 advertisements. The gross ratios in the advertisements are adjusted down to exclude running costs and depreciation, calibrated at 2% p.a., around one-third of the advertized yields. Figure 1.B.12 shows that the newspaper estimates are significantly below the rent-price ratio for the benchmark years between 1910 and 1960. Yet it also suggests that rent-price ratios were generally higher before the mid-1950s. Similarly to Finland, this trajectory may reflect difficulties of the Spanish statistical office to construct a rent index after the introduction of rent freezes in the 1930s and during the years of strong rent regulation after WW2. While the rent freeze was lifted in 1945, these regulations remained effective until the mid-1960s. Specifically, the data suggest that rents between the end of WW2 and the mid-1960s increased substantially less than house prices. To the best of our knowledge, no quantitative or qualitative evidence exists supporting such a pronounced fall in the rent-price ratio in the immediate post-WW2 years or a generally higher level of rental yields prior to the 1960s. To mitigate this bias, we adjust the growth rate in rents between 1910 and 1960 so that the adjusted long-run rent-price ratio concords with the independent estimates obtained from *La Vanguardia*. Figure 1.B.12 displays the resulting adjusted long-run rent-price ratio (green circles), which is the final series we use in this paper.

Sweden

For 2013, the MSCI (2016) reports the rent-price ratio for Swedish residential real estate of 0.036. Applying the rent-price approach to this benchmark gives us the long-run net rent-price ratio series depicted as green circles in in Figure 1.B.13, which are the estimates used in this paper.

We obtain three independent estimates of rent-price ratios for Sweden. First, we compute net rental yields based on the balance-sheet approach as total rental expenditure less running costs and depreciation, as a share of housing wealth, drawing on a variety of sources. The modern-day rental expenditure data are obtained from OECD (2016b), and further data back to 1969 were provided by Birgitta Magnusson Wärmark at Statistics Sweden. These are extrapolated back to 1931 using data on total housing expenditure from Dahlman and Klevmarken (1971). The data on running costs are a weighted average of total repairs of dwellings (data provided by Jonas Zeed at Statistics Sweden), and maintenance costs on rentals reported by (OECD, 2016b) scaled up to capture owner-occupied dwellings. Data on depreciation were provided by Jonas Zeed at Statistics Sweden, and were extrapolated back using dwellings depreciation in Edvinsson (2016). Before 1995, running costs are

Figure 1.B.13. Sweden: plausibility of rent-price ratio



assumed to have evolved in line with depreciation. The long-run housing wealth data are sourced from Waldenström (2017). Both the level and the time trend in the resulting long-run rent-price ratio are in line with the historical balance-sheet approach estimates.

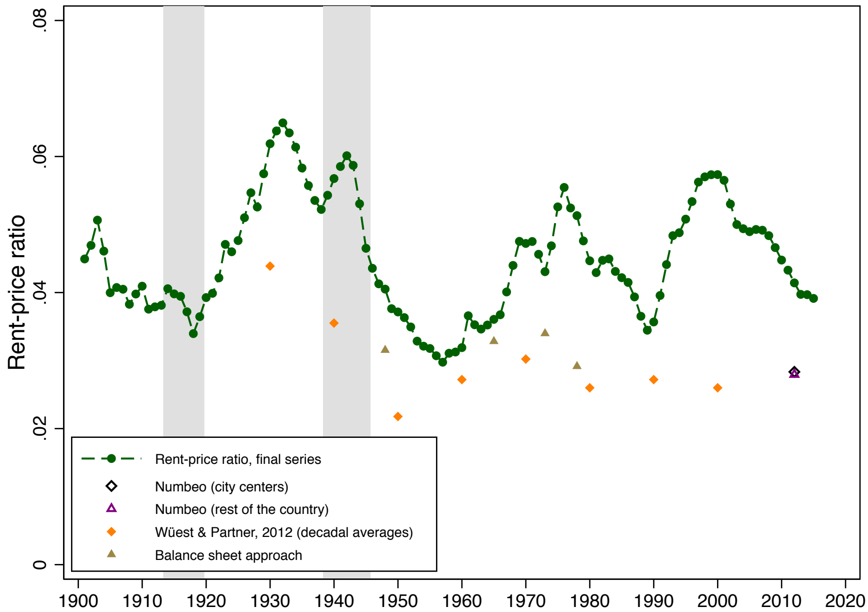
Second, the rent-price ratio in the late 19th / early 20th century is in line with those reported in several newspaper advertisements and articles. According to these sources, gross rent-price ratios were in the range of 0.07 to 0.1, and residential real estate was perceived as highly profitable investment (Dagens Nyheter, 1892; Dagens Nyheter, 1897; Dagens Nyheter, 1899). Given that running costs and depreciation amounted to around 2% p.a. of property value in Sweden during the period 1930–2015, this leads us to conjecture that net rent-price ratios were around 0.05–0.08, in line with our estimates.

Finally, estimates of modern-day rent-price ratios from Numbeo.com are somewhat below both our benchmark ratio and the balance sheet approach. However these are not based on a representative or matched sample of properties for sale and for rent, and are therefore less reliable than the alternative estimates.

Switzerland

For 2013, the MSCI (2016) reports the rent-price ratio for Swiss residential real estate of 0.040. Applying the rent-price approach to this benchmark gives us the long-run net rent-price ratio series depicted as green circles in in Figure 1.B.14, which are the estimates used in this paper.

Figure 1.B.14. Switzerland: plausibility of rent-price ratio



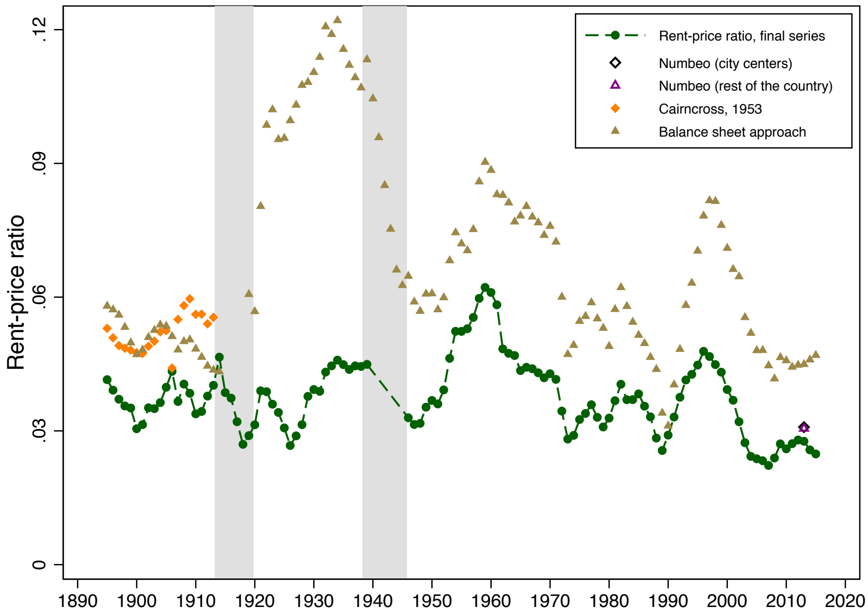
To check the plausibility of the long-run rent-price ratio, we obtain four independent estimates. First, Real (1950) reports real returns on residential real estate in Zurich of 6 percent in 1927 and 7.3 percent in 1933. These data are—by and large—in line with the estimates of housing returns constructed by merging the indices of house prices and rents. Second, Wüest and Partner (2012) estimate 10-year averages of real rental yields in Switzerland for 1920–2000. Assuming around one-third of gross rent goes to running costs and depreciation, the resulting net rental yield estiamtes are broadly consistent with the long-run rent-price ratio (Figure 1.B.14), taking into account the various estimation uncertainties. For the post-WW2 period, we calculate rent-price ratios using the balance sheet approach for benchmark years (1948, 1965, 1973, 1978) drawing on data on housing wealth from Goldsmith (1985a), rental expenditure from Statistics Switzerland (2014), and assuming one-third of gross rent is taken up by runnign costs and depreciation. Again, the resulting estimates are broadly consistent with the long-run rent-price ratio (Figure 1.B.14).

Finally, estimates of rent-price ratios based on data from Numbeo.com are somewhat below, but within a reasonable error margin of the MSCI (2016) benchmark ratio.

United Kingdom

For 2013, the MSCI (2016) reports the rent-price ratio for U.K. residential real estate of 0.032. Applying the rent-price approach to this benchmark gives us the long-run

Figure 1.B.15. United Kingdom: plausibility of rent-price ratio



net rent-price ratio series depicted as green circles in in Figure 1.B.15, which are the estimates used in this paper. Please note that for years 1947–1955, no rental index data were available, and we extrapolated the rent-price ratio series using the growth in the "balance sheet approach" measure, benchmarking against rental index values in 1946 and 1956.[58]

We construct several alternative estimates of the rent-price ratio for the period going back to 1900. First, we construct the net rental yield based on the balance-sheet approach using data on total rental expenditure less running costs and depreciation, in proportion to housing wealth, based on a variety of sources. For rents, we rely on historical series of housing and rental expenditure from Mitchell (1988), Sefton and Weale (1995) and Piketty and Zucman (2014), combined with recent Office for National Statistics (ONS) data, and historical data from the ONS shared with us by Amanda Bell. Estimates of costs and depreciation are available from the UK National Accounts, and housing wealth is taken from Piketty and Zucman (2014). It is worth noting that the estimates of rental expenditure for the UK are subject to large uncertainty: the ONS updated the methodology for rent imputation in 2016, resulting in large upward revisions to historical imputed rent estimates (by as large as a factor of three). It is possible that some of the historical data are subject to similar uncertainties, which helps explain why the rental yield levels using the bal-

58. We assume that the 1956 index value is correct, but correct the 1946 rental index value for possible biases arising from the wartime rent controls, such that the trend in the rent-price ratios matches that in the balance sheet approach measure, and the 1956 rent-price approach estimate.

ance sheet approach are so much higher than the extrapolated rent-price ratio, even though the time trend is similar.

Some additional scattered data on rent-price ratios are available for the pre-WW2 period. For England, Cairncross (1975) reports an average gross rent-price ratio of 0.068 between 1895 and 1913, or around 0.05 in net terms. Offer (1981) estimates slightly higher rent-price ratios for selected years between 1892 and 1913 for occupied leasehold dwellings in London. As Figure 1.B.15 shows, these data are slightly higher, but broadly consistent with the our long-run rent-price ratio estimates (an average of 0.037 during 1900–1913). Tarbuck (1938) states that high-quality freehold houses were valued at 25 to 16 years purchase and lower quality freehold houses at 14 to 11 years purchase in the 1930s, again broadly consistent with our estimates.

Overall, these estimates suggest that our rental yields for the UK are somewhat conservative, but fit the time pattern and broad levels found in the alternative historical sources.

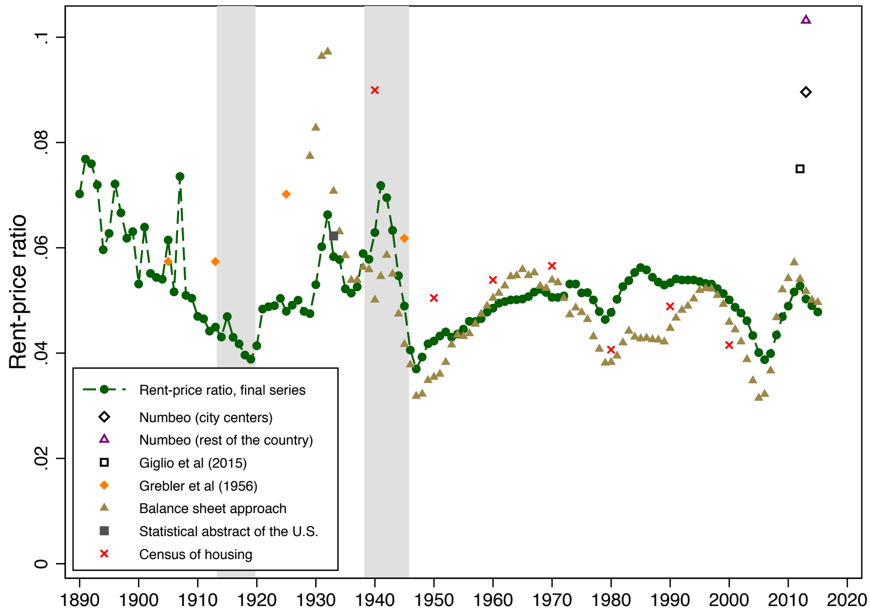
Concerning the modern period, estimates of the rent-price ratio based on data from www.Numbeo.com are very similar to the MSCI (2016) benchmark. Additionally, Bracke (2015) estimates a gross rental yield of 0.05 on central London properties over the period 2006–2012, based on a matched micro-level dataset of around 2000 properties. Again, these estimates are consistent with our data.

United States

For 2014, the MSCI (2016) reports the rent-price ratio for U.S. residential real estate of 0.049. Applying the rent-price approach to this benchmark gives us the long-run net rent-price ratio series depicted as green circles in in Figure 1.B.16, which are the estimates used in this paper.

We obtain independent estimates of U.S. rent-price ratios from five additional sources. First, decadal averages of gross price-rent ratios are available for 1899–1938 from Grebler, Blank, and Winnick (1956) ranging between 10.4 and 12.6. Second, estimates of gross rents paid and home values are available from various issues of the U.S. Census and Statistical Abstract, published by U.S. Census Bureau (1942) and U.S. Census Bureau (2013). Once adjusted for estimates of running costs and depreciation, the estimates from these sources are similar to the price-rent ratios resulting from merging the indices of house prices and rents (see Figure 1.B.16). Third, we calculate the rent-price ratio using the balance sheet approach, as total rental expenditure less housing running costs—estimated as 2/3 of total housing intermediate consumption—in proportion to total housing value, using expenditure data from Bureau of Economic Analysis (2014) and housing wealth estimates in Saez and Zucman (2016). Reassuringly, the resulting estimates are very close to the long-run rent-price ratio. Estimates of the rent-price ratio for 2012 are also available from the real estate portal Trulia, as used by Giglio, Maggiori, and Stroebel (2015). The

Figure 1.B.16. United States: plausibility of rent-price ratio



resulting net rent-price ratio of 0.075 is higher than the figures from MSCI (2016) and the balance sheet approach. This may be because the Trulia ratios are not market cap weighted, and may overweigh the high-yield low-housing-wealth areas outside of cities. Alternatively, the MSCI (2016) IPD ratio could understate the rental yield because investor portfolios tend to be concentrated in cities. To be consistent with the balance sheet approach and to remain conservative, we use the IPD ratio as our benchmark.

Finally, estimates of the rent-price ratio based on data from www.Numbeo.com are higher than our benchmark estimate and similar to the Trulia transaction-level data. As with the Trulia data, these are not market-capitalization weighted, which may bias the rental yield estimates upwards. Given the similarity to the balance-sheet approach yields and the historical estimates from Grebler, Blank, and Winnick (1956), the rent-price approach estimates stemming from the MSCI (2016) benchmark should provide the most accurate picture of the historical rental returns on housing in the US. Still, given the higher alternative benchmark yield estimates of Trulia and Numbeo.com, our housing return series for the US should be viewed as conservative compared to other possible alternatives.

1.B.3 Rent indices: methodology

Rent indices measure the change in 'pure' rents for primary residences, i.e., net of house furnishings, maintenance costs, and utilities. For modern rent indices included in CPIs, data are usually collected by statistical offices through surveys of housing authorities, landlords, households, or real estate agents (International Labour Organization, International Monetary Fund, Organization for Economic Co-operation and Development, Statistical Office of the European Communities, United Nations, et al., 2004).

Rental units are heterogeneous goods.[59] Consequently, there are several main challenges involved when constructing consistent long-run rent indices. First, rent indices may be national or cover several cities or regions. Second, rent indices may cover different housing types ranging from high to low value housing, from new to existing dwellings. Third, rental leases are normally agreed to over longer periods of time. Hence, current rental payments may not reflect the current *market rent* but the *contract rent*, i.e., the rent paid by the renter in the first period after the rental contract has been negotiated.[60] Fourth, if the quality of rental units improve over time, a simple mean or median of observed rents can be upwardly biased. These issues are similar to those when constructing house price indices and the same standard approaches can be applied to adjust for quality and composition changes. For a survey of the different approaches, the reader is referred Knoll, Schularick, and Steger (2017). Yet, as can be seen from the data description that follows, these index construction methods commonly used for house price indices have less often been applied to rents.

Another important question when it comes to rent indices is the treatment of subsidized and controlled rents. Rental units may be private or government owned and hence be subject to different levels of rent controls or subsidies. Since these regulations may apply to a substantial share of the rental market, rent indices typically cover also subsidized and controlled rents (International Labour Organization et al., 2004).[61] It is worth noting that not properly controlling for substantial changes in rent regulation may result in a mis-measurement of rent growth rates. More specifically, if the share of the rental market subject to these regulations sud-

59. Compared to owner-occupied houses, Gordon and Goethem (2007) argue that rental units are, however, less heterogeneous in size at any given time and more homogenous over time. The authors provide also scattered evidence for the U.S. that rental units experience quality change along fewer dimensions than owner-occupied units.

60. Typically, in times of low or moderate general inflation, the market rent will be higher than the contract rent. Yet, the introduction of rent controls or a temporary strong increase in the supply of rental units may result in the market rent being lower than the contract rent (Shimizu, Imai, and Diewert, 2015).

61. Exceptions include, for example, the Canadian rent index where subsidized dwellings are excluded (Statistics Canada, 2015).

denly increases—e.g., during wars and in the immediate post-war years—-the rent index can be downwardly biased.[62]

An additional challenge when constructing rent indices is the treatment of owner-occupied housing. Since a significant share of households in advanced economies are owner-occupants, rent indices typically cover changes in the cost of shelter for both renters and owner-occupiers.[63] The cost for owner-occupied shelter is an estimate of the implicit rent that owner-occupants would have to pay if they were renting their dwellings. Different approaches to estimate the change in implicit rents exist, each with advantages and disadvantages. Most statistical offices rely on the *rental equivalent approach*.[64] The resulting rent index is based on an estimate of how much owner-occupiers would have to pay to rent their dwellings or would earn from renting their home in a competitive market. Data either come from surveys asking owner-occupiers to estimate the units' potential rent or are based on matching owner-occupied units with rented units with similar characteristics.[65] The *user cost approach* assumes that a landlord would charge a rent that at least covers repairs and maintenance, taxes, insurance, and the cost of ownership (i.e., depreciation, mortgage interest, opportunity costs of owning a house). The resulting rent index is a weighted average of the change in the price of these components.[66] The user cost approach is important in its own right (i.e., when the size of the rental market is relatively small, it is not possible to value the services of owner-occupied housing using the *rental equivalence approach*). Nevertheless, the user-cost and rental equivalence approach should, in principle, yield similar results given that capital market theory implies that the price of an asset should equal the discounted value of the flow of income or services (e.g., rents) that it provides over the lifetime of the asset. The *net acquisitions approach* measures the costs associated with the purchase and ongoing ownership of dwellings for own use. Hence it covers the costs of repair and maintenance, taxes, insurances and the change in the cost of the net acquisition of the dwelling, i.e., the change in the total market value (OECD, 2002; International

62. For example, this has been the case for the Australia CPI rent index after WW2 (see Section 1.B.4).

63. Imputed rents of owner-occupied housing are excluded in Belgium and France. In some countries, two rent indices are reported, one for renter-occupied and one for owner-occupied dwellings (OECD, 2002; International Labour Organization et al., 2004).

64. The *rental equivalent approach* is currently used in the U.S., Japan, Denmark, Germany, the Netherlands, Norway, and Switzerland (OECD, 2002).

65. This approach may result in a bias of unknown size and direction if i) owners' assessment of the rental value of their dwelling is unreliable, ii) if the rental market is small and the rental housing stock is not comparable to the owner-occupied housing stock, and ii) if rents set in rental markets are significantly affected by government regulation since subsidized and controlled rents should not be used in calculating an owners' equivalent rent index (OECD, 2002; International Labour Organization et al., 2004; Diewert, 2009).

66. A (partial) *user cost approach* is currently used in Canada, Finland, Sweden, and the United Kingdom (OECD, 2002).

Labour Organization et al., 2004; Diewert, 2009).[67] If rents of owner-occupants are included in rent indices, the combined rent index is a weighted average of rents for rented and owner-occupied dwellings. Weights are based on the share of owner-occupants and tenants in the respective housing market.

67. Hence, a basic requirement of this method is the existence of a constant- quality house price index. The *net acquisitions approach* is currently used in Australia (OECD, 2002).

1.B.4 Data sources for the rental indices

To construct rent indices reaching back to the late 19th century, we rely on two main sources. First, we use the rent components of the cost of living or consumer price indices published by regional or national statistical offices such as Statistics Sweden (1961) and Statistics Norway (2015). The cost of shelter is a major component of household expenditure. Cost of living (COLIs) or consumer price indices (CPIs) therefore typically include a component for housing. In many advanced economies, the construction of COLIs/CPIs was initiated by governments during WW1 to calculate necessary wage adjustments in times of strongly rising price levels. Hence, most countries' statistical offices started to collect data on rents and calculate rent indices in the early 20th century.[68] The Yearbook of Labor Statistics (International Labour Organization, various years) serves as main repository for these data from national statistical offices. Second, to extend these indices back to the late 19th century, we draw on previous work of economic historians, such as Rees (1961) for the U.S., Lewis and Weber (1965) for the U.K., or Curti (1981) for Switzerland.

Australia

Rent data. Historical data on rents in Australia are available for 1901–2015.

For Australia, there are two principal sources for historical rent data. First, the CPI rent component constructed by the Australian Bureau of Statistics covers the period 1901–2015. This rent index is based on data for urban areas and has historically been published in two versions, the *A* and the *C series*.[69]. For the years the two series overlap, the difference appears negligible (Stapledon, 2012). Since 1961, the CPI rent index is based on rent data for 8 capital cities. The sample of dwellings included is stratified according to location, dwelling type and dwelling size based on data from the most recent *Census of Population and Housing* (Australian Bureau of Statistics, 2011). Rent data are collected from real estate agents and state and territory housing authorities (Australian Bureau of Statistics, 2011).

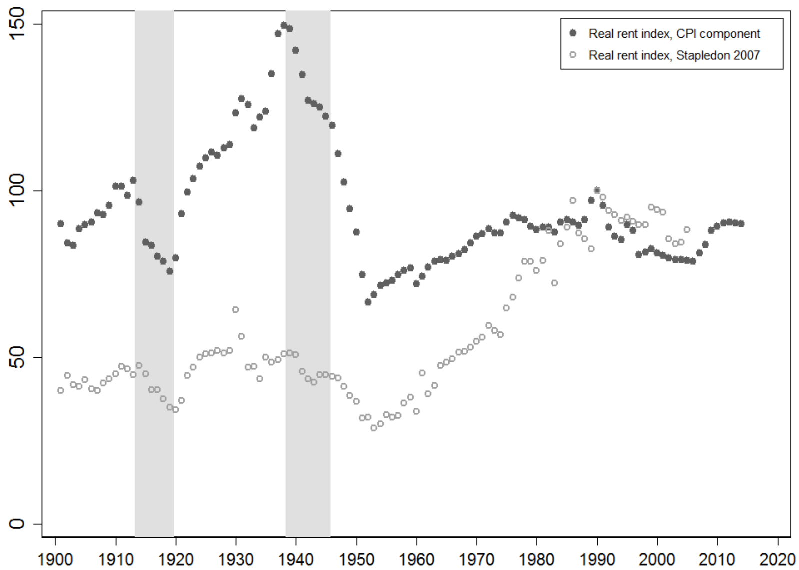
The second source is Stapledon (2007) who presents an index of average rents per dwelling based on census estimates for 1901–2005. The author observes substantial differences between his series and the CPI rent index described above. While for the years prior to WW2, the rent index based on census data and the CPI rent index are highly correlated,[70] the CPI rent index increases much less than the index based on census data during the immediate post-WW2 decades (see Figure 1.B.17). Staple-

68. One exception is Belgium where house rents were only added to the CPI basket in 1989.

69. The *A series* starts in 1901 and refers to average rents of all kinds of dwellings in the 6 capital cities. The series was discontinued in 1938. The *C series* starts in 1920, covers 30 towns (including the 6 capital cities) and is based on rent data for 4- and 5-room houses (Australian Bureau of Statistics, 2011).

70. Correlation coefficient of 0.75.

Figure 1.B.17. Australia: comparison of real rent indices



Note: Indices, 1990=100

don (2007) hypothesizes that this may reflect difficulties of the Australian statistical office to construct a rent index after the introduction of wartime rent controls.

Given this potential bias in the CPI rent index in the post-WW2 period, we rely on the series constructed by Stapledon (2007) for the years 1940–1989 and the CPI rent component before and after.[71]. For the pre-WW2 period, we rely on the *C series* whenever possible as it is based on a more homogeneous dwelling sample and may thus be less affected by shifts in the composition of the sample. The available series are spliced as shown in Table 1.B.3.

The most important limitation of the long-run rent series is the lack of correction for quality changes and sample composition shifts before 1990. As noted above, the latter aspect may be less of a problem for the years 1921–1939 since the index is confined to a specific market segment, i.e., 4- and 5-room dwellings. Note that matching the Australian house price and rent series in terms of geographical coverage has been—by and large—possible. Both series are based on data for capital cities since 1901. Yet, no information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

71. Rent controls were introduced in 1939 and gradually lifted after 1949. According to Stapledon (2007), rent controls affected rent levels well into the 1960s.

Table 1.B.3. Data sources: rent index, Australia

Period	Source	Details
1901–1920	Australian Bureau of Statistics, CPI A series as published in Stapledon (2012)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1921–1939	Australian Bureau of Statistics, CPI C series as published in Stapledon (2012)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: Houses with 4-5 rooms; *Method*: Average rents.
1940–1989	Stapledon (2007)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1990–2015	Australian Bureau of Statistics, CPI series	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Stratification.

Belgium

Rent data. Historical data on rents in Belgium are available for 1890–2015.

The long-run rent index relies on five different sources. First, for the years since 1984, we rely on the CPI rent index constructed by Statistics Belgium.[72] The index covers tenants' rents only, i.e., imputed rents of owner-occupiers are excluded. Second, for 1977–1983, we use the rent index published by the International Labour Organization (2014) which, in turn, is based on data provided by Statistics Belgium. The main characteristics of these two series are summarized in Table 1.B.4.

For earlier periods, data has been drawn from two major historical studies (Buyst, 1994; Segers, 1999) and an unpublished database by Anne Henau.[73] The rent index for seven cities[74] constructed by Segers (1999) for 1890–1920 is based on data from two public institutions for social welfare, the *Burelen van Weldadigheid* and the *Burgerlijke Godshuizen*. The individual city series are constructed as chain indices so as to at least partially account for changes in the underlying sample. The combined index is an unweighted average of the seven city indices. The rent index reported in Buyst (1994) for 1921–1938 is an unweighted average of five city indices[75] combining data drawn from studies by Leeman (1955) and Henau (1991) (see below). The unpublished index constructed by Henau for 1939–1961 covers four cities[76] using records of local Public Welfare Committees (OCMWs).

Three alternative series for the pre-WW2 period are available. Van den Eeckhout and Scholliers (1979) present a rent index for dwellings let by the OCMW in

72. Only in 1989, house rents were added to the CPI basket. Series sent by email, contact person is Erik Vloeberghs, Statistics Belgium.

73. Series sent by email, contact person is Erik Buyst, KU Leuven.

74. These are Antwerp, Brugge, Brussels, Gent, Kortrijk, Leuven, Luik.

75. These are Brussels, Antwerp, Ghent, Leuven, and Luik.

76. These are Leuven, Luik, Ghent, and Antwerp.

Brussels for 1800–1940. Henau (1991), also using records of local OCMWs, constructs rent indices for Leuven, Luik, Ghent, and Antwerp for 1910–1940. Leeman (1955) calculates city indices for a small sample of houses for Brussels, Gent, and Hoei for 1914–1939. As these series, however, are less comprehensive in terms of geographic coverage, we rely on the indices by Segers (1999) and Buyst (1994). The rent indices constructed by Van den Eeckhout and Scholliers (1979), Leeman (1955), Buyst (1994), and Segers (1999) follow a joint, almost identical path for the years they overlap.

The available series are spliced as shown in Table 1.B.4. Since no time series of rents is available for 1961–1977, the two sub-indices (1870–1961 and 1977–2013) are linked using scattered data on rent increases between 1963 and 1982 reported by Van Fulpen (1984).

The resulting index suffers from two weaknesses. The first relates to the lack of correction for quality changes and sample composition shifts. Second, for 1939–161, the series relies on dwellings let by Public Welfare Committees only. It is of course possible that this particular market segment does not perfectly mirror fluctuations in prices of other residential property types. Note further that the matching of the Belgian house price and rent series is imperfect for two reasons. First, the house price index is based on data for the Brussels area prior to 1950. Since the available rent data for the pre-1950 period relies on a rather mall sample, we opted for the

Table 1.B.4. Data sources: rent index, Belgium

Period	Source	Details
1870–1920	Segers (1999)	*Geographic Coverage*: 7 cities ; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1921–1938	Buyst (1994)	*Geographic Coverage*: 5 cities; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1939–1961	Unpublished database by Anne Henau.	*Geographic Coverage*: 4 cities; *Type(s) of Dwellings*: All kinds of dwellings let by Public Welfare Committees; *Method*: Average rents.
1977–1983	International Labour Organization (2014)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: Non-public housing; representative sample of 1,521 apartments and houses of various sizes; *Method*: Average rents.
1984–2013	Statistics Belgium	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: Non-public housing; representative sample of 1,521 apartments and houses of various sizes; *Method*: Average rents.

indices with broader geographic coverage. Second, no information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

Denmark

Rent data. Historical data on rents in Denmark are available for 1870–2015.

For 1870–1926, no rent series for Denmark as a whole exists. We therefore combine three series on rents in Copenhagen to proxy for development of rents in Denmark as a whole. First, for 1870–1911, we rely on an index of average rents for 3 room apartments—which can generally be considered working class or lower middle class dwellings—in Copenhagen (Pedersen, 1930). Second, for 1914–1917, the long-rent index is based on the increase in average rents of 1–8 room houses in Copenhagen as reported in Statistics Copenhagen (1906–1966). Third, for 1918–1926, we rely on the rent component of the cost of living index reported in Statistics Denmark (1925) and Statistics Copenhagen (1906–1966) referring to average rents of 1-5 room houses in Copenhagen.

For 1927–1955, we use the CPI rent index as reported in the Yearbook of Labor Statistics (International Labour Organization, various years) which for the years prior to 1947 is based on average rents in 100 towns and in 200 towns for the years thereafter.

For 1955–1964, to the best of our knowledge, no data on rents for Denmark as a whole are available. we therefore use the increase in average rents of 1–5 room houses in Copenhagen as reported in Statistics Copenhagen (1906–1966) as a proxy for rent increases in Denmark.

For 1965–2015, we rely on the CPI rent index as reported in Statistics Denmark (2003), Statistics Denmark (2015), and the yearbooks of the International Labour Organization (various years). The available series are spliced as shown in Table 1.B.5.

The most important limitation of the long-run rent series is the lack of correction for quality changes and sample composition shifts. To some extent, the latter aspect may be less problematic for 1870–1913 since the index for these years is confined to a specific market segment, i.e., 3-room apartments. It is important to note that the matching of the Danish house price and rent series is imperfect. While the house price index relies on data for dwellings in rural areas prior to 1938, the rent index mostly covers urban areas. Moreover, no information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be inaccurate and we need to assume that changes in rents of different types of houses are strongly correlated.

Table 1.B.5. Data sources: rent index, Denmark

Period	Source	Details
1870–1913	Pedersen (1930)	*Geographic Coverage*: Copenhagen; *Type(s) of Dwellings*: 3 room apartments; *Method*: Average rents.
1914–1917	Statistics Copenhagen (1906–1966)	*Geographic Coverage*: Copenhagen; *Type(s) of Dwellings*: 1-8 room houses; *Method*: Average rents.
1918–1926	Statistics Copenhagen (1906–1966) and Statistics Denmark (1925)	*Geographic Coverage*: Copenhagen; *Type(s) of Dwellings*: 1-5 room houses; *Method*: Average rents.
1927–1954	International Labour Organization (various years)	*Geographic Coverage*: Danish towns; *Type(s) of Dwellings*: New and existing dwellings; *Method*: Average rents.
1955–1964	Statistics Copenhagen (1906–1966)	*Geographic Coverage*: Copenhagen; *Type(s) of Dwellings*: 1-5 room houses; *Method*: Average rents.
1965–2015	International Labour Organization (various years), Statistics Denmark (2003, 2015)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: New and existing dwellings; *Method*: Average rents.

Finland

Rent data. Historical data on rents in Finland are available for 1920–2015.

The long-run rent index relies on the rent component of the consumer price index as published by the Ministry for Social Affairs (1920–1929), the International Labour Organization (various years), and Statistics Finland (2009). The main characteristics of the rent series are summarized in Table 1.B.6.

The main weakness of the long-run rent series relates to the lack of correction for quality changes and sample composition shifts. These aspects may be somewhat less problematic for the post-1964 period since the index is adjusted for the size

Table 1.B.6. Data sources: rent index, Finland

Period	Source	Details
1920–1926	Ministry for Social Affairs (1920–1929)	*Geographic Coverage*: 21 towns; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1927–1965	International Labour Organization (various years)	*Geographic Coverage*: 21 towns (1927–1936), 36 towns (1937–1965); *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1964–2015	Statistics Finland (2009)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents per sqm.

of the dwelling. Unfortunately, due to data limitations, the matching of the Finnish house price and rent series is imperfect. While the house price index relies on data for Helskinki prior to 1969, the rent index also covers more urban areas but is based on a larger city sample. In addition, no information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be inaccurate and we need to assume that changes in rents of different types of houses are strongly correlated.

France

Rent data. Historical data on rents in France are available for 1870–2015.

The long-run rent index relies on two main sources. For 1870–1948, we use an average rent index for Paris constructed by Marnata (1961). The index is based on a sample of more than 10,000 dwellings. Data come from lease management books from residential neighbourhoods in Paris and mostly refer to dwellings of relatively high quality. After 1949, we rely on national estimates, measured by the rent component of the CPI from the Statistics France (2015). The index covers tenants' rents only, i.e., imputed rents of owner-occupiers are excluded.

For the years prior to 1949, data on rents are also available for Paris (1914–1962) from the yearbooks of the International Labour Organization (various years). Reassuringly, the series by Marnata (1961) and the series published by the International Labour Organization (various years) are highly correlated for the years the overlap.[77] In addition, the International Labour Organization (various years) also presents a series for 45 departments for 1930–1937. For the years the series for Paris and the series for 45 departments overlap, they show similar rent increases. Note, however, that the house price index also relies on data for Paris only prior to 1936. For this reason, we use the Paris series throughout for the years prior to 1949 (Marnata, 1961). The available series are spliced as shown in Table 1.B.7.

Table 1.B.7. Data sources: rent index, France

Period	Source	Details
1870–1948	Marnata (1961)	*Geographic Coverage*: Paris; *Type(s) of Dwellings*: High-quality existing dwellings; *Method*: Average rents.
1949–2015	Statistics France (2015) as published in Conseil General de l'Environnement et du Developpement Durable (2013)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.

77. Correlation coefficient of 0.98.

The most important drawback of the long-run rent series is again the lack of correction for quality changes and sample composition shifts. Both aspects may be less problematic for the pre-WW2 years since the rent index is confined to a specific market segment, i.e., high-quality existing dwellings in Paris. Note further that the matching of the French house price and rent series in terms of geographical coverage has been generally possible. Both series are based on data for Paris prior to WW2 and on data for France as a whole for the second half of the 20th century. Yet, no additional information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

Germany

Rent data. Historical data on rents in Germany are available for 1870–2015.

The earliest data on rents in Germany comes from Hoffmann (1965). Hoffmann (1965) presents a rent index for 1850–1959. For 1850–1913, Hoffmann (1965) calculates a rent index using data on long-term interest rates and the replacement value of residential buildings, hence assuming that rents only depend on replacement costs and interest rates.

There are two additional sources on rents prior to WW1, both providing data on average rents in (parts of) Berlin. Bernhardt (1997) presents data on average rents for 1- and 2-room apartments between 1890 and 1910, and for 1-6 room apartments (separately for each size) in Berlin-Wilmersdorf between 1906–1913. Kuczynski (1947) provides an average rent based on scattered data for a number of larger German cities[78] for 1820–1913. Both sources, however, only report data for some years, not for the full period. For the 1895–1913 period, Kuczynski (1947) suggests a substantially stronger rise in nominal rents (42 percent) when compared to the index constructed by Hoffmann (1965) (22 percent). According to Hoffmann (1965), this can be explained by the fact that the index by Kuczynski (1947) does not account for quality improvements and may hence be upwardly biased. To be precise, the same bias should be present in Bernhardt (1997) as the data also refers to average rents. Yet, during the period they overlap (1890–1910), the series by Hoffmann (1965) and Bernhardt (1997) show about the same increase in rents while Kuczynski (1947) again suggests a significantly steeper rise.

For the years after 1913, Hoffmann (1965) relies on the rent component of the consumer price index as published by the Statistics Germany (1924–1935) (for 1913–1934) and Statistics Germany (various years) (for 1934–1959). The CPI rent

78. These include Berlin, Halle, Hamburg, Leipzig, Breslau, Dresden, Magdeburg, Barmen, Chemnitz, Jena, Lübeck, Magdeburg, Strassburg, and Stuttgart.

index is a weighted average of rents in 72 municipalities (with population used as weights) including small, medium, and large cities. It is based on data for working class family dwellings, typically 2 rooms with a kitchen. The index refers to existing dwellings, i.e., built prior to WW1, throughout. This, however, should not underestimate increases in rents given that dwellings built after WW1 only accounted for about 15 percent of all rental dwellings in 1934 (Statistics Germany, 1925; Statistics Germany, 1934).

Statistics Germany (various years) reports the CPI rent index for the years since 1948. The index relies on a survey of households and landlords and covers 3-4 room apartments in more than 100 German municipalities. Subsidized apartments are included. The index is calculated as a matched-models index and adjusts for major renovations (Angermann, 1985; Kurz and Hofmann, 2004).[79]

The long-run index is constructed as shown in the Table 1.B.8. For 1870–1912, we use the rent index constructed by Hoffmann (1965). For the years since 1913, we rely on the rent component of the consumer price index as published in Statistics Germany (1924–1935) and Statistics Germany (various years).

The long-run rent index has two main weaknesses. First, for the years prior to WW2, the index neither controls for quality changes nor for sample composition shifts. The latter aspect may be less of a problem for the interwar period ince the index is confined to a specific and presumably relatively homogeneous market segment, i.e., working class dwellings. Second, data prior to WW1 are not based on actual observed rents but have been estimated using data in replacement values and long-term interest rates.

Matching the German house price and rent series in terms of geographical coverage has been largely possible for the post-WW2 period. In both cases, data refers to Germany as a whole or at least covers a substantial share of the German housing market. This is unfortunately not the case for the pre-WW2 period. House price data for the pre-WW1 years only reflects trends in Berlin and Hamburg but the rent index covers all of Germany. For the interwar period, the house price index refers to urban real estate while the rent index provides a somewhat broader coverage. Moreover, no information on differences between the characteristics of the dwellings in the house price and the dwellings included in the rent index exist. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

79. The matched models method aims to control for quality changes by matching rents collected for a sample of models (or varieties of selected apartments) in a baseline period with rents of these same matched models in subsequent periods (Kurz and Hofmann, 2004).

Table 1.B.8. Data sources: rent index, Germany

Period	Source	Details
1870–1912	Hoffmann (1965)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Imputed rent based on long-term rates and replacement values of residential buildings.
1913–1947	Statistics Germany (1924–1935, various years)	*Geographic Coverage*: 72 municipalities; *Type(s) of Dwellings*: Working class dwellings; *Method*: Weighted average rents.
1948–2015	Statistics Germany (various years)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: 3-4 room apartments; *Method*: Matched models index.

Italy

House price data. Historical data on house prices in Italy are available for 1927–2015.

We rely on the long-run house price index constructed by Cannari and D'Alessio (2016) throughout. For 1927–1941, Cannari and D'Alessio (2016) rely on a series published in Statistics Italy's statistical yearbooks which, in turn, are based on house price indices constructed by the *Federazione Nazionale Fascista di Proprietari di Fabbricati*. The series is based on data for existing dwellings and reflects average transaction prices per room. For the years since 1966, the index relies on average transaction prices per square meter of new and existing dwellings in provincial capitals before 1997 and average transaction prices per square meter of new and existing dwellings in municipal districts after 1998. Data are drawn from publications of the *Consulente Immobiliare*.

Unfortunately, no price data are available for the period 1941–1961. To obtain a long-run index, Cannari and D'Alessio (2016) link average prices per room in eight cities (Turin, Genoa, Milan, Trieste, Bologna, Rome, Naples and Palermo) in 1941 with average transaction prices per room in these cities in 1966 assuming an average room size of 18 square meters. To obtain an annual house price series for 1941–1966, Cannari and D'Alessio (2016) interpolate using data on year-to-year increases in construction costs.

Rent data. Historical data on rents in Italy are available for 1927–2015. The long-run index relies on the CPI rent component throughout and spliced as shown in Table 1.B.9. Data are drawn from International Labour Organization (various years) and reflect average rents. The index covers tenants' rents only, i.e., imputed rents of owner-occupiers are excluded. Due to data availability, geographic coverage varies over time. The series reflects average rents in Milan (pre-1938), in 62 cities (1938–

Table 1.B.9. Data sources: rent index, Italy

Period	Source	Details
1927–1937	International Labour Organization (various years)	*Geographic Coverage*: Milan; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1938–1955	International Labour Organization (various years)	*Geographic Coverage*: 62 cities; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1956–2015	International Labour Organization (various years)	*Geographic Coverage*: 92 cities; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.

1955), and 92 cities (post-1955). The series has a gap between 1939 and 1945. Since, to the best of the author's knowledge, no data on rents are available for this period, we link the pre-1939 and post-1945 series assuming that rents increased in lockstep with house prices, i.e., by a factor of about 1.6 adjusted for inflation.

The single most important drawback of the long-run rent series is again the lack of correction for quality changes and sample composition shifts. Moreover, the matching of the Italian house price and rent series is unfortunately imperfect. While the rent index is only based on data for Milan before 1937 and for urban areas more generally thereafter, the house price index offers a more comprehensive geographic coverage. Second, no additional information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be inaccurate and we need to assume that changes in rents of different types of houses are strongly correlated.

Japan

Rent data. Historical data on rents in Japan are available for 1931–2015.

The long-run rent index relies on the rent component of the consumer price index throughout. For 1931–1946, the CPI rent index is reported in the yearbooks of the International Labour Organization (various years). The index covers 13 cities through 1936 and 24 cities thereafter and refers to average rents of wooden houses.

For the years since 1947, the rent component of the CPI is published by Statistics Japan (2012). Data are collected as part of the *Retail Price Survey* in more than 1200 districts. The rent index covers small and medium-sized wooden houses as well as non-wooden houses and refers to the average rent per sqm. Subsidized dwellings are included. Imputed rents for owner-occupiers are included since 1970 (Shiratsuka, 1999; International Labour Organization, 2013). The available series are spliced as shown in Table 1.B.10.

Table 1.B.10. Data sources: rent index, Japan

Period	Source	Details
1931–1946	International Labour Organization (various years)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: Wooden houses; *Method*: Average rents.
1947–2015	Statistics Japan (2012)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: Small and medium-sized wooden houses, non-wooden houses; *Method*: Average rents per sqm.

The most important limitation of the long-run rent index is the lack of correction for quality improvements and sample composition shifts. Particularly the latter aspect may be somewhat less problematic for the post-WW2 years since the series controls for the size of the dwelling. Matching the Japanese house price and rent series in terms of geographical coverage has been partly possible. For the pre-WW2 years both series are based on data for urban dwellings only. Yet for the second half of the 20th century, the rent index offers a somewhat broader coverage. In addition, the house price index reflects residential land prices only whereas the rent index naturally is based on rents for dwellings.

Netherlands

Rent data. Historical data on rents in the Netherlands are available for 1870–2015.

We rely on the long-run rent index constructed by Ambrose, Eichholtz, and Lindenthal (2013) throughout. The series is based on two main sources. For 1870–1913, it uses the rent component of the cost of living index calculated by Van Riel (2006). This pre-WW1 series refers to imputed rents of owner-occupied houses. Data comes from tax authorities and are estimated relying on average rents of comparable renter-occupied dwellings in the vicinity. For the post-WW1 period, Ambrose, Eichholtz, and Lindenthal (2013) draw data from various publications of Statistics Netherlands. Statistics Netherlands collects data through annual rent surveys and covers more than two thirds of Dutch municipalities. The nationwide index is a weighted average of rent changes by region. It is adjusted for the effect of major renovations (Statistics Netherlands, 2010; Statistics Netherlands, 2014). The main characteristics of the series are summarized in 1.B.11.

One alternative series for the pre-WW2 period is available which can be used as comparative to the index presented by Ambrose, Eichholtz, and Lindenthal (2013). For 1909–1944, Statistics Amsterdam (1916–1944) reports average rents of working class in Amsterdam that have not undergone significant alteration or renova-

Table 1.B.11. Data sources: rent index, Netherlands

Period	Source	Details
1870–1913	Van Riel (2006) as published in Ambrose, Eichholtz, and Lindenthal (2013)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1914–2015	Statistics Netherlands as published in Ambrose, Eichholtz, and Lindenthal (2013)	*Geographic Coverage*: Nationwide ; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Weighted average rents.

tion.[80] Both series, i.e., the index constructed by Ambrose, Eichholtz, and Lindenthal (2013) and the series published in the Statistics Amsterdam (1916–1944) are strongly correlated for the years they overlap.[81] This is reassuring since the long-run house price index only relies on data for Amsterdam prior to 1970 (Knoll, Schularick, and Steger, 2017).

The main weakness of the long-run rent series is again the lack of correction for quality changes and sample composition shifts. Moreover, it is important to note that the matching of the Dutch rent and house price series is unfortunately imperfect. This is mainly for two reasons. First, while the house price index relies on data for Amsterdam only prior to 1970, the rent index offers a broader geographical coverage. Yet, the evidence suggests that at least during the first half of the 20th century, rents in Amsterdam and the rest of the country moved closely together. Second, no information exists on the extent to which characteristics of the dwellings included in the house price index differ from those included in the rent index. The matching of the series with respect to the exact type of dwelling covered may hence be inaccurate and we need to assume that changes in rents of different types of houses are strongly correlated.

Norway

Rent data. Historical data on rents in Norway are available for 1871–2015.

For the period 1871–1978, the long-run index relies on a rent index presented by Jurgilas and Lansing (2012).[82] The series uses the rent component of the consumer price index since 1914[83] which for the years since 1920 is based on data for 26 towns

80. For 1909 to 1928, Statistics Amsterdam (1916–1944) provides only scattered evidence, i.e., data on 1909, 1912, 1918. The series are continuous after 1928. Statistics Amsterdam (1916–1944) also presents data on average rents of middle class dwellings. Yet, this series is based on a significantly smaller sample compared to the one for working class dwellings. According to the 1936–37 yearbook, for example, the data covers 1719 working class dwellings but only 110 middle class dwellings.

81. Correlation coefficient of 0.92 for 1909–1940.

82. The series were constructed by Ola Grytten, Norwegian School of Economics, and sent by email. Contact person is Marius Jurgilas, Norges Bank.

83. See for example the rent index for 1914–1948 as reported in Statistics Norway (1949, Table 185) and for 1924–1959 as reported in Statistics Norway (1978, Table 287) for comparison.

Table 1.B.12. Data sources: rent index, Norway

Period	Source	Details
1871–1978	Rent index underlying by the price to rent ratio reported in Jurgilas and Lansing (2012)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Weighted average rents.
1979–2013	Statistics Norway (2015)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Weighted average rents.

and 5 industrial centers across Norway and on data for Oslo only for 1914–1919. For the pre-WW1 period, the index is constructed as a weighted average of average rents in 32 cities and towns.[84] Data comes from consumption surveys conducted by Statistics Norway.

For the years prior to WW1, an additional series is available in Statistics Oslo (1915) covering average expenditures for rents of a family of four in Oslo for 1901–1914. Both series, i.e., the rent index by Jurgilas and Lansing (2012) and the data published in Statistics Oslo (1915), depict a similar trend for the years they overlap. For 1979–2015, the long-run rent index relies on the rent component of the consumer price index as published by Statistics Norway (2015). The series is based on a sample of about 2000 rented dwellings that are classified according to their age. The aggregate index is calculated as a weighted average rent index (Statistics Norway, 1991). The available series are spliced as shown in Table 1.B.12.

The main weakness of the long-run rent series is the lack of adjustment for quality changes and sample composition shifts. On the upside, the matching of the Norwegian house price and rent series in terms of geographic coverage has been generally possible. Both series rely on data for urban areas. Yet the coverage of the rent series is relatively more comprehensive. Unfortunately, no information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

Portugal

House price data. Historical data on house prices in Portugal are available for 1931–2015.

We rely on the long-run house price index constructed by Azevedo (2016). The author relies on the total number and value of transactions of new and existing real estate as reported to the land registry and collected by the Ministry of Justice to

84. Population is used as weights.

construct a weighted average house price index.[85] The number of transactions is used as weights. The data cover Portugal as a whole and are published in yearbooks and monthly bulletins by Statistics Portugal.[86]

Rent data. Historical data on rents in Portugal are available for 1948–2015.

The long-run rent index is based on the rent component of the consumer price index as published in International Labour Organization (various years). Data are collected by personal or phone interviews. The index covers tenants' rents only, i.e., imputed rents of owner-occupiers are excluded. The main characteristics of the series are summarized in Table 1.B.13.

Table 1.B.13. Data sources: rent index, Portugal

Period	Source	Details
1948–2015	International Labour Organization (various years)	*Geographic Coverage*: 1948–1950: Lisbon, 1951–1953: Lisbon and Porto, 1954–1961: 5 cities, 1962–1976: 6 cities, 1976–2015: 41 cities; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.

The main weakness of the long-run rent series is again the lack of correction for quality changes and sample composition shifts. Moreover, the matching of the Portuguese house price and rent series is unfortunately imperfect. While the rent index is only based on data for urban areas throughout, the house price index consistently offers a more comprehensive geographic coverage. Second, no additional information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be inaccurate and we need to assume that changes in rents of different types of houses are strongly correlated.

Spain

House price data. Historical data on house prices in Spain are available for 1900–2015.

We rely on the long-run house price index constructed by Amaral (2016) throughout. The author combines data from various sources to arrive at a long-run index. For 1900–1904, the series are based on average transaction prices of new and existing dwellings in Madrid and Barcelona. Data are collected from newspaper

85. While the data also includes commercial real estate, Azevedo (2016) argues based on evidence presented by Evangelista and Teixeira (2014) that commercial property transactions only account for a small share of all transactions recorded.

86. Sources are the various issues of the *Annuário estatístico de Portugal*, the *Estatísticas Monetárias e Financeiras*, and the *Boletins Mensais de Estística*.

advertisements.[87] For 1905–1933, Amaral (2016) uses an average transaction price index constructed by Carmona, Kampe, and Rosés (2017) based on data for all kinds of existing dwellings drawn from the *Registrars Yearbooks*. For 1934–1975, Amaral (2016) uses transaction price data for new and existing dwellings collected from the *Registrars Yearbooks* to construct a weighted average house price index covering Spain as a whole. For 1976–1986, the authors relies on a series of average transaction prices per square meter of new dwellings in Madrid constructed by the real estate agency *Tecnigrama*. For 1987–1994, the series is based on weighted average transaction prices per square meter of new and existing dwellings collected by the Spanish Ministry of Housing covering Spain as a whole. For the years after 1995, he relies on a nationwide index published by the Spanish Ministry of Public Works and Transports which reflects average transaction prices per square meter for new and existing dwellings.

Rent data. Historical data on rents in Spain are available for 1870–2015.

The earliest source for data on rents in Spain is Maluquer de Motes (2013) covering average rents of all kinds of dwellings in Catalunya between 1870 and 1933. Data are drawn from archival records and from the *Registrars Yearbooks*. For the years since 1935, the long-run rent index is based on the CPI rent index as published in the yearbooks of the International Labour Organization (various years) and Statistics Spain (2016). The index covers tenants' rents only, i.e., imputed rents of owner-occupiers are excluded. The available series are spliced as shown in Table 1.B.14.

The single most important drawback of the long-run rent series is again the lack of correction for quality changes and sample composition shifts. Moreover, the matching of the Spanish house price and rent series is unfortunately imperfect. While the rent index is only based on data for urban areas before 1976, the house

Table 1.B.14. Data sources: rent index, Spain

Period	Source	Details
1870–1936	Maluquer de Motes (2013)	*Geographic Coverage*: Catalunya; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1937–1976	International Labour Organization (various years)	*Geographic Coverage*: 1937–1956: 50 cities; 1957–1976: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Weighted average rents.
1977–2015	Statistics Spain (2016)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Weighted average rents.

87. On average, more than 120 observations per year were collected.

price data covers the whole of Spain. The opposite is true for the years between 1987 and 1994. After 1994, both series provide nationwide coverage. Second, no additional information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be inaccurate and we need to assume that changes in rents of different types of houses are strongly correlated.

Sweden

Rent data. Historical data on rents in Sweden are available for 1883–2015.

The earliest source for data on rents in Sweden is Myrdal (1933). For 1883–1913, Myrdal (1933) reports an index of average rents per room in Stockholm based on data published in the *Stockholm list of houses to let (Stockholms hyreslista)*, a publication advertising dwellings to let edited by the *Stockholms Intecknings Garanti Aktiebolag*. For 1913/14–1931, Myrdal (1933) reports the rent component of the cost of living index of the Social Board based on housing surveys and covering working or lower middle class dwellings in more than 40, predominantly urban, municipalities (Statistics Sweden, 1933).

For the years since 1932, the long-run rent index is based on the rent component of the consumer price index as published in International Labour Organization (various years) and Statistics Sweden (1961) and Statistics Sweden (1933). The main characteristics of this series are summarized in Table 1.B.15. The available series are spliced as shown in Table 1.B.15.

Table 1.B.15. Data sources: rent index, Sweden

Period	Source	Details
1882–1913	Myrdal (1933)	*Geographic Coverage*: Stockholm; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents per room.
1914–1931	Myrdal (1933)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents per room.
1932–1959	Statistics Sweden (1933, 1961)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1960–2015	International Labour Organization (various years)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.

The most important drawback of the long-run rent series is again the imperfect of correction for quality changes and sample composition shifts. Both aspects may be less problematic for years prior to 1931 since the rent index reflects average rents per room. Note further that the matching of the Swedish house price and rent series in terms of geographical coverage has been largely possible. For the years prior to 1960, both series are based on for urban areas. For the years after 1960, however, the rent index provides a more comprehensive geographical coverage compared to the house price series. Moreover, no additional information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

Switzerland

Rent data. Historical data on rents in Switzerland are available for 1890–2015.

The earliest source for rent data in Switzerland is Curti (1981). Curti (1981) separately calculates indices of rents for 3-room apartments for five cities (Zurich, Winterthur, Bern, Biel, and Basel) and the Zurich highlands for 1890–1910. Data are collected from newspaper advertisements.[88] For 1908–1920, Curti (1981) relies on data from the city of Zurich housing authority (as collected by Statistics Zurich). Curti (1981) adjusts the 3-year moving average of the spliced series so as to conform with the average rents of 3 room apartments according to the housing censuses of 1896, 1910 and 1920. Since for the years prior to 1930 the house price index for Switzerland is based on data for Zurich only (Knoll, Schularick, and Steger, 2017), we use the city index for Zurich for 1890–1910 to construct a long-run rent index.

For 1920–1939, we rely on the index of average rents for 3 room apartments in six working class neighborhoods as published by Statistics Zurich (1946–1962).[89]

For 1940–2015, the long-run index is based on the rent component of the consumer price index as published by Statistics Switzerland (2015). The series refers to new and existing 1-5 room apartments in 89 municipalities. Data are collected through surveys of households and the index is calculated as a weighted average.[90] The index is adjusted for major quality changes. The index covers tenants' rents only, i.e., imputed rents of owner-occupiers are excluded. The available series are spliced as shown in Table 1.B.16.

The main weakness of the long-run rent series is the lack of adjustment for quality changes for the pre-WW2 period. Sample composition shifts are unlikely to affect the index since data reflects the rent of 3-room apartments only. Note further, that matching the rent and the house price series with respect to geographic coverage has been largely possible. Both series before the 1930s are based on data for Zurich and for the whole of Switzerland after 1940. Yet, no additional information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with

88. The author collects about 30 advertisements per year from *Tagblatt der Stadt Zürich*.
89. These are Aussersihl, Industriequartier, Wiedikon, Wipkingen, and Unter- and Oberstrass.
90. The number of the different kinds of apartments (new and existing) is used as weights.

Table 1.B.16. Data sources: rent index, Switzerland

Period	Source	Details
1890–1919	Curti (1981)	*Geographic Coverage*: Zurich; *Type(s) of Dwellings*: 3 room apartment; *Method*: Average rent.
1920–1939	Statistics Zurich (1946–1962)	*Geographic Coverage*: Zurich; *Type(s) of Dwellings*: 3 room apartment; *Method*: Average rent.
1940–2015	Statistics Switzerland (2015)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: New and existing 1-5 room apartments; *Method*: Weighted average rent, adjusted for quality changes.

respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

United Kingdom

House price data. We extend the historical house price series in Knoll, Schularick, and Steger (2017) back to 1895 using the new house price index for London constructed by Samy (2015). The index is based on transaction-level data from the London Auction Mart, and constructed using a hedonic regression controlling for quality changes over time.

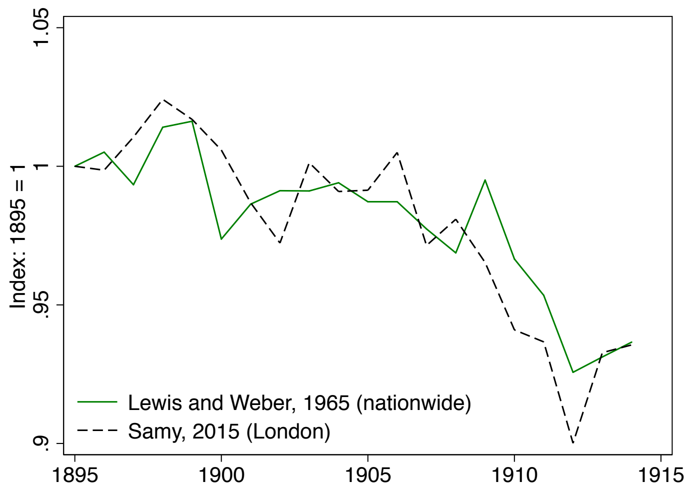
Rent data. Historical data on rents in the United Kingdom are available for 1895–2015. For 1895–1899, we rely on the rental index from the London Auction Mart data constructed by Samy (2015), to be consistent with the house price series. For 1900–1914, we rely on an index of average rents by Lewis and Weber (1965).[91] The series is based on property valuations for the *Inhabited House Duty*, a tax applied to residential houses with an annual rental value of 20 GBP or more. The index may hence include an element of quality increase as well as a true increase in rents, but comparison with the Samy (2015) quality-adjusted index for London, shown in Figure 1.B.18 suggests that the differences are very small.

For 1914–1938, the long-run rent index is based on the rent component of the official cost of living index compiled by the Ministry of Labor (as reported by Holmans (2005) and International Labour Organization (various years)). The series refers to average rents of working class dwellings in more than 500 towns. It is worth noting that the index reflects not only increases in rent proper but also in domestic rates.[92] The index lacks annual rental data during WW1, so we interpolate the an-

91. In principle, the Lewis and Weber (1965) series is available back to 1874, and closely tracks the London index assembled by Samy (2015) from the Auction Mart data.

92. According to Holmans (2005), in the housing market for working class families, dwellings were generally let at a rent that included domestic rates. Landlords recouped the rates they paid to

Figure 1.B.18. Comparison of pre-WW1 rental indices for the UK



Note: Real rental indices. The index in Lewis and Weber (1965) has nationwide coverage, but potentially does not control for quality adjustments. The index in Samy (2015) is for London only, but controls for quality changes.

nual rental changes during WW1 using the London-only index in Samy (2015), for years 1915–1919.

For the post-WW2 period, we use the rent component of the consumer price index as published in the yearbooks of the International Labour Organization (various years). Data are collected through surveys and cover also subsidized dwellings. For the years since 1956, the series includes expenditures on maintenance and repair. To the best of our knowledge, no data on rents exist between 1946 and 1954. To link the pre- and post-WW2 series, we use scattered data on average rents of houses and flats let by local authorities 1936–1957 presented by Holmans (2005). The available series are spliced as shown in Table 1.B.17.

The most important limitation of the long-run rent series is the lack of correction for quality changes and sample composition shifts. As noted above, the latter aspect may be less of a problem for the years 1914–1946 since the index is confined to a specific and presumably relatively homogeneous market segment, i.e., working class dwellings. The matching of the U.K. house price and rent series in terms of geographical coverage has been largely possible. Both series are based on data for the whole of the U.K. after WW2. The house price series reflects urban developments prior to 1930 as does the rent index during the interwar period. Yet, the rent series provides a more comprehensive coverage prior to WW1 compared to the house price series. Moreover, to the best of our knowledge, no information exists on the quality

local authorities through the rents they charged. While the dwellings may have thus been subject to rent controls according to the Rent Restriction Acts, increases in total rents to recoup increases in domestic rates were not limited according to these acts.

Table 1.B.17. Data sources: rent index, United Kingdom

Period	Source	Details
1895–1899	Samy (2015)	*Geographic Coverage*: London; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Hedonic regression.
1900–1913	Lewis and Weber (1965)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Average rents.
1914–1946	Rent component of official consumer price index as published in Holmans (2005) and International Labour Organization (various years)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: Working class dwellings; *Method*: Average rents. *Note:* We interpolate annual changes during 1915–1919 using the London index in Samy (2015).
1954–2013	Rent component of official consumer price index as published in International Labour Organization (various years)	*Geographic Coverage*: Nationwide; *Type(s) of Dwellings*: All kinds of dwellings; *Method*:

differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

United States

Rent data. Historical data on rents in the United States are available for 1890–2015.

For the 1890–1914, the long-run rent index relies on the rent component of the NBER cost of living index for manufacturing wage earners constructed by Rees (1961). The index is based on newspaper advertisements in six cities[93] and is confined to working class dwellings. The aggregate series is a simple average of the city indices. The index controls for differences in size but not for other potential sources of quality differences.

Data for 1915–1940 is available from U.S. Bureau of Labor Statistics (2015) which, in turn, are based on the Bureau of Labor Statistics' rental survey of landlords. The index is based on data on average rents for working class dwellings in 32 shipbuilding and other industrial centers for 1915–1935 and 42 cities with population over 50,000 thereafter. The series is based on comparisons of average rents for identical housing units (Bureau of Labor Statistics, 1966). Yet, several authors made the case for a downward bias of the historical U.S. Bureau of Labor Statistics (2015) rent series (Gordon and Goethem, 2007; Crone, Nakamura, and Voith, 2010), e.g.,

93. These are New York, Chicago, Philadelphia, Boston, Cincinnati, St. Louis.

due to aging bias or omission of new units. To adjust for the downward bias for 1915–1940, we use estimates by Gordon and Goethem (2007).[94]

For 1941–1995, the long-run index relies on the revised CPI for tenant rents constructed by Crone, Nakamura, and Voith (2010). Crone, Nakamura, and Voith (2010) argue that for the post-1995 period, tenant rents should be correctly calculated in the original U.S. Bureau of Labor Statistics (2015) series. For the post-1995 years, we therefore use the CPI rent index as published by U.S. Bureau of Labor Statistics (2015). The available series are spliced as shown in Table 1.B.18.

Compared to data for other countries, the U.S. rent series is relatively well adjusted for quality changes and sample composition shifts. Also, matching the house price and rent series with respect to geographical coverage has been largely possible. Both series rely on data for urban areas prior to WW2. Yet, while this is still true for the post-WW2 rent series, the house price index provides a more comprehensive coverage during that period. Apart from that, to the best of our knowledge, no information exists on the quality differences that may exist between the dwellings included in the house price and the dwellings included in the rent series. The matching of the series with respect to the exact type of dwelling covered may hence be imperfect and we need to assume that changes in rents of different types of houses are strongly correlated.

Table 1.B.18. Data sources: rent index, United States

Period	Source	Details
1890–1914	Rees (1961)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of working class dwellings; *Method*: Stratification.
1915–1940	U.S. Bureau of Labor Statistics (2015), adjusted using estimates by Gordon and Goethem (2007)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: Working class dwellings; *Method*: Average rents.
1941–1995	Crone, Nakamura, and Voith (2010)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Stratification.
1996–2015	U.S. Bureau of Labor Statistics (2015)	*Geographic Coverage*: Urban areas; *Type(s) of Dwellings*: All kinds of dwellings; *Method*: Stratification.

94. Gordon and Goethem (2007) estimate a CPI bias of -0.86 percent per year for 1914–1935 and of -1.04 percent for 1935–1960.

1.B.5 Equity and bond returns

This section details the sources used to construct the total equity and bond return series in this paper.

Australia

Table 1.B.19. Data sources: equity and bond returns, Australia

Year	Data source
Equity returns:	
1870–1881	Sum of capital gains, dividends and gains or losses from stock operations for Australian shares listed in London, weighted by market capitalization. Constructed from *Investor Monthly Manual* (IMM) data, various issues (http://som.yale.edu/imm-issues).
1882–2008	With-dividend return from Brailsford, Handley, and Maheswaran (2012). Note: we use these series rather than the alternative from NERA Economic Consulting (2015) due to greater consistency with the IMM historical series.
2009–2013	Total equity return from NERA Economic Consulting (2015).
2014–2015	ASX 200 total return index, from RBA Statistics Table F7.
Bond returns:	
1900–1925	Total return on Australian government bonds listed in Sydney from Moore (2010b). Converted from pound sterling to Australian Dollar.
1926–1968	Total return on Australian bonds listed in London. Data for 1926–1929 are from Meyer, Reinhart, and Trebesch (2015), shared by Josefin Meyer. Data for 1930–1968 were constructed by the authors.
1969–1987	Implied capital gain + yield from the 10-year government bond yield series published by the Reserve Bank of Australia. Capital gain estimated from movements in yields, using monthly yield data. Spliced with London listings data over 1968–1969.
1988–2015	Average of total returns on individual Australian government bonds, targeting 10-year maturity.

We are grateful to Josefin Meyer and Christoph Trebesch for sharing historical bond return data for Australia.

Belgium

Table 1.B.20. Data sources: equity and bond returns, Belgium

Year	Data source
Equity returns:	
1870–2015	Total return on all common stocks of Belgian companies listed on the Brussels stock exchange, provided by Frans Buelens. Market capitalization weighted. See Annaert, Buelens, Cuyvers, De Ceuster, Deloof, and De Schepper (2011) for further details.
Bond returns:	
1870–1913	Total return on the 3% rente; price and yield data from Drappier (1937), Table II.
1914–1937	Data from the SCOB database shared by Frans Buelens; total return on long-term government bonds, aggregated from individual bond data.
1938–1995	Total return on long-term government bonds, from various issues of National Bank of Belgium *Economic Summaries* and Ten-year Statistics, calculated from monthly data. 1938–1953: 4% perpetual bonds. Spliced with the SCOB data over the period 1938–1940. 1954–1963: 5-20 year 4.5% bond issued before 1962; price changes estimated using movements in yields. 1963–1970: Weighted average of 5-20 year bonds issued before 1962 and 5+ year bonds issued after 1962. 1971–1989: 5+ year maturity bonds, price changes estimated from movements in yields. 1989–1995: basket of 6+ maturity bonds, mean maturity approximately 10 years, price changes estimated from movements in yields.
1996–2015	Total return on 10-year government bonds, National Bank of Belgium online database, price changes estimated from movements in yields.

We are grateful to Frans Buelens for sharing the historical equity and bond return series from the SCOB database of the Brussels stock exchange.

Denmark

Table 1.B.21. Data sources: equity and bond returns, Denmark

Year	Data source
Equity returns:	
1873–1900	Total return on all shares of Danish firms listed on Danish stock exchanges, market cap weighted. Computed from microdata in *Green's Dankse Fonds og Aktier*, various years.
1901–1922	Total return on a broad selection of Danish shares, market cap weighted. We take all shares listed in the statistical yearbooks (Statistisk aarbog, years 1896–1927). For years 1914–1922, we combine the all-share price index in the statistical yearbook with the market cap weighted dividend series based on the smaller selection of stocks.
1923–1999	Combination of dividend yields from Nielsen and Risager (2001b) (market-cap weighted, circa 100 companies), and the share price index from Jordà, Schularick, and Taylor (2017), which is compiled from League of Nations, UN and IMF data.
2000–2001	Returns on the MSCI total return index.
2002–2015	Total return on the OMXCGI index.
Bond returns:	
1870–1990	Total return on long-term government bonds from Statistics Denmark (1969) and various issues of the Danmarks Nationalbank's *Monetary Review*. Perpetuals up to 1923, 10-40 year bonds for 1924–1980, 10-year maturity bonds from 1980 onwards.
1991–2015	Statistics Denmark, total return on the 10-year bullet loan

We are grateful to Kim Abildgren for helpful advice about the historical Danish stock return series.

Finland

Table 1.B.22. Data sources: equity and bond returns, Finland

Year	Data source
Equity returns:	
1895–1912	Total return index from Poutvaara (1996), based on several banks.
1913–1990	Total return index from Nyberg and Vaihekoski (2014), from the data shared with us by Mika Vaihekoski.
1991–2015	OMX Helsinki all-share total return index
Bond returns:	
1870–1925	Total return on long-term Finnish government bonds listed abroad, constructed from individual bond data in Arola (2006) (data from the online appendix of Nyberg and Vaihekoski (2011)).
1926–1991	Total return on approximately 5-year maturity government bonds from Nyberg and Vaihekoski (2011), using price movements implied by changes in market yield.
1992–2016	Average of total returns on individual Finnish government bonds, targeting 10-year maturity.

We are grateful to Mika Vaihekoski for sharing data and assisting with numerous queries regarding the Finnish stock and bond return series.

France

Table 1.B.23. Data sources: equity and bond returns, France

Year	Data source
Equity returns:	
1870–2010	Total return index from Le Bris and Hautcoeur (2010). Index constructed to mirror the methodology of the CAC-40: returns on largest 40 listed French firms weighted by market cap, with a continuously updated sample, market cap weighted.
2011–2015	Total return on the CAC-40 index.
Bond returns:	
1870–1969	Total return on 4% and 5% rente (perpetual bonds). Data provided by David Le-Bris, from Le Bris and Hautcoeur (2010).
1970–2015	Total return on a representative basket of long-term government bonds. Assume 10-year maturity before 1990 and 30-year after; as in Le Bris and Hautcoeur (2010). Price movements estimated from changes in yields at monthly frequency. Data provided by David LeBris, from Le Bris and Hautcoeur (2010).

We are grateful to David Le Bris for sharing data, assisting with numerous queries and providing helpful comments on the paper.

Germany

Table 1.B.24. Data sources: equity and bond returns, Germany

Year	Data source
Equity returns:	
1870–1889	Total return on the value-weighted top-30 blue-chip index from Ronge (2002)
1890–1913	All-share value-weighted performance index from Eube (1998).
1914–1959	Total return on the value-weighted top-30 blue-chip index from Ronge (2002).
1960–1990	Total return index from Gielen (1994), value-weighted, broad coverage. We use the "net" performance index, which excludes the adjustment for dividend income tax credit.
1991–1995	Total return on the DAX index.
1996–2016	Total return on the CDAX index.
Bond returns:	
1870–1903	Total return on listed long-term government bonds, arithmetic average of returns on individual bonds, with price and yield data collected from Homburger (1905) For early years we use regional bonds to fill gaps.
1904–1930	Total return on listed government bonds from the *Berliner Börsenzeitung*. Arithmetic average of individual bond returns. Average maturity generally 5-15 years. No data for the hyperinflation period of 1923–25.
1931–1943	total return on 4.5–6% government bonds (6% until 1935, then converted to 4.5%), aggregated using individual bond data from Papadia and Schioppa (2016), Deutsche Bundesbank (1976) and *Statistisches Jahrbuch für das Deutsche Reich*, various issues. Spliced with the *Berliner Börsenzeitung* series over 1928–1930.
1948–1955	Total return on mortgage bonds (Pfandbriefe, 4% and 5% coupons, from Deutsche Bundesbank (1976) and *Statistisches Jahrbuch für die Bundesrepublik Deutschland*, various issues.
1956–1967	Total return on public bonds from Deutsche Bundesbank (1976), using an average of bond returns for different issue yields. For years where the sample composition changes we use the return implied by yield movements, otherwise we use actual price changes.
1969–2015	REX government bond total return index, Bundesbank database series BBK01.WU046A.

We are grateful to Ulrich Ronge for sharing data and assisting with a number of queries, and to Carsten Burhop for helpful advice. We would also like to thank Andrea Papadia for sharing data.

Italy

Table 1.B.25. Data sources: equity and bond returns, Italy

Year	Data source
Equity returns:	
1870–1887	Capital gain + dividend return on stocks listed on the Genova stock exchange. Cacluated using indices in Da Pozzo and Felloni (1964), which are a book capital weighted average of returns on individual shares.
1888–1912	Total return on shares listed at the Milan Stock Exchange from Baia Curioni (2001). Market cap weighted.
1913–1954	Capital gain + dividend return on a broad index of Italian shares from Rosania (1954). Market cap weighted.
1955–1969	Capital gain on a broad index of Italian shares from Mondani (1978) (capitalization-weighted), plus dividend returns computed using total dividends paid and market capitalization data (as total dividends in lira / market cap), covering the vast majority Italian listed firms. Data sourced from *Mediobanca: indici e dati*, various years.
1970–2015	Total return on the main *Mediobanca* index, from Mediobanca (2013) and Mediobanca (2016).
Bond returns:	
1870–1913	Sum of lagged current yield and capital gain on the 5% perpetual bond (Rendita), computed from data in Bianchi (1979).
1913–1954	Sum of lagged current yield and capital gain on a representative basket of long-term government bonds, computed from data in Rosania (1954).
1955–1987	Total return on listed government bonds using data in various years of *Mediobanca: indici e dati*, targeting a maturity of 10 years. For the 1980s, only data on 3-5 year maturity bonds were used since longer dated government bonds were not typically listed on the stock exchange.
1988–2015	Average of total returns on individual Italian government bonds, targeting 10-year maturity. For 1988–1991, maturity is generally shorter than 10 years since almost all the bonds traded had relatively short maturities.

We are grateful to Stefano Battilossi for helpful advice about the historical series, and Giovanni Pellegrino for help with translating historical sources. We are also grateful to Massimo Caruso, Giuseppe Conte and Roberto Violi at Banca d'Italia for helpful advice and help in accessing historical publications.

Japan

Table 1.B.26. Data sources: equity and bond returns, Japan

Year	Data source
Equity returns:	
1882–1940	Sum of capital gain (Laspeyres index, base 1934–36), dividend return and gain/loss from stock operations, weighted by clearing transaction volumes, from Fujino and Akiyama (1977).
1941–1945	Capital gain from Bank of Japan (1966) + dividend return estimated using 1940 dividend yield, growth in nominal dividends paid by Japanese businesses from Bank of Japan (1966), and share price growth from Bank of Japan (1966) (chain linked).
1946–1947	Stock exchange closed; no data.
1948	Capital gain from Unted Nations' *Monthly Bulletin of Statistics* + dividend return estimated using growth in nominal dividends paid by Japanese businesses, as above.
1949–1951	Capital gain from *Bureau of Statistics Japan*, Table 14-25-a "Transactions and Yields of Listed Stocks, Tokyo Stock Exchange 1st Section" + dividend return from Fujino and Akiyama (1977) + gain/loss from stock operations from Fujino and Akiyama (1977).
1952–2015	Capital gain and dividend return from *Bureau of Statistics Japan* Tables 14-25-a and Table 14-25-b, covering Tokyo Stock Exchange 1st and 2nd section, + gain/loss from stock operations from Fujino and Akiyama (1977) (note: the Fujino and Akiyama (1977) series stop in 1975).
Bond returns:	
1880–1940	Lagged current yield + capital gain on central government bonds, from Fujino and Akiyama (1977). Price index used: Laspeyres, base 1934–36.
1941–1965	Secondary markets for government debt were shut down for a prolonged time after WW2, hence we use government bond yield data (not total returns) for this period. Sources are Homer and Sylla (2005) for 1941–1963 (long-term government bond yield), and IMF's IFS database for 1964–65 (Section "Interest rates", Series "Government Bonds").
1966–1970	Lagged current yield + capital gain on central government bonds, from Fujino and Akiyama (1977). Price index used: Laspeyres, base 1969–71.
1971–1987	Total return on long-term government bonds; 9-10 year maturity, from Hamao (1991).
1988–2015	Average of total returns on individual Japanese government bonds, targeting 10-year maturity.

We are grateful to Ryoji Koike for helpful advice, and to Yuzuru Kumon and Kaspar Zimmermann for assisting with collecting and interpreting the data.

Netherlands

Table 1.B.27. Data sources: equity and bond returns, Netherlands

Year	Data source
Equity returns:	
1900–1995	Total stock return index from Eichholtz, Koedijk, and Otten (2000), based on a selection of Dutch stocks, using data kindly shared with us by Roger Otten. The stock exchange was closed from from August 1944 to April 1946, so the 1945 return covers the period August 1944–April 1946.
1996–2003	CBS total equity return reinvestment index, from *CBS Statline*.
2004–2015	AEX all-share index.
Bond returns:	
1870–1900	Total return on the 2.5% perpetual bond, using data in Albers (2002).
1901–1987	Total return on long-term government bonds from Eichholtz, Koedijk, and Otten (2000), using data kindly shared with us by Roger Otten.
1988–2003	CBS total bond return reinvestment index, bonds of 8 years and above maturity, from *CBS Statline*.
2004–2015	Average of total returns on individual Dutch government bonds, targeting 10-year maturity.

We are grateful to Roger Otten for sharing the data on historical stock and bond returns in the Netherlands.

Norway

Table 1.B.28. Data sources: equity and bond returns, Norway

Year	Data source
Equity returns:	
1881–1920	Total return on all stocks listed on the Oslo stock exchange, market cap weighted. Constructed from share-level microdata collected from the following publications: *Kurslisten over Vaerdipapier* (the stock listing), *Farmand* magazine, and *Kierulfs haandbok over aktier og obligationer*, various years.
1921–1969	Capital gain from Klovland (2004b) plus dividend return from various issues of Norway's historical statistics and statistical yearbooks (*Historisk Statistikk*, *Statistisk Årbok*).
1970–1983	Capital gain from Klovland (2004b) plus dividend return constructed using the MSCI Norway total return and price return indices.
1984–2000	Capital gain from Klovland (2004b) plus dividend return constructed as total dividends paid by listed firms in proportion to total market capitalization.
2001–2015	Total return on the OSEBX index.
Bond returns:	
1870–1919	Total return on long-term government bonds listed on the Oslo Stock Exchange and major foreign exchanges. We use Oslo data unless there are few bonds being traded, in which case we rely on foreign exchanges. Oslo data come from *Kurslisten over Vaerdipapier*, *Farmand* magazine, and *Kierulfs haandbok over aktier og obligationer*. London data are from the *Investor Monthly Manual* (http://som.yale.edu/imm-issues), various issues. Other major markets' data are from Klovland (2004a), with price movements estimated from changes in yields.
1920–1992	Total return on 10-year government bonds, with price changes estimated from movements in monthly yields in Klovland (2004a).
1993–2015	Average of total returns on individual Norwegian government bonds, targeting 10-year maturity.

We are grateful to Jan Tore Klovland for answering numerous queries and helpful advice, and to the staff at the Oslo Nasjonalbiblioteket for help in locating the historical data sources.

Portugal

Table 1.B.29. Data sources: equity and bond returns, Portugal

Year	Data source
Equity returns:	
1870–1987	Total return on all shares listed on the Lisbon stock exchange, market capitalization weighted. Own calculations using share price, dividend and balance sheet information in the following publications: *Diario do Governo*, *Boletim da Bolsa* and annual reports of public companies, various years. For years 1900–1925, capital for a large number of companies had to be estimated using the trend in capital of a small number of firms. For year 1975, the stock exchange was closed because of the Carnation Revolution. We assumed no dividends were paid, and interpolated the stock prices of firms listed both before and after the closure to compute returns.
1988–2015	Total return on the PSI all-share index.
Bond returns:	
1870–1993	Total return on central government bonds listed on the Lisbon stock exchange. Average maturity around 15–30 years. Computed from bond listings data in *Diario do Governo* and *Boletim da Bolsa*. Weighted by the capitalization of individual bonds. During 1975 the stock exchange was closed, and we used yield data from the Bank of Portugal Statistics, series "Yield on fixed rate treasury bonds—10 years (monthly average)", and estimated price movements from changes in yields.
1994–2015	Average of total returns on individual Portuguese government bonds, targeting 10-year maturity.

We are grateful to Jose Rodrigues da Costa and Maria Eugenia Mata for help and advice in finding and interpreting the data sources for the historical Portuguese data. We are also grateful to staff at the Banco do Portugal archive for helpful advice and sharing data.

Spain

Table 1.B.30. Data sources: equity and bond returns, Spain

Year	Data source
Equity returns:	
1900–1940	Total return on all Spanish ordinary shares listed at the Madrid Stock Exchange, weighted by market capitalization. Data for 1900–1926 were kindly shared with us by Lyndon Moore (see Moore, 2010a; Moore, 2010b). Data for 1926–1936 were collected at the archive of the Banco de España, using stock exchange listings in various issues of the *Boletin de Cotization Oficial* of the Madrid stock exchange. The stock exchange was closed during the Spanish Civil war years 1937–1939. For these years, we calculated the returns using the average return on shares listed both before and after the exchange was closed, and assumed no dividends were paid (this seems reasonable since even in 1940, very few companies paid our dividends).
1940–2015	IGBM and Historical IGBM total return index for the Madrid stock exchange. Sources: López, Carreras, and Tafunell (2005), Chapter 10, “Empresa y Bolsa”, Table 10.33; Fernandez, Carabias, and Miguel (2007), European Federation of Exchanges. All shares, market capitalization weighted.
Bond returns:	
1900–1936	Total return on long-term government bonds listed on the Madrid Stock Exchange, market capitalization weighted, average maturity around 25 years. Data for 1900–1926 were kindly shared with us by Lyndon Moore (see Moore, 2010a,b).
1940– 1972	Total return on long-term government bonds from various issues of statistical bulletins, *Anuario Estadístico da España* (http://www.ine.es/inebaseweb/25687.do).
1973–1990	Total return on government bonds traded on the Barcelona stock exchange, from the *La Vanguardia* newspaper, various issues. Spliced with the series from statistical bulletins over years 1973–1975.
1989–2015	Total return on medium- and long-term government bonds from various issues of the *Banco de España Statistical Bulletin*. 1988–1994: maturity of less than 5 years; 1995–2015: maturity of 7–8 years.

We are grateful to Lyndon Moore for sharing data and providing helpful advice. We would also like to thank Stefano Battilossi for help with locating the historical data sources, and staff at the Banco de España archive for assisting with our queries.

Sweden

Table 1.B.31. Data sources: equity and bond returns, Sweden

Year	Data source
Equity returns:	
1871–2002	Total equity return index from Waldenström (2014).
2003–2015	OMXSGI total return index.
Bond returns:	
1870–1874	Total return on 4% and 5% perpetuals, using individual bond data in the online appendix of Waldenström (2014).
1874–2015	Holding period return on long-term government bonds from Waldenström (2014), generally targeting 10-year maturity. Extended to 2015 using own data.

We are grateful to Daniel Waldenström for helpful advice regarding the historical Swedish returns data.

Switzerland

Table 1.B.32. Data sources: equity and bond returns, Switzerland

Year	Data source
Equity returns:	
1900–1925	Total return on all Swiss stocks listed in Zurich, capitalization-weighted. Calculated using individual stock price and dividend data kindly shared with us by Lyndon Moore (see Moore, 2010a,b). The stock exchange closed from mid-1914 to mid-1916, and the 1915 return covers the period July 1914 to July 1916.
1926–1959	Total return on Swiss equities from Pictet & Cie (1998).
1960–1983	SBC total return index from Pictet & Cie (1998) and Swiss National Bank's *Kapitalmarkt* statistics.
1984–2015	SPI total return index from Pictet & Cie (1998), Swiss National Bank's *Kapitalmarkt* statistics and the SIX stock exchange statistics (six-group.com).
Bond returns:	
1899–1926	Total return on all Swiss government bonds listed on the Zurich stock exchange, capitalization-weighted. Calculated using individual bond price and yield data kindly shared with us by Lyndon Moore (see Moore, 2010a,b).
1927–1995	Total return on Swiss bonds from Pictet & Cie (1998).
1996–2015	SBI total bond return index from the SIX stock exchange statistics (six-group.com). 7+ year maturity before 2007 and 7–10 year maturity afterwards.

We are grateful to Lyndon Moore for sharing data and providing helpful advice, and to Rebekka Schefer for hepling us locate the historical sources.

United Kingdom

Table 1.B.33. Data sources: equity and bond returns, United Kingdom

Year	Data source
Equity returns:	
1870–1907	Total return on all UK stocks listed on the London stock exchange, capitalization weighted, from Grossman (2002).
1908–1963	Blue-chip market capitalization weighted index based on the largest 30 stocks listed on the London stock exchange, from Barclays (2016).
1964–2015	FTSE Actuaries all-share index, from Barclays (2016).
Bond returns:	
1870–1901	Total return on 3% and 2.75% consols from the *Statistical abstract for the UK*, various issues.
1902–2015	Total return on gilts (price change + lagged yield) from Barclays (2016). Targetting 20-year maturity before 1990 and 15-year maturity afterwards.

We are grateful to Richard Grossman and John Turner for helpful advice regarding historical UK stock and bond return data.

United States

Table 1.B.34. Data sources: equity and bond returns, United States

Year	Data source
Equity returns:	
1870–2015	Capital gain + dividend return from Shiller (2000) (up-to-date data from http://www.econ.yale.edu/~shiller/data.htm)
Bond returns:	
1870–1926	Total return on a basket of central government bonds around 10-year maturity. Calculated from prices of individual bonds in the *Commercial & Financial Chronicle*, various issues.
1927–1928	Total return on 10-year government bonds, price changes imputed from yields. Source: Aswath Damodaran database (http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/histretSP.html).
1929–2015	Total return on US long-term government bonds, from Barclays (2016).

We are grateful to Josefin Meyer for helpful advice concerning the historical bond return data for the US.

1.B.6 Taxes on real estate

Although the extent of real estate taxation varies widely across countries, real estate is taxed nearly everywhere in the developed world. International comparisons of housing taxation levels are, however, difficult since tax laws, tax rates, assessment rules vary over time and within countries. Typically, real estate is subject to four different kinds of taxes. First, in most countries, transfer taxes or stamp duties are levied when real estate is purchased. Second, in some cases capital gains from property sales are taxed. Often, the tax rates depend on the holding period. Third, income taxes typically also apply to rental income. Fourth, owners' of real estate may be subject to property taxes and/or wealth taxes where the tax is based upon the (assessed) value of the property.

This section briefly describes the current property tax regimes by country and provides estimates of the tax impact on real estate returns. With few exceptions, the tax impact on real estate returns can be considered to be less than 1 percentage point per annum.

Australia

Two kinds of property taxes exist. First, all but one Australian states/territories levy a land tax (no land tax is imposed in the Northern Territory). Typically, land tax is calculated by reference to the site value of the land (i.e., excluding buildings). Tax rates vary depending on the property value between 0.1% and 3.7%. Yet, the land tax is a narrow-based tax, i.e., many states apply substantial minimum thresholds and several land uses—such as owner-occupied housing—are exempt. Consequently, I will not consider any tax impact of land taxes on housing returns. Second, council rates are levied by local governments. Rates vary across localities rates and are set based on local budgetary requirements. Some councils base the tax on the assessed value of the land, others base it on the assessed value of the property as a whole (i.e., land and buildings) (Commonwealth of Australia, 2010). While all these specific make it difficult to determine an average or exemplary tax impact on returns, it can generally be considered to be well below 1%. Capital gains taxes apply only to investment properties, not to primary residences. Rates are higher the shorter the holding period. All Australian states levy stamp duties on property transfers. Rates vary across states and different types of property and may amount up to 6% of the property value (Commonwealth of Australia, 2010).

Belgium

Property taxes (*Onroerende voorheffing*) are levied on the cadastral value, i.e., the notional rental value, of the property. Rates range between 1.25% in Wallonia and Brussels and 2.5% in Flanders (Deloitte, 2016a). Using a tax rate 2.5% and a rent-price ratio of 0.045 (2012) the implied tax impact is $0.025 \times 0.045 \times 100 = 0.11\%$.

Capital gains taxes of 16.5% are levied if the property has been owned for less than five years. Property transfer taxes amount to 12.5% of the property value in Wallonia and Brussels and 10% in Flanders (Deloitte, 2016a).

Denmark

Two kinds of property taxes exist. First, the national property tax (*Ejen-domsvÃŠrdiskat*). The tax rate is 1% of the assessed property value if the property value is below DKK 3,040,000 and 3% above. The tax is not based on current assessed property values but on 2002 values. Second, a municipal land tax (*Grundskyld* or *Daekningsafgifter*) is levied on the land value. Rates vary across municipalities and range between 1.6% and 3.4% (Skatteministeriet, 2016). According to Pedersen and Isaksen (2015) the national property tax amounted to a little below 0.6% of property values in 2014 and municipal land taxes to about 0.07% giving us a combined tax impact of about 1.35% (Pedersen and Isaksen, 2015). No capital gains tax is payable if the property was the owners' principal residence. Stamp duties are levied on property transfers and amount to 0.6% of the purchase prices plus DKK 1,660.

Finland

Property taxes (*Kiinteistövero*) are levied by municipalities. Tax rates for permanent residences range between 0.37% and 0.8% of the taxable value where the taxable value is about 70% of the property's market value (KTI, 2015). The implied tax impact is therefore $0.8 \times 0.7 = 0.56\%$. Capital gains from property sales are taxed at progressive rates, from 30% to 33%. There is a 4% property transfer tax for property. First-time homebuyers are exempt from transfer taxes (KTI, 2015).

France

Property taxes (*taxe foncière sur les propriétés bâties*) are levied by municipalities. The tax base is the cadastral income, equal to 50% of the notional rental value (Public Finances Directorate General, 2015). Tax rates in 2014 ranged between 0.84% and 3.34% (OECD, 2016a). Using the rent-price ratio of 0.045 in 2012 and assuming a tax rate of 3.34%, the implied tax impact therefore is $0.045 \times 0.5 \times 0.034 \times 100 = 0.08\%$. Capital gains from property sales are taxed at 19%. Property transfer taxes amount to about 5% of the property value (Deloitte, 2015a).

Germany

Property laxes (*Grundsteuer*) are levied by federal states. Tax rates vary between 0.26% and 0.1% of the assessed value (*Einheitswert*) of the property and are multiplied by a municipal factor (*Hebesatz*). Since assessed values are based on historic values, they are significantly below market values. In 2010, assessed values were about 5% of market values (Wissenschaftlicher Beirat beim Bundesministerium der Finanzen, 2010). Municipal factors in 2015 ranged between 260% and 855% (median value of 470%) (Deutscher Industrie- und Handelskammertag, 2016). Using a tax rate of 0.5%, the implied tax impact is $0.05 \times 0.005 \times 4.7 = 0.12\%$. Capital gains from property sales are taxed if the property has been owned for less than 10 years (*Abgeltungssteuer*). Property transfer taxes are levied on the state level and range between 3.5% and 6.5% of the property value.

Japan

Two kinds of property taxes exist. First, a fixed assets tax is levied at the municipal level with rates ranging from 1.4 to 2.1 of the assessed taxable property value. The taxable property value is 33% of the total assessed property value for residential properties and 16% if the land plot is smaller than 200 sqm. Second, the city planning tax amounts to 0.3% of the assessed taxable property value. The taxable property value is 66% of the total assessed property value for residential properties and 33% if the land plot is smaller than 200 sqm (Ministry of Land, Infrastructure, Transport, and Tourism, 2016b). The implied tax impact is therefore $0.33 \times 2.1 + 0.66 \times 0.3 = 0.89\%$. Capital gains from property sales are taxed at 20% if the property has been owned for more than five years and at 39% if the property has been owned for less than five years. Owner-occupiers are given a deduction of JPY 30 mio. There is a national stamp duty (*Registered Licence Tax*) of 1% of the assessed property value and a prefectural real estate acquisition tax of 3% of the property value (Ministry of Land, Infrastructure, Transport, and Tourism, 2016a).

Netherlands

Property taxes (*Onroerendezaakbelasting*) are levied at the municipal level. Tax rates range between 0.0453% and 0.2636% (average of 0.1259%) of the assessed property value (*Waardering Onroerende Zaak (WOZ) value*) (Centrum voor Onderzoek van de Economie van de Lagere Overheden, 2016; Deloitte, 2016c). The tax impact on returns therefore ranges between about 0.05% and 0.26%. No capital gains tax is payable if the property was the owners' principal residence. Property transfer taxes amount to 2% of the property value (Deloitte, 2016c).

Norway

Property taxes are levied at the municipal level. Tax rates range between 0.2% and 0.7% of the tax value of the property. Typically, the tax value of a dwelling is about 25% of its assessed market value if the dwelling is the primary residence. Higher values apply for secondary residences. In addition, wealth taxes are levied at a rate of 0.85% (tax-free threshold is NOK 1.2 mio) on the tax value of the property (Norwegian Tax Administration, 2016). The implied tax impact therefore is

$0.25 \times 0.7 + 0.25 \times 0.85 = 0.39\%$. Capital gains from the sale of real estate property are taxed as ordinary income at 27%. A stamp duty of 2.5% applies to the transfer of real property (Deloitte, 2016b).

Sweden

Property taxes (*kommunal fastighetsavgift*) are levied at the municipal level. For residential properties, the tax rate is 0.75% of the taxable property value with taxable values amounting to about 75% of the property's market value. Fees are reduced for newly built dwellings (Swedish Tax Agency, 2012). The implied tax impact is therefore $0.75 \times 0.75 = 0.56\%$. Capital gains from sales of private dwellings are taxed at a rate of 22%. Stamp duties amount to 1.5% of the property value (Swedish Tax Agency, 2012).

Switzerland

Most Swiss municipalities and some cantons levy property taxes (*Liegenschaftssteuer*) with rates varying across cantons between 0.2% and 3% (property taxes are not levied in the cantons Zurich, Schwyz, Glarus, Zug, Solothurn, Basel-Landschaft, and Aargau). The tax is levied on the estimated market value of the property (Deloitte, 2015b). The tax impact on returns therefore ranges between 0.2% and 3%. Capital gains from property sales are taxed in all Swiss cantons (*Grundstückgewinnssteuer*). Tax rates depend on the holding period and range from 30% (if the property is sold within 1 year) and 1% (if the property has been owned for more than 25 years) of the property value. In addition, almost all cantons levy property transfer taxes (*Handänderungssteuer*). Tax rates vary between 10% and 33% (Eidgenössische Steuerverwaltung, 2013; ch.ch, 2016).

United Kingdom

Property taxes (*Council tax*) are levied by local authorities. Each property is allocated to one of eight valuation bands based on its assessed capital value (as of 1 April 1991 in England and Scotland, 1 April 2003 in Wales). Taxes on properties in Band D (properties valued between GBP 68,001 and GBP 88,000 in 1991) amounted to GBP 1484 in 2015 (Department for Communities and Local Government, 2016). Since 1991, nominal house prices have increased by a factor of about 2.5. The implied tax impact in 2015 for a property valued at GBP 68,001 in 1991 is $1484/(68,001 \times 2.5) \times 100 = 0.87\%$. No capital gains tax is payable if the property was the owners' principal residence. Property transfer tax rates (*Stamp Duty Land Tax*) depend on the value of the property sold and range between 0% (less than GBP 125,000) and 12.5% (more than GBP 1.5 m.) (Deloitte, 2016d).

United States

Property taxes in the U.S. are levied at the state level and are deductible from federal income taxes. Generally, tax rates are about 1% of real estate values, with rates varying across states. Giglio, Maggiori, and Stroebel (2015) assume the deductibility reflects a marginal federal income tax rate of 33%. The tax impact is thus $(1 - 0.33) \times 0.01 = 0.67\%$. Property transfer taxes are levied at the state level and range between 0.01% and 3% of property value (Federation of Tax Administrators, 2006).

References

Abildgren, Kim. 2016. “The National Wealth of Denmark 1845–2013 in a European Perspective.” *Danish Journal of Economics* 154 (1): 1–19.

Akbulut-Yuksel, Mevlude. 2014. “Children of War. The Long-Run Effects of Large-Scale Physical Destruction and Warfare on Children.” *Journal of Human Resources* 49 (3): 634–662.

Albers, Ronald Martin. 2002. *Machinery Investment and Economic Growth: The Dynamics of Dutch Development 1800–1913*. Aksant Academic Publishers.

Amaral, Francisco. 2016. “House Prices in Spain, 1870–2015.” Mathesis. University of Bonn.

Ambrose, Brent W., Piet Eichholtz, and Thies Lindenthal. 2013. “House Prices and Fundamentals: 355 Years of Evidence.” *Journal of Money, Credit and Banking* 45 (2–3): 477–491.

Angermann, Oswald. 1985. “Weiterentwicklung des Mietenindex in der Verbraucherpreisstatistik mit Hilfe von Ergebnisse der Wohnungszählung.” *Wirtschaft und Statistik* 6/1985: 505–508.

Annaert, Jan, Frans Buelens, Ludo Cuyvers, Marc De Ceuster, Marc Deloof, and Ann De Schepper. 2011. “Are Blue Chip Stock Market Indices Good Proxies for All-Shares Market Indices? The Case of the Brussels Stock Exchange 1833–2005.” *Financial History Review* 18 (3): 277–308.

Annaert, Jan, Frans Buelens, and Marc De Ceuster. 2012. “New Belgian Stock Market Returns: 1832–1914.” *Explorations in Economic History* 49 (2): 189–204.

Arola, Mika. 2006. “Foreign Capital and Finland: Central Government’s First Period of Reliance on International Financial Markets 1862–1938.” Bank of Finland Scientific Monograph E:37–2006.

Artola Blanco, Miguel, Luis E. Bauluz, and Clara Martínez-Toledano. 2017. “Wealth in Spain, 1900–2014: A Country of Two Lands.” Working Paper.

Australian Bureau of Statistics. 2011. “Consumer Price Index: Concepts, Sources and Methods.” *Information Paper* 6461.0:

Australian Bureau of Statistics. 2014. “Australian National Accounts: National Income, Expenditure and Product. Table 8: Household Final Consumption Expenditure.”

Azevedo, Joao. 2016. “House Prices in Portugal, 1930 to 2015.” Mathesis. University of Bonn.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2016. “Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy.” CEPR Discussion Paper 11734.

Baia Curioni, Stefano. 2001. *Modernizzazione e Mercato. La Borsa di Milano Nella "Nuova Economia" Dell’età Giolittiana (1888–1914).* Milan: EGEA.

Ball, Laurence, Douglas W. Elmendorf, and N. Gregory Mankiw. 1998. “The Deficit Gamble.” *Journal of Money, Credit and Banking* 30 (4): 699–720.

Bank of Japan. 1966. *Hundred-Year Statistics of the Japanese Economy*.

Barclays. 2016. “UK Equity and Gilt Study 2016.”

Barro, Robert J. 2006. “Rare Disasters and Asset Markets in the Twentieth Century.” *Quarterly Journal of Economics* 121 (3): 823–866.

Barro, Robert J., and José Ursúa. 2008. “Macroeconomic Crises since 1870.” *Brookings Papers on Economic Activity* 39 (1): 225–350.

Barro, Robert J., and Jose F. Ursua. 2008. “Consumption Disasters in the Twentieth Century.” *American Economic Review* 98 (2): 58–63.

Benhabib, Jess, and Alberto Bisin. 2016. “Skewed Wealth Distributions: Theory and Empirics.” NBER Working Paper 21924.

Bernhardt, Christoph. 1997. *Bauplatz Gross-Berlin: Wohnungsmärkte, Terraingewerbe und Kommunalpolitik im Städtewachstum der Hochindustrialisierung (1871–1918)*. Berlin: De Gruyter.

Bianchi, Bruno. 1979. "Appendice Statistica: Il Rendimento del Consolidato dal 1862 al 1946." In *Capitale Industriale e Capitale Finanziaro: Il Caso Italiano. Bologna: Il Mulino*.

Birck, Laurits Vilhelm. 1912. *Ejendomsskatter Og Eijendomspriser: En Studie*. Copenhagen: G.E.C. Gad.

Blancheton, Bertrand, Hubert Bonin, and David Le Bris. 2014. "The French Paradox: A Financial Crisis During the Golden Age of the 1960s." *Business History* 56 (3): 391–413.

Bracke, Philippe. 2015. "House Prices and Rents: Microevidence from a Matched Data Set in Central London." *Real Estate Economics* 43 (2): 403–431.

Brailsford, Tim, John C. Handley, and Krishnan Maheswaran. 2012. "The Historical Equity Risk Premium in Australia: Post-GFC and 128 Years of Data." *Accounting & Finance* 52 (1): 237–247.

Bureau of Economic Analysis. 2014. "Personal Consumption Expenditures by Major Type of Product."

Bureau of Labor Statistics. 1966. "The Consumer Price Index: History and Techniques." *Bureau of Labor Statistics Bulletin* 1517:

Butlin, N. G. 1985. "Australian National Accounts 1788–1983." Source Papers in Economic History 6. Australian National University.

Buyst, Erik. 1994. "Het Inkomen uit Onroerend Vermogen toevloeiend aan Particulieren: Belgie, 1920-1939." *Workshop in Quantitative Economic History Research Paper* 1994-01:

Caballero, Ricardo J., and Emmanuel Farhi. 2017. "The Safety Trap." *Review of Economic Studies*. Forthcoming:

Cabinet Office. Government of Japan. 2012. "Composition of Final Consumption Expenditure of Households Classified by Purpose."

Cairncross, Alexander K. 1975. *Home and Foreign Investment, 1870–1913: Studies in Capital Accumulation*. Edited by reprint of the 1953 ed. Clifton, N.J.: Augustus M. Kelley Publishers.

Campbell, John Y. 1999. "Asset Prices, Consumption, and the Business Cycle." In. Vol. 1, Handbook of Macroeconomics. Elsevier. Chapter 19, 1231–1303.

Cannari, Luigi, and Giovanni D'Alessio. 2016. "I Prezzi delle Abitazioni in Italia, 1927–2012." *Questioni di Economia e Finanza (Occasional Papers)* 333:

Cardoso, Abilio. 1983. "State Intervention in Housing in Portugal 1960–1980." Doctoral dissertation. University of Reading.

Cardoso, Fátima, Luıúsa Farinha, and Rita Lameira. 2008. "Household Wealth in Portugal: Revised Series." Banco de Portugal Occasional Papers 1-2008.

Carmona, Juan, Markus Kampe, and Joan Rosés. 2017. *Economic History Review* 70 (2): 632–658.

Centrum voor Onderzoek van de Economie van de Lagere Overheden. 2016. "Tarievenoverzicht 2016."

ch.ch. 2016. "Besteuerung Von Immobilien."

Cheng, Ping, Zhenguo Lin, and Yingchun Liu. 2008. "The Real Estate Risk Premium Puzzle: A Solution." Working Paper.

Commonwealth of Australia. 2010. "Australia's Future Tax System: Report to the Treasurer."

Conseil General de l'Environnement et du Developpement Durable. 2013. "House Prices in France: Property Price Index, French Real Estate Market Trends, 1200–2013."

Crafts, Nicholas. 2016. "Reducing High Public Debt Ratios: Lessons from UK Experience." *Fiscal Studies* 37 (2): 201–223.

Crone, Theodore M., Leonard I. Nakamura, and Richard Voith. 2010. "Rents Have Been Rising, Not Falling, in the Postwar Period." *Review of Economics and Statistics* 3: 628–642.

Curti, Marco. 1981. *Reallöhne schweizerischer Industriearbeiter von 1890 bis 1921: Ergebnisse 3 - Die Entwicklung der Wohnungsmieten*. Zurich: University of Zurich.

Da Pozzo, Mario, and Giuseppe Felloni. 1964. *La Borsa Valori di Genova nel Secolo XIX*. ILTE.

Dagens Nyheter. 1892. “Annonsering: 2 Stenhus Till Salu.” *Dagens Nyheter*, November 5, 1892.

Dagens Nyheter. 1897. “Annonsering: Hus.” *Dagens Nyheter*, September 3, 1897.

Dagens Nyheter. 1899. “Anonsering: Hrr Kapitalister.” *Dagens Nyheter*, December, 20, 1899.

Dahlman, Carl Johan, and Anders Klevmarken. 1971. “Private Consumption in Sweden, 1931–1975.” Uppsala.

De Telegraaf. 1939. “Stijgende Woningbouw in Ons Land.” *De Telegraaf*, (21): 47, January 21, 1939.

Deloitte. 2015a. “Taxation and Investment in France 2015: Reach, Relevance, and Reliability.”

Deloitte. 2015b. “Taxation and Investment in Switzerland: Reach, Relevance, and Reliability.”

Deloitte. 2016a. “Taxation and Investment in Belgium 2015: Reach, Relevance and Reliability.”

Deloitte. 2016b. “Taxation and Investment in Norway 2015: Reach, Relevance, and Reliability.”

Deloitte. 2016c. “Taxation and Investment in the Netherlands: Reach, Relevance, and Reliability.”

Deloitte. 2016d. “Taxation and Investment in United Kingdom 2015: Reach, Relevance, and Reliability.”

Department for Communities and Local Government. 2016. “Council Tax Levels Set by Local Authorities in England 2015-16 (Revised).”

Deutsche Bundesbank. 1976. *Deutsches Geld-Und Bankwesen in Zahlen, 1876-1975*. Knapp.

Deutscher Industrie- und Handelskammertag. 2016. “Realsteuer-Hebesätze.”

Diefendorf, Jeffry M. 1993. *In the Wake of War: The Reconstruction of German Cities After World War II*. Oxford: Oxford University Press.

Diewert, Erwin D. 2009. “Durables and Owner-Occupied Housing in a Consumer Price Index.” In *Price Index Concepts and Measurement*. Edited by Erwin W. Diewert, John S. Greenlees, and Charles R. Hulten. Chicago: University of Chicago Press, 445–500.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2009. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, N.J.: Princeton University Press.

Drappier, Jean-Marie. 1937. “La Conjoncture des Cours des Valeurs Mobilières, de Leurs Dividendes et des Taux d’Intérêt en Belgique de 1830 à 1913.” *Recherches Économiques de Louvain* 8 (4): 391–449.

Edvinsson, Rodney. 2016. “Historical National Accounts for Sweden 1800–2000.”

Eggertsson, Gauti B., and Neil R. Mehrotra. 2014. “A Model of Secular Stagnation.” NBER Working Paper 20574.

Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins. 2017. “A Model of Secular Stagnation: Theory and Quantitative Evaluation.” NBER Working Paper 23093.

Eichholtz, Piet M. A., C. G. Koedijk, and Roger Otten. 2000. “De Eeuw Van Het Aandeel.” *Economisch-statistische berichten* 85:

Eidgenössische Steuerverwaltung. 2013. *Die Handänderungssteuer*. Bern: Eidgenössische Steuerverwaltung.

Eisfeldt, Andrea, and Andrew Demers. 2015. “Total Returns to Single Family Rentals.” NBER Working Paper 21804.

Eube, Steffen. 1998. *Der Aktienmarkt in Deutschland vor dem Ersten Weltkrieg: Eine Indexanalyse*. Frankfurt am Main: Knapp.

Evangelista, R., and A. Teixeira. 2014. “Using Different Administrative Data Sources to Develop House Price Indexes for Portugal.” *INE Working Paper*,

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016. “Heterogeneity and Persistence in Returns to Wealth.” NBER Working Paper 22822.

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh. 2017. "The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk Sharing in General Equilibrium." *Journal of Political Economy* 125 (1): 140–223.

Federation of Tax Administrators. 2006. "State Real Estate Transfer Taxes."

Fernandez, Pablo, Jose M. Carabias, and Lucia Miguel. 2007. "Rentabilidad de los Fondos de Inversión de Renta Variable Nacional en España (1991–2006)." IESE Research Papers D/695.

Fishback, Price V., and Trevor Kollmann. 2015. "Hedonic Housing Indexes During the Great Depression." unpublished manuscript.

Fox, Ryan, and Peter Tulip. 2014. "Is Housing Overvalued?" RBA Research Discussion Paper 2014-06.

Francis, Jack Clark, and Roger G. Ibbotson. 2009. "Contrasting Real Estate with Comparable Investments, 1978 to 2008." *Journal of Portfolio Management* 36 (1): 141–155.

Fujino, Shozaburo, and Ryoko Akiyama. 1977. *Security Prices and Rates of Interest in Japan: 1874–1975*. Tokyo: Hitotsubashi University.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty. 2017a. "Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800–2014)." CEPR Discussion Paper 11848.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty. 2017b. "Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014)." BANK OF FRANCE WORKING PAPER No. WP 633.

Garland, John. M., and Raymond W. Goldsmith. 1959. "The National Wealth of Australia." In *The Measurement of National Wealth*. Edited by Raymond W. Goldsmith and Christopher Saunders. Income and Wealth Series VIII. Chicago, Ill.: Quadrangle Books, pp. 323–364.

Giacomini, Emanuela, David C. Ling, and Andy Naranjo. 2015. "Leverage and Returns: A Cross-Country Analysis of Public Real Estate Markets." *Journal of Real Estate Finance and Economics* 51 (2): 125–159.

Gielen, Gregor. 1994. *Können Aktienkurse Noch Steigen?: Langfristige Trendanalyse Des Deutschen Aktienmarktes*. Wiesbaden: Gabler-Verlag.

Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel. 2015. "Very Long-Run Discount Rates." *Quarterly Journal of Economics* 130 (1): 1–53.

Goldsmith, R. W. 1962. *The National Wealth of the United States in the Postwar Period*. Princeton, N.J.: Princeton University Press.

Goldsmith, Raymond W. 1985a. *Comparative National Balance Sheets: A Study of Twenty Countries, 1688–1978*. eng. Chicago, Ill.: University of Chicago Press.

Goldsmith, Raymond W. 1985b. *Comparative National Balance Sheets: A Study of Twenty Countries, 1688–1978*. Chicago, Ill.: University of Chicago Press.

Goldsmith, Raymond W., and A.C. Frijdal. 1975. "Le Bilan National de la Belgique de 1948 à 1971." *Cahiers Economiques de Bruxelles* 66: 191–200.

Gordon, Robert J., and Todd van Goethem. 2007. "Downward Bias in the Most Important CPI Component: The Case of Rental Shelter, 1914–2003." In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*. Edited by Ernst R. Berndt and Charles R. Hulten. Chicago, IL: University of Chicago Press, 153–195.

Graham, John R., Mark T. Leary, and Michael R. Roberts. 2015. "A Century of Capital Structure: The Leveraging of Corporate America." *Journal of Financial Economics* 118 (3): 658–683.

Grebler, Leo, David M. Blank, and Louis Winnick. 1956. *Capital Formation in Residential Real Estate: Trends and Prospects*. Princeton, N.J.: Princeton University Press.

Groote, Peter, Ronald Albers, and Herman De Jong. 1996. *A Standardised Time Series of the Stock of Fixed Capital in the Netherlands, 1900–1995*. Groningen Growth, and Development Centre, Faculty of Economics, University of Groningen.

Grossman, Richard S. 2017. "Stocks for the Long Run: New Monthly Indices of British Equities, 1869–1929." CEPR Discussion Paper 12042.

Grossman, Richard S. 2002. "New Indices of British Equity Prices, 1870–1913." *Journal of Economic History* 62 (1): 121–146.

Hamao, Yasushi. 1991. "A Standard Data Base for the Analysis of Japanese Security Markets." *Journal of Business* 64 (1): 87–102.

Hansen, Alvin H. 1939. "Economic Progress and Declining Population Growth." *American Economic Review* 29 (1): 1–15.

Hansen, Svend Aage. 1976. *Økonomisk Vækst I Danmark*. 6. Akademisk forlag.

Haynie, Henry. 1903. "Paris Past and Present." *New York Times.* January 10, 1903:

Henau, Anne. 1991. "De Belgische Huishuren Gedurende Het Interbellum." *Workshop in Quantitative Economic History Research Paper* 1991-01:

Hjerppe, Riitta. 1989. *The Finnish Economy 1860–1985: Growth and Structural Change*. Studies on Finland's Economic Growth. Helsinki: Bank of Finland.

Hoffmann, Walther G. 1965. *Das Wachstum der Deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts*. Berlin: Springer.

Holmans, A.E. 2005. *Historical Statistics of Housing in Britain*. Cambridge: Cambridge Center for Housing, and Planning Research.

Holston, Kathryn, Thomas Laubach, and John C. Williams. 2017. "Measuring the Natural Rate of Interest: International Trends and Determinants." *Journal of International Economics* 108 (S1): 59–75.

Homburger, Paul. 1905. *Die Entwicklung Des Zinsfusses in Deutschland Von 1870–1903*. Frankfurt am Main: Sauerländer.

Homer, Sidney, and Richard E. Sylla. 2005. *A History of Interest Rates*. English. 4th. Hoboken, N.J.: Wiley, xx, 710 p. :

Ilmanen, Antti. 2011. *Expected Returns: An Investor's Guide to Harvesting Market Rewards*. Hoboken, N.J.: Wiley.

International Labour Organization. various years. *Year-Book of Labour Statistics*. Geneva: International Labour Organization.

International Labour Organization. 2013. "Sources and Methods: Labour Statistics. Volume 1: Consumer Price Indices."

International Labour Organization. 2014. "LABORSTA Internet. Table 7F: Consumer Prices, Rent Indices."

International Labour Organization, International Monetary Fund, Organization for Economic Cooperation and Development, Statistical Office of the European Communities, United Nations, The International Bank for Reconstruction and Development, and The World Bank. 2004. *Consumer Price Index Manual: Theory and Practice*. Geneva: International Labour Organization.

Istat. 2016. "National Accounts, Final Consumption Expenditure of Households; Consumption of Fixed Capital by Industry."

Jones, Charles M. 2002. "A Century of Stock Market Liquidity and Trading Costs." Working paper. New York.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2015. "Betting the House." *Journal of International Economics* 96 (S1): 2–18.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2016. "The Great Mortgaging: Housing Finance, Crises and Business Cycles." *Economic Policy* 31 (85): 107–152.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." In *NBER Macroeconomics Annual 2016, Volume 31*. Edited by Jonathan A. Parker Martin Eichenbaum. Chicago, Ill.: University of Chicago Press, pp. 213–263.

Jurgilas, Marius, and Kevin J. Lansing. 2012. "Housing Bubbles and Homeownership Returns." *FRBSF Economic Letter* 2012-19:

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.

Klovland, Jan Tore. 2004a. "Bond Markets and Bond Yields in Norway 1820–2003." In *Historical Monetary Statistics for Norway 1819–2003. Norges Bank Occasional Paper No. 35*. Edited by Jan T. Klovland Øyvind Eitrheim and Jan F. Qvigstad. Chapter 4, pp. 99–181.

Klovland, Jan Tore. 2004b. "Historical Stock Price Indices in Norway 1914–2003." In *Historical Monetary Statistics for Norway 1819–2003. Norges Bank Occasional Paper No. 35*. Edited by Jan T. Klovland Øyvind Eitrheim and Jan F. Qvigstad. Chapter 8, pp. 329–349.

Knoll, Katharina. 2017. "Our Home in Days Gone By: Housing Markets in Advanced Economies in Historical Perspective." Doctoral dissertation. Free University of Berlin. Chapter 3. As Volatile As Houses: Return Predictability in International Housing Markets, 1870–2015.

Knoll, Katharina, Moritz Schularick, and Thomas M. Steger. 2017. "No Price like Home: Global House Prices, 1870–2012." *American Economic Review* 107 (2): 331–352.

Krishnamurthy, Arvind, and Tyler Muir. 2017. "How Credit Cycles Across a Financial Crisis." NBER Working Paper 23850.

KTI. 2015. "The Finnish Property Market 2015."

Kuczynski, Jürgen. 1947. *Die Geschichte der Lage der Arbeiter in Deutschland von 1800 bis in die Gegenwart: 1800 bis 1932*. Vol. 1, Berlin: Verlag die freie Gewerkschaft.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2017. "Income and Wealth Inequality in America , 1949–2013." CEPR Discussion Paper 20547.

Kurz, Claudia, and Johannes Hofmann. 2004. "A Rental-Equivalence Index for Owner-Occupied Housing in West Germany 1985–1998." *Deutsche Bundesbank Discussion Paper, Series 1: Studies of the Economic Research Center* 2004-08:

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2018. "The Big Bang: Stock Market Capitalization in the Long Run." EHES Working Paper 136.

Kuznets, Simon. 1941. *National Income and its Composition*. Cambridge, Mass.: National Bureau of Economic Research.

Le Bris, David. 2012. "Wars, Inflation and Stock Market Returns in France, 1870–1945." *Financial History Review* 19 (3): 337–361.

Le Bris, David, and Pierre-Cyrille Hautcoeur. 2010. "A Challenge to Triumphant Optimists? A Blue Chips Index for the Paris Stock Exchange, 1854–2007." *Financial History Review* 17 (2): 141–183.

Leeman, August. 1955. *De Woningmarkt in Belgie 1890–1950*. Kortrijk: Uitgeverij Jos. Vermaut.

Leroy-Beaulieu, Paul. 1906. *L'Art de Placer et Gerer sa Fortune*. Paris: Libraire Ch. Delagrave.

Lewis, Parry J., and Bernard Weber. 1965. *Building Cycles and Britain's Growth*. London, Macmillan.

Limburgsch Dagblaad. 1935. "Advertentie: Steenen Devalueeren Niet." *Limburgsch Dagblaad*, (222): 18, September 21, 1935.

López, Carlos Barciela, Albert Carreras, and Xavier Tafunell. 2005. *Estadıústicas Históricas De España: Siglos XIX–XX*. Madrid: Fundacion BBVA.

Lustig, Hanno, Stijn Van Nieuwerburgh, and Adrien Verdelhan. 2013. "The Wealth-Consumption Ratio." *Review of Asset Pricing Studies* 3 (1): 38–94.

Maluquer de Motes, J. 2013. "La Inflacion en Espana. Un Indice de Precios de Consumo, 1830–2012." *Estudios de Historia Economica* 64:

Marnata, F. 1961. *Les Loyers des Bourgeois de Paris, 1860–1958*. Paris: A. Colin.

Mediobanca. 2013. "La Borsa Italiana dal 1928."

Mediobanca. 2016. "Indici e Dati Relativi ad Investimenti in Titoli Quotati."

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15 (2): 145–161.

Mehrotra, Neil. 2017. "Debt Sustainability in a Low Interest Rate World." Hutchins Center Working Paper 32.

Meyer, Josefin, Carmen C. Reinhart, and Christoph Trebesch. 2015. "200 Years of Sovereign Haircuts and Bond Returns." Working Paper.

Ministry for Social Affairs. 1920–1929. *Social Tidskrift (various issues)*. Helsinki: Stadsradets Tryckeri.

Ministry of Land, Infrastructure, Transport, and Tourism. 2016a. "Tax System on Acquisition of Land."

Ministry of Land, Infrastructure, Transport, and Tourism. 2016b. "Tax System on Possession of Land."

Mitchell, B.R. 1988. *British Historical Statistics*. Cambridge: Cambridge University Press.

Mondani, A. 1978. "Aspetti Metodologici dell'indagine Mediobanca Sull'andamento dei Corsi e sul Movimento dei Capitali delle Società Quotate in Borsa Dal 1928 al 1977." *Risparmio*, 1566–84.

Moore, Lyndon. 2010a. "Financial Market Liquidity, Returns and Market Growth: Evidence from Bolsa and Börse, 1902–1925." *Financial History Review* 17 (1): 73–98.

Moore, Lyndon. 2010b. "World Financial Markets 1900–25." Working Paper.

Moskowitz, Tobias J., and Annette Vissing-Jørgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92 (4): 745–778.

MSCI. 2016. "Real Estate Analytics Portal."

Myrdal, Gunnar. 1933. *The Cost of Living in Sweden, 1830–1930*. London: P.S. King & Son.

Nakamura, Emi, Jón Steinsson, Robert Barro, and José Ursúa. 2013. "Crises and Recoveries in an Empirical Model of Consumption Disasters." *American Economic Journal: Macroeconomics* 5 (3): 35–74.

NERA Economic Consulting. 2015. "Historical Estimates of the Market Risk Premium."

Nielsen, Steen, and Ole Risager. 2001a. "Stock Returns and Bond Yields in Denmark, 1922–1999." *Scandinavian Economic History Review* 49 (1): 63–82.

Nielsen, Steen, and Ole Risager. 2001b. "Stock Returns and Bond Yields in Denmark, 1922–1999." *Scandinavian Economic History Review* 49 (1): 63–82.

Nieuwe Tilburgsche Courant. 1934. "Advertentie: Geldbelegging." *Nieuwe Tilburgsche Courant*, (11946): 56, March 31, 1934.

Nieuwe Tilburgsche Courant. 1936. "Advertentie: Geldbelegging." *Nieuwe Tilburgsche Courant*, 58, August 14, 1936.

Norwegian Tax Administration. 2016. "Municipal Property Tax."

Nyberg, Peter M., and Mika Vaihekoski. 2011. "Descriptive Analysis of Finnish Equity, Bond and Money Market Returns." BANK OF FINLAND DISCUSSION PAPER SERIES 14/2011.

Nyberg, Peter M., and Mika Vaihekoski. 2014. "Equity Premium in Finland and Long-Term Performance of the Finnish Equity and Money Markets." *Cliometrica* 8 (2): 241–269.

OECD. 2002. *Comparative Methodological Analysis: Consumer and Producer Price Indices*. MEI Methodoligical Analysis Supplement 2. Paris: OECD.

OECD. 2012. *OECD Economic Surveys: European Union 2012*. Paris: OECD Publishing.

OECD. 2016a. "OECD Fiscal Decentralization Database: Recurrent Tax on Immovable Property."

OECD. 2016b. "OECD Statistics. 5. Final Consumption Expenditure of Households."

OECD. 2016c. "OECD Statistics. Table 9B. Balance-Sheets for Non-Financial Assets."

Offer, Avner. 1981. *Property and Politics 1870–1914: Landownership, Law, Ideology, and Urban Development in England*. Cambridge: Cambridge University Press.

Office for National Statistics. 2016. "Changes to National Accounts: Imputed Rental."

Papadia, Andrea, and Claudio A. Schioppa. 2016. "Foreign Debt and Secondary Markets: The Case of Interwar Germany." unpublished.

Pedersen, Erik, and Jacob Isaksen. 2015. "Recent Housing Market Trends." *Danmarks Nationalbank Monetary Review*, (3): 51–62.

Pedersen, Jorgen. 1930. *Arbejdslonnen i Danmark*. Copenhagen: Gyldendalske Boghandel.

Peeters, Stef, Martine Goossens, and Erik Buyst. 2005. *Belgian National Income During the Interwar Period: Reconstruction of the Database*. Leuven: Leuven University Press.

Pictet & Cie. 1998. "The Performance of Shares and Bonds in Switzerland: An Empirical Study Covering the Years Since 1925."

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Harvard University Press.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics* 133 (2): 553–609.

Piketty, Thomas, and Gabriel Zucman. 2014. "Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010." *Quarterly Journal of Economics* 129 (3): 1255–1310.

Poullet, Gh. 2013. "Real Estate Wealth by Institutional Sector." *NBB Economic Review* Spring 2013: 79–93.

Poutvaara, Panu. 1996. "Pörssikurssien Kehitys Suomessa 1896–1929: Uudet Indeksisarjat Ja Niiden Tulkinta." Bank of Finland Discussion Paper.

Public Finances Directorate General. 2015. "Overview of the French Tax System."

Quinn, Dennis P., and Hans-Joachim Voth. 2008. "A Century of Global Equity Market Correlations." *American Economic Review* 98 (2): 535–540.

Rachel, Lukasz, and Thomas Smith. 2015. "Secular Drivers of the Global Real Interest Rate." Bank of England Working Paper 571.

Real, Werner Hermann. 1950. *Erfahrungen und Möglichkeiten bei der Aufstellung von Richtlinen für die Stadtplanung: Unter Besonderer Berücksichtigung der Verhältnisse in der Stadt Zürich*. Zürich: Eidgenössische Technische Hochschule.

Rees, Albert. 1961. *Real Wages in Manufacturing, 1890–1914*. Princeton, NJ: Princeton University Press.

Rognlie, Matthew. 2015. "Deciphering the Fall and Rise in the Net Capital Share." *Brookings Papers on Economic Activity* 46 (1): 1–69.

Ronge, Ulrich. 2002. *Die Langfristige Rendite Deutscher Standardaktien: Konstruktion eines Historischen Aktienindex ab Ultimo 1870 bis Ultimo 1959*. Frankfurt am Main: Lang.

Rosania, L. 1954. "Indice del Corso Secco e Rendimento dei Titoli Quotati in Borsa." *Banca d'Italia, Bollettino* 9: 539–71.

Ruff, Jon. 2007. "Commercial Real Estate: New Paradigm or Old Story?" *Journal of Portfolio Management*, 27.

Saez, Emmanuel, and Gabriel Zucman. 2016. “Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data.” *Quarterly Journal of Economics* 131 (2): 519–578.

Samy, Luke. 2015. “Indices of House Prices and Rent Prices of Residential Property in London, 1895-1939.” University of Oxford Discussion Paper in Economic and Social History 134.

Sefton, James, and Martin Weale. 1995. *Reconciliation of National Income and Expenditure: Balanced Estimates of National Income for the United Kingdom, 1920–1990*. Vol. 7, Cambridge: Cambridge University Press.

Segers, Yves. 1999. “De Huishuren in Belgie, 1800-1920. Voorstelling van een Databank.” *Centre for Economic Studies Discussion Paper Series* 1999-14:

Shiller, Robert J. 1981. “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71 (3): 421–436.

Shiller, Robert J. 2000. *Irrational Exuberance*. Princeton, N.J.: Princeton University Press.

Shilling, James D. 2003. “Is There a Risk Premium Puzzle in Real Estate?” *Real Estate Economics* 31 (4): 501–525.

Shimizu, Chihiro, Satoshi Imai, and Erwin Diewert. 2015. “Housing Rent and Japanese CPI: Bias from Nominal Rigidity of Rents.” *IRES Working Paper Series* 9:

Shinohara, Miyohei. 1967. *Estimates of Long-Term Economic Statistics of Japan Since 1868*. Edited by Kazushi Ohkawa, Miyohei Shinohara, and Mataji Umemura. Volume 6: Personal Consumption Expenditure. Tokyo: Tokyo Keizai Shinposha.

Shiratsuka, Shigenori. 1999. “Measurement Errors in the Japanese Consumer Price Index.” *Monetary and Economic Studies* 17 (3): 69–102.

Shumway, Tyler. 1997. “The Delisting Bias in CRSP Data.” *Journal of Finance* 52 (1): 327–340.

Shumway, Tyler, and Vincent A. Warther. 1999. “The Delisting Bias in CRSP’s Nasdaq Data and Its Implications for the Size Effect.” *Journal of Finance* 54 (6): 2361–2379.

Simonnet, François, Georges Gallais-Hamonno, and Pedro Arbulu. 1998. “Un Siècle de Placement Immobilier. L’exemple de La Fourmi Immobilière.” *Journal de la Société Française de Statistique* 139 (2): 95–135.

Skatteministeriet. 2016. “Ejendomsvaerdiskat og Ejendomsskat (grundskyld).”

Stapledon, Nigel D. 2007. “Long Term Housing Prices in Australia and Some Economic Perspectives.” Doctoral dissertation. Australian School of Business at the University of New South Wales.

Stapledon, Nigel D. 2012. “Historical Housing-Related Statistics for Australia 1881–2011 – A Short Note.” *UNSW Australian School of Business Research Paper* 2012-52 (12):

Statistics Amsterdam. 1916–1944. *Amsterdam Statistical Yearbook*. Amsterdam: J.M. Meulenhoff.

Statistics Belgium. 2013a. “Final Consumption Expenditure of Households (P.3), Estimates at Current Prices.”

Statistics Belgium. 2013b. “Huishoudbudgetonderzoek.”

Statistics Belgium. 2015. “Bouw en Industrie - Verkoop van Onroerende Goederen.”

Statistics Canada. 2015. “A Revision of the Methodology of the Rent Component of the Consumer Price Index (CPI), beginning with the July 2009 CPI.”

Statistics Copenhagen. 1906–1966. *Statistical Yearbook of Copenhagen and Frederiksberg*. Copenhagen: Gyldendalske Boghandel.

Statistics Denmark. 1919. *Vurderingen til Ejendomsskyld Pr. 1 Juli 1916*. Statistisk Tabelvaerk, 5. Raekke, Litra E 10. Copenhagen: Bianco Lunoc Bogtrykkeri.

Statistics Denmark. 1923. *Vurderingen til Eijendomsskyld Pr. 1 Juli 1920*. Statistisk Tabelvaerk, 5. Raekke, Litra E 12. Copenhagen: Bianco Lunoc Bogtrykkeri.

Statistics Denmark. 1925. *Statistisk Arbog*. Copenhagen: Gyldendalske Boghandel.
Statistics Denmark. 1948. *Vurderingen til Grundskyld Og Ejendomsskyld Pr. 1 Oktober 1945*. Statistisk Tabelvaerk, 5. Raekke, Litra E 21. Copenhagen: Bianco Lunoc Bogtrykkeri.
Statistics Denmark. 1954. *Vurderingen til Grundskyld Og Ejendomsskyld Pr. 1 Oktober 1950*. Statistisk Tabelvaerk, 5. Raekke, Litra E 23. Copenhagen: Bianco Lunoc Bogtrykkeri.
Statistics Denmark. 1969. "Kreditmarkedsstatistik." *Statistiske undersøgelser* 24:
Statistics Denmark. 2003. "Consumer price index (1980=100) by Commodity Group."
Statistics Denmark. 2015. "Consumer price index (2000=100) by Commodity Group and Unit."
Statistics Denmark. 2017a. "Annual National Accounts."
Statistics Denmark. 2017b. "Private Consumption (DKK Million) by Group of Consumption and Price Unit."
Statistics Finland. 1920. *The Republic of Finland: An Economic and Financial Survey*. Helsinki: Statistics Finland.
Statistics Finland. 2009. "Average Rents in 1964–2008."
Statistics France. 2015. "Banque de Donnees Macroeconomiques: Serie 000637649."
Statistics France. 2016a. "National Accounts. 6.461 Consumption of Fixed Capital at Current Prices (Billions of Euros)."
Statistics France. 2016b. "National Accounts. Actual Final Consumption of Households by Purpose at Current Prices (Billions of Euros)."
Statistics Germany. 1924–1935. *Statistisches Jahrbuch für das Deutsche Reich*. Berlin: Verlag von Puttkammer und Mühlbrecht.
Statistics Germany. various years. *Statistical Yearbook Germany*. Wiesbaden: Statistics Germany.
Statistics Germany. 1925. *Wirtschaft und Statistik*. Vol. 5, 5. Berlin: Reimar Hobbing.
Statistics Germany. 1934. *Vierteljahreshefte zur Statistik des Deutschen Reichs*. Vol. 43, 4. Berlin: Verlag für Sozialpolitik, Wirtschaft und Statistik G.m.b.H.
Statistics Germany. 2013. *Volkswirtschaftliche Gesamtrechnungen: Private Konsumausgaben Und Verfügbares Einkommen*. Beiheft zur Fachserie 18, 3. Vierteljahr 2013. Wiesbaden: Statistics Germany.
Statistics Japan. 2012. "Historical Statistics of Japan: Chapter 22 - Prices - Consumer Price Index."
Statistics Netherlands. 1959. "The Preparation of a National Balance Sheet: Experience in the Netherlands." In *The Measurement of National Wealth*. Edited by Raymond W. Goldsmith and Christopher Saunders. Income and Wealth Series VIII. Chicago, Ill.: Quadrangle Books, pp. 119–146.
Statistics Netherlands. 2010. "Rent Increase for Dwellings."
Statistics Netherlands. 2014. "Consumentenprijzen; Huurverhoging Woningen vanaf 1959."
Statistics Norway. 1949. *Statistical Survey 1948*. Oslo: H. Aschehoug & Co.
Statistics Norway. 1954. "Nasjonalregnskap 1938 Og 1948–1953."
Statistics Norway. 1978. *Historical Statistics 1978*. Oslo: Statistics Norway.
Statistics Norway. 1991. "Konsumprisindeksen." *Rapporter fra Statistisk Sentrabyra* 91/8:
Statistics Norway. 2014. "Annual National Accounts."
Statistics Norway. 2015. "Consumer Price Index for Goods and Services by Delivery Sector, 1979M01 - Latest Month."
Statistics Oslo. 1915. *Husholdningsregnskaper fort av Endel Mindre Bemidlede Familier i Kristiania, Bergen, Trondhjem, Drammen, Kristianssand of Hamar i Aaret 1912/13*. Oslo: J. Chr. Gundersens Boktrykkeri.
Statistics Spain. 2016. "Standard and Life Conditions (CPI)."

Statistics Sweden. 1933. *Detaljpriser och Indexberäkningar Aren 1913-1930*. Stockholm: A. B. Hasse W. Tullbergs Boktbyckeri.

Statistics Sweden. 1961. *Konsumentpriser och Indexberäkningar Aren 1931–1959*. Stockholm: Kungl. Sociastyrelsen.

Statistics Switzerland. 2014. "Haushaltungsrechnungen von Unselbständigerwerbenden: Ausgabenstruktur nach Sozialklassen 1912-1988 (ausgewählte Erhebungen)."

Statistics Switzerland. 2015. "Rent Index."

Statistics Zurich. 1946–1962. *Statistical Yearbook of Zurich*. Zurich: City of Zurich.

Summers, Lawrence H. 2014. "US Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound." *Business Economics* 49 (2): 65–73.

Swedish Tax Agency. 2012. "Taxes in Sweden: An English Summary of the Tax Statistical Yearbook of Sweden."

Tarbuck, Edward Lance. 1938. *Handbook of House Property: A Popular and Practical Guide to the Purchase, Mortgage, Tenancy and Compulsory Sale of Houses and Land, Including Dilapidations and Fixtures; with Examples of All Kinds of Valuations, Information on Building and on the Right Use of Decorative Art*. London: Technical Press.

Thwaites, Gregory. 2015. "Why are Real Interest Rates So Low? Secular Stagnation and the Relative Price of Investment Goods." Bank of England Working Paper 564.

Tilly, Richard H. 1986. "Wohnungsbauinvestitionen während des Urbanisierungsprozesses im Deutschen Reich, 1870–1913." In *Stadtwachstum, Industrialisierung, Sozialer Wandel: Beiträge zur Erforschung der Urbanisierung im 19. und 20. Jahrhundert*. Edited by H.-J. Teuteberg. Berlin: Duncker, and Humblot, pp. 61–99.

U.S. Bureau of Labor Statistics. 2015. "Consumer Price Index for All Urban Consumers: Rent of Primary Residence."

U.S. Census Bureau. 1942. *Statistical Abstract of the United States*. Washington, D.C.: US Government Printing Office.

U.S. Census Bureau. 2013. "Census of Housing, Tables on Gross Rents and Home Values."

Van den Eeckhout, P., and Peter Scholliers. 1979. *De Brusselse Huishuren: 1800–1940*. Lonen en Prijzen in Belgie in de 19e en 20e Eeuw. Brussels: VUB.

Van Fulpen, H. 1984. "Een International Vergelijking van Woonuitgaven." *Technische Hogeschool Delft Working Paper* 13:

Van Riel, A. 2006. "Constructing the Nineteenth-Century Cost of Living Deflator (1800-1913)."

Villa, Piere. 1994. *Un Siècle de Données Macro-Économiques*. INSEE résultats 86-87. Paris: INSEE.

Waldenström, Daniel. 2014. "Swedish Stock and Bond Returns, 1856–2012." In *Historical Monetary and Financial Statistics for Sweden, Volume 2: House Prices, Stock Returns, National Accounts and the Riksbank Balance Sheet, 1860-2012*. Edited by Tor Jacobson Rodney Edvinsson and Daniel Waldenström. Stockholm: Sveriges Riksbank, and Ekerlids förlag, pp. 223–293.

Waldenström, Daniel. 2017. "Wealth-Income Ratios in a Small, Developing Economy: Sweden, 1810-2014." *Journal of Economic History* 77: 285–313.

Whitehead, Christine, editor. 2012. *The Private Rented Sector in the New Century: A Comparative Approach*. Copenhagen: Boligokonimisk Videncenter.

Wissenschaftlicher Beirat beim Bundesministerium der Finanzen. 2010. "Reform Der Grundsteuer: Stellungnahme Des Wissenschaftlichen Beirats Beim Bundesministerium Der Finanzen."

Wüest, and Partner. 2012. "Immo-Monitoring 2012-1."

Zimmermann, Kaspar. 2017. “Breaking Banks? Bank Profitability and Monetary Policy.” unpublished.

Chapter 2

The Time Varying Risk Puzzle*

2.1 Introduction

Prices of risky assets fluctuate substantially over time. Much of this variation is difficult to square with changes in fundamentals such as dividends, rents or corporate default rates (Shiller, 1981; Greenwood and Hanson, 2013; Knoll, 2017). Instead, the dominant explanation for asset price volatility is time variation in discount rates – the idea that investors are sometimes less willing to bear risk, hence eliciting a low price of risky assets. Cochrane (2011) calls the understanding of why discount rates vary over time "the central organising question of current asset-pricing research".

This paper re-examines the contribution of time varying discount rates to asset price fluctuations using a new long-run dataset which covers three major risky asset classes – equity, housing and corporate bonds – in 17 advanced economies between 1870 and today. The importance of discount rate variation can be assessed by analysing the co-movement of expected returns – or "asset-specific discount rates" – across these three asset categories. Standard asset pricing theory stipulates that in the absence of arbitrage, prices of all assets should be proportional to a single discount factor. Time variation in this factor should, therefore, induce a positive co-movement in expected returns of different risky assets.

I show that this co-movement is absent in the data. Figure 2.1.1 displays the rolling decadal correlations between discount rate proxies for the three asset classes:

* This work is part of a larger project kindly supported by a research grant from the Bundesministerium für Bildung und Forschung (BMBF). I am grateful to Christian Bayer, Johannes Beutel, John Cochrane, Òscar Jordà, Moritz Kuhn, Gernot Müller, Björn Richter, Moritz Schularick, Alan Taylor, Kaspar Zimmermann, the participants of seminars in Bonn, ESWM 2018, the EAYE Workshop on Housing and Macroeconomy and the Bundesbank conference on Financial Cycles and Regulation for valuable comments. I would also like to thank numerous faculty who provided helpful comments during my job market interviews and flyouts. I would like to thank the researchers who helped with data for individual countries, Katharina Knoll who shared her code, and Francisco Amaral for helpful comments and help collecting the data. Felix Rhiel, Thomas Schwarz, Maira Talha and Yevhenii Usenko have provided invaluable research assistance.

the dividend-price ratio, rent-price ratio and the corporate bond spread. The correlation coefficients are roughly equal to zero and, for the most part, statistically insignificant.[1] For example, knowing that equity valuations are high tells us nothing about housing or corporate bond valuations.

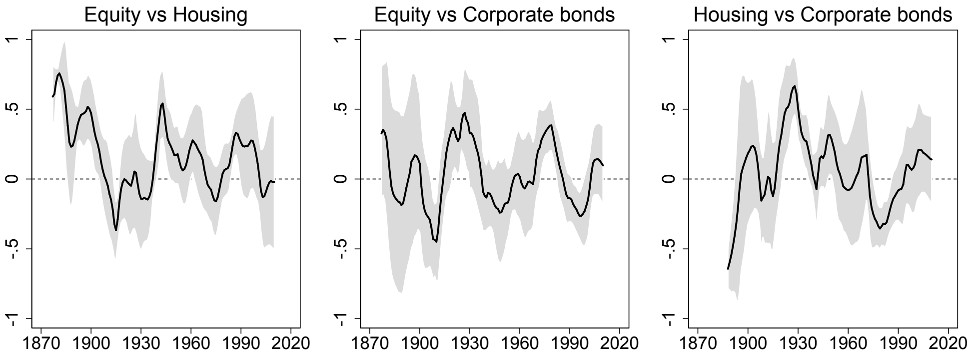
This lack of co-movement presents a new "time varying risk puzzle" for the asset pricing literature. Variation in the discount rate – or the "price of risk" – is central to most standard asset pricing theories including time varying risk aversion (Campbell and Cochrane, 1999; Piazzesi, Schneider, and Tuzel, 2007), disaster risk (Barro, 2006; Gabaix, 2012), long-run risk (Bansal and Yaron, 2004) and intermediary risk appetite (He and Krishnamurthy, 2013). All these factors are simply different proxies for the aggregate price of risk (Cochrane, 2017). The low co-movement fact is difficult to reconcile with these theories, and suggests that discount rate variation may not be the key driver of asset price volatility. This fact does, however, favour another set of theories which rely on volatile expectations rather than discount rates. This expectation volatility can come about from behavioural biases relating to extrapolation of asset-specific outcomes such as past returns (Barberis, Greenwood, Jin, and Shleifer, 2015; Adam, Beutel, and Marcet, 2017), return surprises (Bordalo, Gennaioli, La Porta, and Shleifer, 2017), or relative performance of an asset class (Barberis and Shleifer, 2003). Since there is little reason for such expectation biases – and the associated asset price variation – to co-move across asset classes, these theories offer a potential resolution to the puzzle.

Much of the paper is spent documenting the time varying risk puzzle more rigorously and searching for potential explanations. Because the discount rate proxies in Figure 2.1.1 depend on expectations of fundamentals as well as returns, the lack of co-movement could simply be attributable to idiosyncratic cashflows. I test for the relative importance of expected returns and cashflows using the standard framework of return predictability regressions (see Cochrane, 2008). If asset valuations predict future returns, they are partly driven by non-fundamental factors such as discount rates or expectations.

Within a single asset class, however, variation in discount rates and expectations is observationally equivalent (Cochrane, 2011). But the same is not true for multiple classes of risky assets. For example, equity price could be low either because investors are risk averse, or because they are overly pessimistic. But if investors are risk averse, house prices should also be low, whereas excessive pessimism about future dividends does not imply pessimism about future rents. To test for such cross-asset co-dependence, I test for predictive power of asset valuations and other macro-financial risk factors such as consumption and bank leverage *across* asset classes. If a discount rate proxy for one asset class predicts returns across the risky asset spec-

1. The correlation becomes even weaker when I consider changes in discount rates, rather than levels (Appendix Figure 2.A.1), and when using a cleaner discount rate measure (Section 2.6 Figure 2.6.1).

Figure 2.1.1. Comovement of discount rate proxies across asset classes



Note: Pairwise correlation coefficients between the dividend-price ratio, rent-price ratio and corporate bond spread over rolling decadal windows (e.g. the 1875 value covers the window 1870–1880). Shaded areas are 90% confidence intervals, using country-clustered standard errors. Underlying data are 3-year moving averages, to smooth over timing idiosyncracies across assets, demeaned at country level.

trum, it measures an aggregate factor such as the discount rate. If it does not, it measures asset-specific factors such as expectations.

After establishing the facts, I explore various theoretical explanations that could rationalise the lack of co-movement. These include variation in asset-specific risk, investor heterogeneity, market segmentation and volatility of expectations. The final part of the paper assesses whether these asset-specific factors still matter for real economic outcomes, despite being unrelated to aggregate risk appetite. To do this, I estimate the effect of asset-specific sentiment variation on future GDP, and analyse the links between systematic shifts in sentiment and financial crises.

To study these questions, I introduce a new long-run dataset of returns, cashflows and discount rates on all major risky asset classes – equity, housing and corporate bonds – in 17 advanced economies between 1870 and today. The comprehensive country and time dimension enables me to reliably distinguish variation in future cashflows from that in future discount rates or expectations, with these series observed over a horizon long enough to reliably approximate discounted present values, within a sample that is representative and covers less successful asset markets and disaster events.[2] The multi-asset nature of these data makes it possible to examine discount rate co-movement, and ultimately distinguish variation in discount rates from that in expectations. The dataset builds on the housing and equity data in the recent work of Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2017), and adds to it a new series of returns and spreads on corporate bonds. The new corporate bond data capture a market which accounts for a sizeable proportion of risky wealth, and is seen as an important indicator of financial crises and macroeconomic

2. See Stambaugh (1999) and Goyal and Welch (2008) for a further discussions on the need for representative long data samples in the return predictability regressions.

fluctuations (Biais and Green, 2007; Gilchrist and Zakrajšek, 2012; Krishnamurthy and Muir, 2017). For most of the 17 countries in my sample, this paper presents the first historical database of corporate bond spreads and returns, with most series constructed from new hand-collected archival sources.

It turns out that the discount rate proxies in Figure 2.1.1 do reflect variation in expected returns as well as fundamentals. Each of the three valuation measures – the dividend-price ratio, rent-price ratio and the corporate bond spread – predicts the future return on the respective asset class. The relationship is both economically and statistically significant, with a 1 percentage point increase in the asset-specific discount rate predicting roughly 1.5 ppts p.a. lower returns one year ahead, and 5 ppts lower cumulative returns 5 years ahead. The fact that high asset valuations are associated with low future returns means that investors either knowingly accept lower future payoffs because of a low discount rate, or are overoptimistic about future cashflows whenever valuations are high, with these expectations subsequently corrected through a fall in prices. This non-fundamental volatility is modest for equities, and strongest for the relatively unexplored asset categories of housing and corporate bonds. My analysis shows that the seminal "excess volatility" puzzle of Shiller (1981) is a salient feature of risky asset markets across 17 advanced economies and three major asset classes, over the last century and a half.

This paper's central finding is that discount rates of different asset classes do not co-move. Put differently, what looks like discount rate variation for a single asset class cannot actually be interpreted as time-varying discount rates from the broader cross-asset perspective. To test for the extent of co-movement, I first construct a clean discount rate proxy for each asset class, equal to the discount rate news component of unexpected returns (Campbell, 1991).[3] I find that the correlation between discount rate news for different asset classes is near zero and not statistically significant, both in the full sample and across different countries, subsamples and data definitions. Cashflow news, on the contrary, show a significantly positive correlation. This turns the notion of idiosyncratic risk on its head. Asset pricing models typically assume that cashflows have a strong asset-specific component, while discount rates on all assets are driven by a single common factor. The reality is the opposite: it is asset-specific cashflows that co-move, and discount rates that do not. The benefits of diversification across different asset classes – a standard finding in finance – then also largely come about from hedging against asset-specific discount rate changes, rather than fundamentals.

The low co-movement finding holds within the more formal framework of cross-asset predictability regressions. The discount rate proxies for one asset class do not predict returns on other assets. For example, the dividend-price ratio does not predict future housing or corporate bond returns. The same is true for other macro-

3. Campbell (1991) uses a VAR to estimate the discount rate component for equities, and I extend the methodology by employing a similar technique for housing and corporate bonds.

financial factors which have been used as proxies for the discount rate in the literature. I test the cross-asset predictive power of 8 macro-financial variables relating to consumption, such as surplus consumption (Campbell and Cochrane, 1999) and the consumption-wealth ratio (Lettau and Ludvigson, 2002); financial intermediary risk appetite, such as bank leverage and asset growth (Baron and Muir, 2018); and other empirical predictors of equity returns such as the term spread (Campbell, 1991). None of these variables predict returns across all three asset categories. The absence of cross-asset return predictability, and the low predictive power of macro-financial risk factors hold across different time periods, data definitions and empirical specifications. For example, they hold for five-year-ahead returns, nominal and real returns, when limiting the sample to recent decades, and both during economic recessions and expansions.

Taken together, the absence of co-movement in expected returns across asset classes is a robust and persistent feature of the data. This feature poses a challenge to discount rate based explanations of financial volatility, regardless of whether this discount rate variation is ultimately driven by time varying risk aversion, disaster risk or intermediary risk appetite. I consider several potential modifications which could improve the fit between the theory and these new data, relating to asset-specific frictions, investor heterogeneity and volatility of expectations.

Variation in asset-specific risk, or non-monetary payoffs not captured in my cash-flow measure – such as liquidity services – can delink the movements in asset-specific expected returns from those in the aggregate discount factor. For example, if housing becomes more risky or less liquid relative to equities, expected returns on housing should increase relative to those of equities even in the presence of a common pricing kernel. To explain the puzzle, the time variation in these frictions needs to be asset-specific and quantitatively large. The data, however, only suggest a modest degree of such variation. For example, I show that a proxy for asset riskiness – the covariance between each risky return and the macro discount factor – is relatively stable over time and across asset classes.

If markets are segmented and investors in these markets are highly heterogeneous, the pricing kernel in each asset class will be asset-specific, offering another potential resolution to the puzzle. But the extent of both segmentation and heterogeneity in the data is relatively modest. While several studies have documented existence of segmentation and cross-asset arbitrage opportunities, their quantitative importance is typically an order of magnitude smaller than the expected return differentials in my data.[4] The extent of investor discount rate heterogeneity also appears limited, with a large share of both bond equity securities held by institutional

4. For example, Fleckenstein, Longstaff, and Lustig (2010) document a 20 bps arbitrage between TIPS and government bonds plus an inflation hedge during normal times, whereas a one standard deviation higher dividend-price ratio in my data predicts 150 bps higher relative returns one year ahead and 500 bps higher returns five years ahead.

investors (Gompers and Metrick, 2001; Biais and Green, 2007), and discount rate proxies for housing and equity investors showing high co-movement.[5]

Expectation volatility can, on the other hand, not only generate the low co-movement in theory, but also match several other salient features of the data. To test for evidence of extrapolative and style investing behaviour, I construct a proxy for sentiment – or expected return – on each asset class following the methodology of López-Salido, Stein, and Zakrajšek (2017). The behaviour of this sentiment variable is consistent with return extrapolation: a high, or higher than expected return predicts an increase in sentiment for the specific asset class one year ahead. It is also consistent with investors switching between styles, with an increase in sentiment on one asset class predicting lower sentiment on other asset classes.

Despite not being driven by aggregate risk appetite, this asset-specific expectation volatility has real effects. Elevated sentiment in year t predicts low GDP growth in years $t+1$ to $t+5$, with the effect strongest and most persistent in the housing and corporate bond markets. Asset-specific sentiment also displays systematic variation around the onset of financial crises. An average crisis event is preceded by elevated sentiment – or "froth" – in equity or corporate bond markets, followed by sharp sentiment reversals when the crisis starts. Housing sentiment peaks after the crisis and slowly unwinds thereafter. But even during financial crises, sentiment fluctuations remain uncorrelated across asset classes, suggesting that they are, again, driven by expectations rather than a discount rate. The overall pattern in the data fits that of a Minsky (1977) financial instability cycle, with elevated sentiment accompanying an economic boom, followed by an eventual bust once the optimistic expectations are reversed and sentiment unwinds.

My findings complement the following three strands of existing literature. The first strand relates to the study of return predictability and price volatility for individual asset classes. A large literature, starting with Shiller (1981), has documented that US equity prices are excessively volatile, and returns – predictable (see Cochrane, 2008, for a summary). Knoll (2017) studies housing return predictability in a historical sample similar to this paper, and Greenwood and Hanson (2013) show that corporate bond returns in the US can be predicted by past issuance. I show that the evidence for predictability is both much more extensive, stretching to multiple countries and risky asset classes, and much more pervasive in the less studied markets of housing and corporate bonds.

The second strand relates to the study of return and discount rate co-movement across asset classes. Shiller (1982) proposed that the divergent patterns in US equity, housing and corporate bond prices pose a challenge to discount rate theories, but lacked the necessary data to formally test this hypothesis. Since then, several

5. More precisely, housing is primarily held by the middle class, while equities are held by the top of the income distribution (Kuhn, Schularick, and Steins, 2017) – but the income growth of these two income groups in the US data is strongly positively correlated.

studies have reported a positive co-movement between or within expected returns on equities, corporate and government bonds in the post-1925 US data. Fama and French (1989) have documented this co-movement for corporate bonds and equities, Campbell and Ammer (1993) – for equities and government bonds, and Cochrane and Piazzesi (2005) – across different maturities of government bonds. A number of recent papers in the intermediary asset pricing literature find that proxies for intermediary risk appetite can predict returns on equities, bonds and foreign exchange (He, Kelly, and Manela, 2017; Baron and Muir, 2018; Haddad and Muir, 2018).

Despite the above evidence, the extent of bond and equity discount rate co-movement – even in the recent US data – remains subject to debate. In recent work, Giglio and Kelly (2018) and Haddad, Kozak, and Santosh (2017) show that the variation in expected returns across a wide range of financial securities cannot be explained by a single discount factor. Further to this, even though several papers have studied the co-movement of *realised* returns on equity and housing (Ibbotson and Siegel, 1984; Gyourko and Keim, 1992; Liu and Mei, 2003), there is, to the best of my knowledge, no evidence on the co-movement in *expected* returns across housing and other risky asset classes. My paper is the first study discount rate co-movement across all major risky asset categories, or in cross-country long-run data. The cross-asset discount rate correlation turns out to be very low – even lower than that in returns or cashflows – a puzzling new stylised fact which has deep implications for asset pricing theory.

The third strand of the literature relates to real effects of time varying sentiment. Most studies have focussed on the US corporate bond market, and find that increasing excess bond premiums or predictable reversals in bond sentiment forecast low GDP growth (Gilchrist and Zakrajšek, 2012; López-Salido, Stein, and Zakrajšek, 2017). Muir (2017) and Krishnamurthy and Muir (2017) find that equity and corporate bond premiums tend to spike around financial crises, using historical data for a number of advanced economies. There is little agreement, however, about whether these bond spread movements and the associated real effects can be attributed to changing credit supply or sentiment. My paper shows that these changes in spreads are best thought of as time-varying expectations or sentiment, that the real effects of this sentiment variation are present in the 150-year cross section of advanced economies and asset classes, and that housing sentiment matters at least as much as that in the corporate bond market.

Taken together, my findings have the following implication for existing literature. Asset price volatility does not, primarily, come about from volatile discount rates. Rather, it comes about from volatile expectations. It is the understanding of these volatile expectations which could form the central organising question of future asset pricing research.

2.2 Testing for volatile discount rates

The central question addressed by this paper is whether asset price volatility is driven by time varying discount rates or not. To delineate the different determinants of asset price variation, consider a risky asset *i* which yields a stream of stochastic payoffs or cashflows CF_t. These cashflows correspond to dividend payments for equity, rental payments for housing and coupon payments for bonds.[6] To make the marginal investor indifferent to owning the asset, the asset price should equal the present value of these payoffs:

$$P_{i,t} \;=\; \mathbb{E}_t\left(\sum_{s=1}^{\infty} m_{t+s} CF_{i,t+s}\right) \tag{2.2.1}$$

The stochastic discount factor m, inversely proportional to the discount *rate DR*, captures investor willingness to save in risky assets, and is a combination of the ex ante safe rate – measuring the general willingness to save – and the risk premium – measuring the willingness to accept uncertain payoffs on savings.[7] Most asset pricing theories work with a recursive version of equation (2.2.1), which takes the form of:

$$1 \;=\; \mathbb{E}_t\left(m_{t+1} R_{i,t+1}\right) \tag{2.2.2}$$

where R is period-ahead total return, $R_{i,t+1} = (P_{i,t+1} + CF_{i,t+1})/P_{i,t+1}$. This "fundamental asset pricing equation" allows to classify the drivers of asset price volatility into the following three categories:

1: Time-varying fundamentals. or changes in future cashflows *CF*. The cashflows on each asset class in my study are by nature risky: dividends and rents can vary from year to year, and corporate bonds can default, ending the coupon payments. Fluctuations in future cashflows should be reflected in asset values. For example, if corporate profitability declines, dividends should fall and default risk should rise, reducing the value of equity and corporate bonds. Under this explanation, discount rates are constant, $m_{t+1} = M \;\forall\; t$, and hence so are expected returns:

$$1 \;=\; \mathbb{E}_t(MR_{i,t+1}) \tag{2.2.2F}$$

6. For owner-occupied housing, the payoff stream instead corresponds to "imputed rents", which capture the flow value of housing services, approximated by the hypothetical rental income the home-owner would have earned, had the house been placed on the rental market.

7. Both in my data and in the existing literature on the US equity returns, changes in safe interest rates have a relatively limited effect on the cyclical variation in risky asset prices. Therefore, for the purposes of my analysis it is convenient to think of variations in m as being mainly driven by the risk premium.

2: Time-varying discount rates. or changes in m. The discount rate m_{t+1} reflects the desire to hold risky assets and accept uncertain future payoffs. Put differently, it measures the price of risk. When the desire to save in risky assets is low, prices of these assets will also be low to incentivise investors to hold them. To rule out arbitrage, the discount factor m has to be the same across all risky assets (Ross, 1977). In the simple consumption asset pricing model, m is equal to the marginal rate of substitution between consuming today and tomorrow, $u'(c_{t+1})/u'(c_t)$. When consumption is high, investors are willing to save and take risks, and prices of risky assets are high.[8] Recent theories have proposed a number of modifications to define "good" and "bad" times – periods when the demand for risky savings is, respectively, low or high (see Cochrane, 2017, for a summary). These give m a more general formulation, such that it depends not only on consumption, but also on an another economic state variable Y_t, such as habit (Campbell and Cochrane, 1999), disaster risk (Barro, 2006; Gabaix, 2012) or financial intermediary risk appetite (He and Krishnamurthy, 2013):

$$1 = \mathbb{E}_t(\underbrace{Y_t * u'(c_{t+1})/u'(c_t)}_{m_{t+1}} R_{i,t+1}) \tag{2.2.2DR}$$

3: Time-varying sentiment. or changes in $\mathbb{E}$. If agents form their expectations according to some statistical operator $\mathbb{E}^* \neq \mathbb{E}$, prices of risky assets can fluctuate even in the absence of shocks to future cashflows *CF* or the price of risk m. In line with the literature (Baker and Wurgler, 2000; Greenwood and Hanson, 2013), I label such deviations in expectations from their rational value as "investor sentiment". If sentiment is elevated, investors are overoptimistic and $\mathbb{E}_t^* > \mathbb{E}_t$, hence asset prices will be high.[9]

Differences in aggregate sentiment are indistinguishable from changes in discount rates in my data. If expected returns on all three risky asset classes are high, this could be either because m is high, or $\mathbb{E}^*$ is high. But in most theoretical models, sentiment is an asset-specific phenomenon, driven by extrapolation of past returns and fundamentals, or forecasts that overweigh certain asset-specific outcomes (Barberis and Shleifer, 2003; Greenwood and Hanson, 2013; Barberis, Greenwood, et al., 2015; Adam, Beutel, and Marcet, 2017; Bordalo et al., 2017). This means

8. This can be seen more clearly by reformulating equation (2.2.2) as $\mathbb{E}_t(R_{i,t+1}) = (1 - cov(R_{i,t+1}, m_{t+1})/\mathbb{E}_t(m_{t+1})$. Risky asset returns tend to be high during good times, when m is low, hence $cov(R_{i,t+1}, m_{t+1}) < 0$. Hence, periods of low m are associated with high expected risky asset returns $E(R)$.

9. Note that the variation in sentiment includes "irrational" bubbles driven by biases in investor expectations. Strictly speaking, "rational" bubbles would appear as an extra term on the right-hand side of equation (2.2.1), equal to the discounted value of the terminal price of asset i, but given that the variation in terminal prices appears to be relatively limited in the data, and to ease the exposition, in this paper I interpret both irrational and rational bubbles as changes in sentiment.

that there is no reason for expected returns across assets to co-move positively, or at all. Sentiment-based explanations imply the following reformulation of equation (2.2.2):

$$1 = \mathbb{E}^*_{i,t}\left(m_{t+1}R_{i,t+1}\right) \tag{2.2.2SENT}$$

Each of the three explanations above carries a testable hypothesis about the time variation in expected returns on different asset classes, $\mathbb{E}_t(R_{i,t+1})$, summarised below.

Hypothesis 1. Fundamentals: Expected returns on all asset classes are constant.

Hypothesis 2. Discount rates: Expected returns on all asset classes vary over time, and exhibit high cross-asset co-movement.

Hypothesis 3. Sentiment: Expected returns vary over time, but expected returns on different asset classes can show zero or negative co-movement.

Within a single asset class, variation in discount rates and expectations is observationally equivalent. When it comes to multiple asset classes, discount rate variation induces co-movement in expected returns on different assets whereas expectation variation generally does not. To test for, first, the time variation in expected returns (hypothesis 2) and, second, for the co-movement in expected returns (hypothesis 3) in the data, I rely on the standard framework of return predictability regressions, extended somewhat to cover multiple classes of risky assets. Campbell and Shiller (1988) show that the discount rate $DR = 1/m$ can be approximated by the ratio of fundamentals CF to asset prices P. This "inverse" valuation ratio, dp, approximately equals the present value of expected returns and cashflows:

$$dp_{i,t} \approx \mathbb{E}\sum_{s=0}^{\infty}\rho_i^s r_{i,t+1+s} - \mathbb{E}\sum_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s} \tag{2.2.3}$$

Here, dp are (inverted) asset valuations, such as the equity dividend-price ratio, r are holding period returns, and dg is the growth in cashflows for each asset class, all expressed in logs:

$$\begin{aligned}
dp_{i,t} &= log\left(CF_{i,t}/P_{i,t}\right) \\
r_{i,t+1} &= log\left((P_{i,t+1} + CF_{i,t+1})/P_{i,t}\right) - log(1 + \pi_{t+1}) \\
dg_{t+1} &= log\left(CF_{i,t+1}/CF_{i,t}\right) - log(1 + \pi_{t+1}),
\end{aligned}$$

Equation (2.2.3) holds, up to a small approximation, for equity and housing, with cashflows measured as dividends or rental income. It holds somewhat more approximately for bonds, with r the excess return over government bonds, dp the price premium compared to government bonds, and CF – the default-adjusted expected coupon payments (Nozawa, 2017).

If asset prices vary because of future cashflows, valuations *dp* should predict future cashflow growth *dg*. If they vary because of expected returns, *dp* should be predictive of future returns *r*. If time varying discount rats are important, this expected return variation should be common to all risky asset classes: i.e. valuations of one asset class should predict returns on all three risky assets. Both of these propositions can be tested by running return predictability regressions. To test for the importance of fundamentals and expected returns, I regress the return and cashflow growth of asset i at $t+1$ on the valuation of that asset at t:

$$r_{i,j,t+1} = \beta_{i,j,1} + \beta_{i,2} dp_{i,j,t} + u_{i,j,t} \tag{2.2.4}$$

$$dg_{i,j,t+1} = \gamma_{i,j,1} + \gamma_{i,2} dp_{i,j,t} + e_{i,j,t}, \tag{2.2.5}$$

where i is an index which denotes different asset classes. $j = \{1, 17\}$ is a country index, and $t = \{1870, 2015\}$ denotes the year of the observation. The return predictability regressions, with some straightforward modifications, also allow me to compute proxies for asset-specific expected returns $\mathbb{E}\sum_{j=0}^{\infty} \rho^j r_{i,t+1+j}$, or "discount rate news" in each year of the cross-country panel dataset, and estimate their contribution to the asset price volatility (i.e., variation in *dp*) and co-movement across asset classes.

To assess the importance of discount rates and expectations, I test whether valuations of asset class i predict not only returns on asset i, but also on the other two asset classes j and k:

$$r_{i,j,t+1} = \beta_{i,j,1} + \beta_{k1\neq i} dp_{k1,j,t} + \beta_{k2\neq i} dp_{k2,j,t} + u_{i,j,t} \quad i = \{eq, hous, bond\} \tag{2.2.6}$$

Finally, the absence of cross-asset predictability may arise because asset-specific valuations *dp* are a poor proxy for the aggregate discount rate. To check whether this is the case, I also test for predictive power of a broad range of macro-financial risk factors F, mostly related to consumption and financial intermediary balance sheets:

$$r_{i,j,t+1} = \beta_{i,1} + \sum_{f=1}^{F} \beta_{f,i} F_{j,t} + e_{i,t} \quad i = \{eq, hous, bond\} \tag{2.2.7}$$

The tests of hypotheses 1–3 can then be summarised as follows:

Test 1. Fundamentals vs expected returns: return predictability regressions with one risk factor. If $\beta_{i,2} > 0$ in regression (2.2.4), expected returns for asset i vary over time.

Test 2. Importance of discount rates: return predictability regressions with cross-asset risk factors. If valuations on one asset i, or a macro-financial risk factor F, predict returns on all three risky asset classes, the variation in the aggregate price of risk contributes to time varying expected returns. If no single asset valuation or risk factor predicts returns on all three asset classes, the expected return variation is driven by asset-specific factors such as sentiment.

Testing these hypotheses requires data on the discount rate proxies, or asset valuations *dp*, risky returns *r* and cashflow growth *dg*. I describe the construction of the corresponding dataset in the following section.

2.3 New data on historical returns and discount rates

The dataset in this paper consists of valuations, cashflows and returns on the three major risky asset-classes: equity, housing and corporate bonds. The data are annual, and cover 17 advanced economies over the period 1870 to 2015. Table 2.3.1 summarises the data coverage by country and asset class. For most countries and series, the data go back to the late 19th century. The corporate bond data, by their nature, have slightly lower coverage, because in a number of countries, during some historical periods the private unsecured corporate debt markets were more or less non-existent.[10] My study is the first to document the joint historical evolution of corporate bond spreads and returns for every country in the sample apart from the US.[11]

Because asset prices depend on the value of all future fundamentals, and predictability tests lack statistical power and representativeness over small time or country samples (Stambaugh, 1999; Goyal and Welch, 2008; Lettau and Van Nieuwerburgh, 2008), the long time and country dimension of the dataset is an essential prerequisite for the statistical analysis described in Section 4.3. The novel cross-asset dimension of the dataset also enables me to differentiate the time variation in discount rates from that in asset-specific expectations. I next describe the sources for each series.

Equity. The data consist of total returns, dividends, and dividend-price ratios of listed equities, all taken from Jordà, Knoll, et al. (2017), with the addition of a new data series covering Canada. The return, price and dividend data mostly consist of value-weighted all-share indices. The dividend-price ratio is computed as dividend income over the course of the year in proportion to the year-end share price. Returns are a sum of capital gain and dividend income, in proportion to previous year's price, net of inflation. Cashflows are measured as dividend growth in a given year, in excess of inflation. For more details on the sources and accuracy checks, see Jordà, Knoll, et al. (2017).

10. For example, in Denmark, all the private sector bonds on the market, until the recent decades, consisted of mortgage bonds, which were considered safe assets on par with government bonds, and were issued under a strict set of regulations.

11. Greenwood and Hanson (2013) study the joint evolution of US corporate bond spreads and returns between 1926 and 2008. Compared to their study, my paper adds roughly another 60 years of data.

Table 2.3.1. Data coverage

Country	Equity	Housing	Corporate Bonds
Australia	1870–2016	1901–2015	1915–2016
Belgium	1870–2015	1890–2015	1870–2016
Canada	1870–2015		1905–2015
Denmark	1872–2016	1875–2015	
Finland	1912–2012	1920–2015	1960–2015
France	1870–2016	1870–2015	1913–2016
Germany	1870–2016	1870–2015	1870–2016
Italy	1870–2015	1927–2015	1873–2016
Japan	1886–2015	1931–2015	1900–2016
Netherlands	1900–2015	1870–2015	1975–2003
Norway	1880–2016	1871–2015	1903–2004
Portugal	1870–2015	1948–2015	1905–1988
Spain	1899–2016	1900–2015	1913–2016
Sweden	1871–2012	1883–2015	1871–1990
Switzerland	1900–2015	1901–2015	1935–2002
UK	1871–2015	1895–2015	1870–2014
USA	1871–2015	1890–2015	1870–2015

Housing. The data consist of total returns, rents, and rent-price ratios for residential real estate, all taken from Jordà, Knoll, et al. (2017). The return, price and rent data are constructed to, wherever possible, cover both owner occupiers and renters, cover the national housing stock, and adjust for quality changes, maintenance costs, depreciation and other non-tax housing expenses. The rent-price ratio is calculated as net rent received over the course of the year in proportion to the house price. Total return is the sum of capital gain and rental income, and cashflows are measured as rental growth, both net of inflation. For more details on the sources and accuracy checks, see Jordà, Knoll, et al. (2017).

Corporate bonds. This paper introduces a dataset of yields, spreads, and holding period returns on bonds issued by private sector creditors, targeting 10-year maturity. Corporate bond valuations *dp* are proxied by the spread, which equals the yield to maturity differential between corporate and government bonds:

$$\text{spread}_t \;=\; YTM_{\text{corporate},t} - YTM_{\text{government},t} \tag{2.3.1}$$

The yield to maturity measures the implicit discount rate which would make the present value of future coupon payments equal the observed bond price, and the spread, therefore, measures the forward-looking risk premium embedded in the prices of corporate and government issued securities.

I construct two measures of the corporate bond return r. The first measure is the holding period return, which is the sum of capital appreciation ΔP and coupon payments C received during year t, in proportion to the previous year's bond price:

$$r^{holding}_{bond,t+1} = C_{bond,t+1}/P_{bond,t} + (P_{bond,t+1} - P_{bond,t})/P_{bond,t} \tag{2.3.2}$$

Because of a lack of secondary sources for corporate bond return data, I sometimes estimate the price change from the change in yields using duration approximation. Results are, however, robust to only using non-approximated bond returns. The second bond return measure does not rely on such approximation, and simply uses the change in spreads as a proxy for returns:

$$r^{implied}_{bond,t+1} = -(\text{spread}_{t+1} - \text{spread}_t) \tag{2.3.3}$$

The $r^{implied}_{bond,t+1}$ proxies both the direction and magnitude of the bond price change, and is commonly used in the literature (López-Salido, Stein, and Zakrajšek, 2017). Comparing the current spread to future spread growth also allows to directly estimate the degree of predictable mean reversion in bond prices, with this mean reversion equivalent to expected return variation in the Campbell-Shiller decomposition equation (2.2.3). I use $r^{holding}_{bond,t+1}$ to summarise the moments of bond returns, and use $r^{implied}_{bond,t+1}$ as the main variable for the regression analysis. The use of either measure has no material bearing on the results.

Most of the corporate bond data were constructed from primary sources, by aggregating yields and returns on individual bonds listed on the domestic stock exchange. The bulk of the data comes from domestic stock exchange listings, complemented with bonds listed on major foreign exchanges (e.g. London and New York) and over the counter transactions. I weight the average by the market capitalization of individual bonds – unless these data are missing or the sample size is small, in which case I use equal-weighted averages to avoid biasing the series towards any individual bond. The individual listings data are complemented by a rich selection of secondary sources from publications of statistical agencies, international organisations and central banks, as well as financial history books and research articles.

The corporate bond sample covers all private sector fixed-rate bonds traded on the secondary market of the respective country with a maturity close to 10 years. I exclude foreign company bonds, foreign currency bonds of domestic companies, bonds with explicit government guarantees, and mortgage bonds issued by credit institutions or special purpose vehicles.[12] For some historical periods, most listed bonds had very long maturities, or there were relatively few bonds listed and traded. In these cases I extend the maturity window, sometimes including all private sector listed bonds, in order to obtain a comprehensive sample coverage. For periods where

12. For most of my sample, mortgage bonds were government guaranteed or strictly regulated, and generally considered safe assets on par with government debt. I do, however, include private non-financial company bonds backed by mortgages or property, which were generally not government-guaranteed, not strictly regulated, and considered risky.

secondary markets were thin but primary markets were active, I rely on issue yields instead of secondary market yields. Where maturity data are missing, I use current yields – the ratio of coupon to bond price – instead of yields to maturity. The government bond yield data are an extended and updated version of those in Jordà, Knoll, et al. (2017). The government bond dataset also excludes foreign currency bonds and targets a maturity of 10 years.

The extensive use of these new and previously uncovered data sources allows me to guard against a number of biases in estimating a consistent corporate bond yield series. The potential biases and accuracy issues for the corporate bond series are discussed in Appendix 2.A.1. The most pressing bias refers to the time variation in selection and credit quality of the corporate bond index, and is largely dealt with by using a large sample of bonds, comparing the bond yields on different bonds within the microdata, and tests for subperiod stability and robustness of results.

Comparison to existing datasets. The data in this paper have a much broader cross-asset coverage compared to existing studies. They also significantly extend the return predictability and excess volatility analysis for individual asset classes. The notable exception is the PhD thesis of Knoll (2017) analyses excess volatility in housing markets within a historical dataset similar to this paper. For equities, Engsted and Pedersen (2010) assess return predictability using data for 5 countries' going back to the early 20th century. Muir (2017) documents dividend-price ratio movements around financial crises and deep recessions in a cross-country historical dataset sourced largely from *Global Financial Data*.

Most of the excess volatility analysis for corporate bonds relies on historical or contemporary US data (Greenwood and Hanson, 2013; López-Salido, Stein, and Zakrajšek, 2017). Muir (2017) and Krishnamurthy and Muir (2017) document corporate spread movements around financial crises in a broad historical cross-country dataset, and use these to shed light on the fundamental asset price determinants. The historical data in Krishnamurthy and Muir (2017) largely consist of foreign bonds listed on the London stock exchange over the period 1869 – 1929. The focus on domestic exchanges in my data allows for a considerably broader selection of bonds, and guards against a number of potential selection biases discussed above.[13] My data also have a broader time and country coverage, and provide estimates of corporate bond returns and yields, as well as the spread.

These new data mark a significant step forward in documenting the returns and riskiness of different classes of risky wealth. Going beyond US equities market reduces the selection bias of this relatively successful asset market. Adding housing and corporate bonds documents the returns and risks on the largest component

13. Foreign listings typically only include the largest companies in the country, with many of the international bond issues also guaranteed or backed by the domestic government. Bonds listed on foreign exchanges were also often denominated in foreign currency.

of household wealth (residential real estate), and the most macro-informative asset (corporate bonds). The presence of both housing and corporate bond data also allows me to narrow down the range of potential explanations behind the low comovement. For example, since corporate bonds and equities are traded on the same market while housing is traded over the counter, market segmentation could delink the expected returns on housing and equities, but not on equities and corporate bonds. The risky asset classes which are not included the analysis consist of commercial property, agricultural land, unlisted equity and business capital. Housing and equity return data provide rough proxies for these missing asset classes, but the more detailed analysis is left to future research.

2.4 Stylised facts

Before proceeding to the formal analysis of return predictability regressions, this section documents a number of facts relating to the volatility of asset prices in these new historical data, and its potential underlying drivers.

Fact 1: Each risky asset earns a positive excess return premium. Risky assets should earn a positive return premium to compensate investors for the uncertain nature of the cashflows. This is, indeed, the case for each of the three asset classes in my dataset.

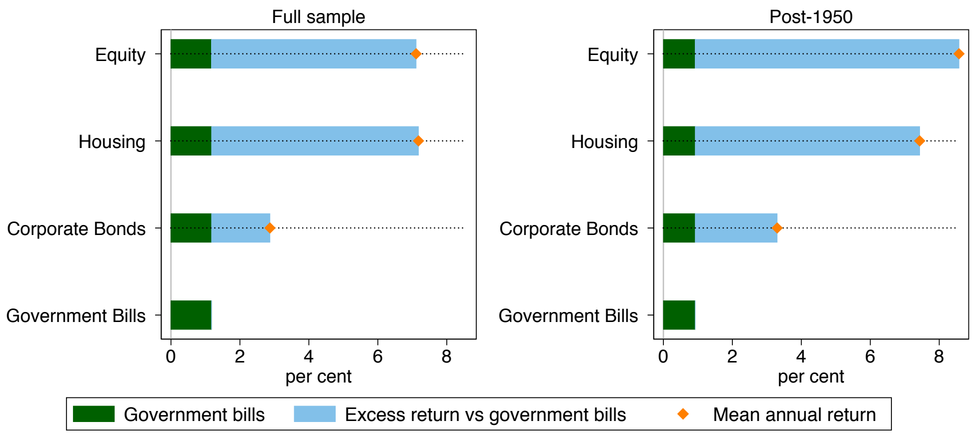
Figure 2.4.1 and Table 2.4.1 Panel 1 show the mean annual return for each asset class compared to a safe rate benchmark, taken to be the short-term government bill rate. Excess returns are positive for all three asset classes, in the full sample and after 1950. They are relatively high for housing and equity – around 6% p.a. – and moderate for corporate bonds, at around 2% p.a. Since all three risky assets have long durations, part of these excess returns actually consists of a term premium, which amounts to around 1% p.a. in these historical data.[14] After 1950, risky returns on corporate sector assets – equity and corporate bonds – are somewhat higher than the full-sample averages, while housing returns are similar to the full sample (Figure 2.4.1, right-hand panel).

Fact 2: Returns on all three risky asset classes are volatile. One reason why these assets command a risk premium is that their cashflows are uncertain, and returns – volatile. Table 2.4.1 Panel 1 shows that the annual standard deviation of real returns is 21 ppts p.a. for equities, and 10–11 ppts p.a. for housing and corporate bonds. A two-sigma return movement would, thus, lead to equity returns of -35%, housing returns of -13%, and corporate bond returns of -20%.

Figure 2.4.2a shows that risky asset returns are not only volatile in the sense of high annual standard deviation, but also show significant cyclical variability and tail

14. The difference between returns on short and long dated government securities in this historical sample is, on average, 1% p.a..

Figure 2.4.1. Risky returns compared to a safe benchmark



Note: Arithmetic average real total return on equity, housing and corporate bonds, compared to the real short-term government bill rate. Annual average of 17 countries, unweighted. Consistent coverage within countries: each country-year observation used to compute the average has data for all four asset returns.

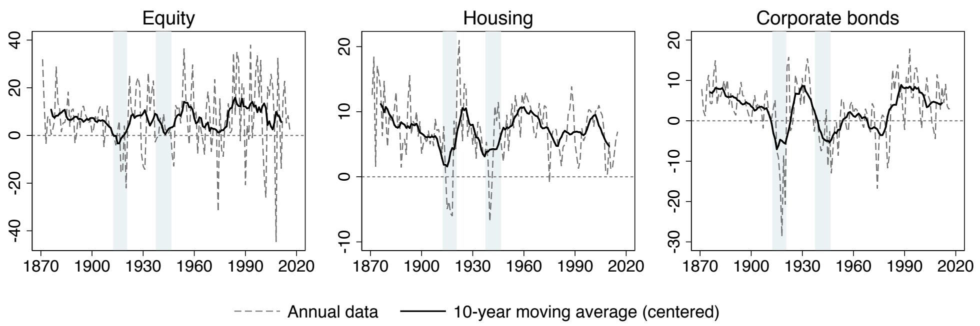
Table 2.4.1. Risky asset returns, cashflows and discount rates

	Panel 1: *Real total returns*		
	Equity	Housing	Corporate bonds
Mean	7.19	7.04	2.37
Standard deviation	21.97	9.92	10.73
Geometric mean	4.98	6.61	1.70
Excess return over bills	6.52	6.51	1.79
	Panel 2: *Cashflows*		
	Real dividend growth	Real rent growth	Corporate default rate
Mean	3.34	1.18	0.99
Standard deviation	29.08	8.02	1.51
	Panel 3: *Discount rates*		
	Dividend-price ratio	Rent-price ratio	Corporate bond spread
Mean	3.92	5.31	0.90
Standard deviation	1.66	2.01	0.93
Observations	1381	1381	1381

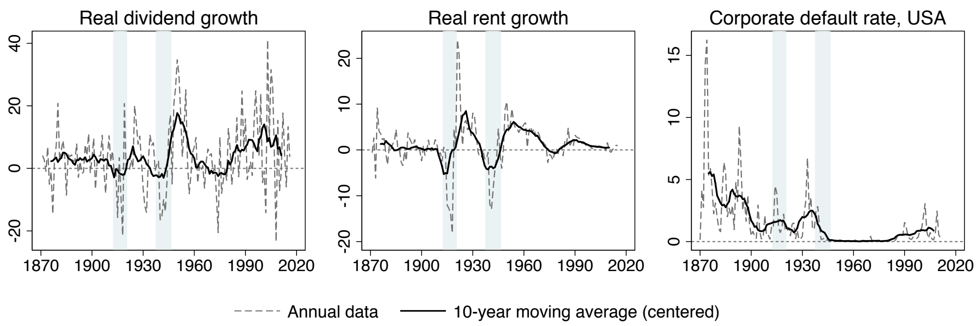
Note: Risky asset total returns, cashflows and discount rate proxies. Averages and standard deviations of pooled cross-country annual data, in percentage points. The corporate default rate measures the par value of bonds in default relative to total. Consistent sample: each country-year observation includes the data on all four asset classes.

Figure 2.4.2. Risky asset returns, cashflows and discount rates over time

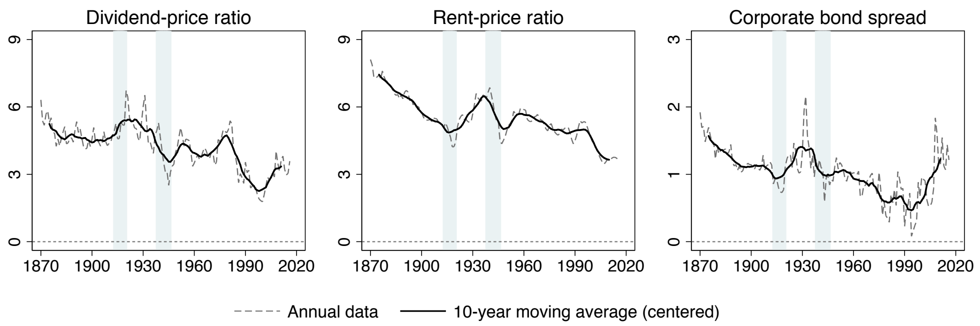
(a) Real total returns



(b) Cashflows



(c) Discount rates



Note: Unweighted averages of 17 countries. Shaded areas indicate world wars. Corporate default rate data are from Giesecke, Longstaff, Schaefer, and Strebulaev (2014).

risk. The figure displays the 17-country annual average return (grey dashed line), and its smoothed 10-year moving average time trend (solid black line) for each of the three asset categories.[15] The annual return series displays substantial volatility, as does the medium-term moving average trend. For example, the Global Financial Crisis saw a -40% return on equities in 2008, World War 1 saw real corporate bond returns of -30%, and both world wars saw below-zero total housing returns.

Fact 3: Risky cashflows are also volatile, but much less so than returns. Some of the volatility in returns can be accounted for by the uncertain nature of the risky cashflows. Panel 2 of Table 2.4.1 shows that dividend growth has an annual standard deviation of 26 percentage points. As shown in Figure 2.4.2b, equity cashflows fell during world war 1 and the recent crisis, when equity returns were also low, and grew strongly during the dot-com boom of the 1990s when equity returns were high. Time variation in cashflows seems much less important when it comes to housing and corporate bonds. Rental growth shows almost no variation during peacetime, and corporate default rate – based on the data for US only – has varied little since the late 19th century. Housing and corporate bond returns have, however, shown substantial variability during these time periods.

This fact suggest that at least some of the time variation in discount rate proxies – the inverted asset valuations dp shown in Figure 2.4.2c – corresponds to future returns rather than cashflows. For example, recent decades have seen a sharp increase in housing valuations while rental growth has slowed down. And bond spreads have varied substantially during the second half of the 20th century, while the default rate has remained flat.

Fact 4: Risky asset prices fluctuate more than future fundamentals. To see whether asset values, indeed, vary more than cashflows, I next compute the present value of future fundamentals for each year and asset class. For equity and housing, this "fundamental price" $P^*_{i,t}$ simply equals the discounted sum of future dividends or rents, discounted at a constant rate:

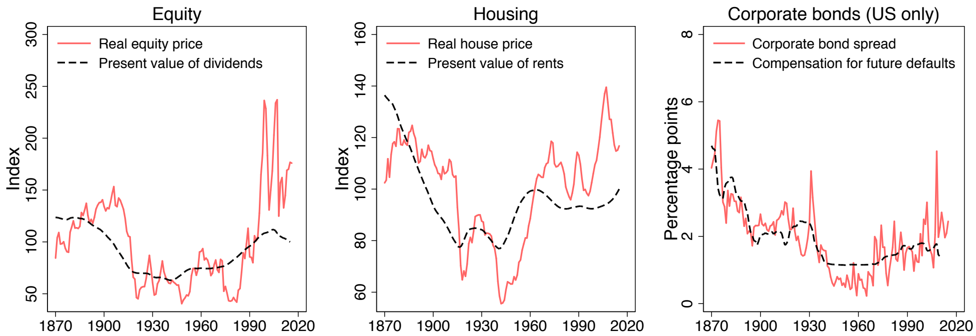
$$P^*_{i,t} \;=\; \mathbb{E}_t\left(\sum_{s=1}^{T-t}\frac{CF_{i,t+s}}{DR_i^s}\right) \tag{2.4.1}$$

Above, the discount factor m is simply the average of the ex-post rate observed in the data, $\overline{m} = 1/(1+\overline{dp_i})$, where dp_i is the dividend or rent to price ratio for equity and housing respectively. For corporate bonds, I compute the implicit default premium by regressing the bond spread on 15-year ahead defaults, and using the predicted spread value from that regression.[16]

15. The time trend is centered, such that the 1875 data point corresponds to the 1871–1880 average.

16. Specifically, I first regress spreads on 15-year ahead average default rate, $spread_t = \eta_1 + \eta_2\overline{Defrate}_{t,t+15} + e_t$. The rational spread is the predicted value from this regression $spread^*_t =$

Figure 2.4.3. Risky asset prices and fundamentals



Note: The equity and house price comparison follows Shiller (1981). Real equity and house prices are unweighted averages of the 17 countries in the sample, detrended using an exponential time trend. The present value of cashflows is the discounted sum of dividends or rents between year *t* and 2015, discounted at constant rate $1/(1 + dp)$, where dp is the long-run average rent-price or dividend-price ratio. Terminal value of discounted cashflows is set to equal the long-run average between 1870 and 2015. The compensation for future defaults is constructed by regressing spreads on a constant and the 15-year ahead default rate, and using the predicted value of this regression.

Figure 2.4.3 show the corresponding observed asset price or spread (solid red line) and the present value of fundamentals (dashed black line). I follow Shiller (1981) for the details of the calculation: real prices are detrended using an exponential trend, and both observed prices and discounted cashflows equal 100 on average throughout the sample. The equity and housing data are an unweighted average of the 17 countries in my sample, whereas the corporate bond data cover the US – the only country for which historical default data are available (the default rate is sourced from Giesecke et al., 2014).

Figure 2.4.3 shows that prices of all three risky assets move a lot more than future fundamentals. Reassuringly, prices and fundamentals do track each other over the medium to long run. Both dividends and rents were relatively high in the late 19th and late 20th centuries and low in-between, and so were the real equity and housing prices. This U-shaped pattern mirrors the U-shaped evolution of wealth-to-income ratios documented by Piketty and Zucman (2014), and suggests that even at long horizons, fluctuations in national are affected by swings in asset prices and valuations.[17] The corporate bond spreads, and their rational counterpart also follow a U shape – but since prices are inversely proportional to spreads, this actually means

$\hat{\eta}_2\overline{Defrate}_{t,t+15}$, and proxies the spread that would have been demanded by a rational investor with a constant discount rate, had the investor known defaults 15 years ahead – the approximate maturity of the bond.

17. Notably, the U-shaped real price pattern only becomes apparent when I detrend the real price series. Without adjusting for the exponential trend, the evolution of real equity and housing prices resembles a hockey stick, consistent with findings of Kuvshinov and Zimmermann (2018) and Knoll, Schularick, and Steger (2017).

Table 2.4.2. Correlations of discount rate proxies across asset classes

	Equity	Housing	Corporate bonds
Equity	1		
Housing	0.17	1	
Corporate bonds	0.12**	0.08	1

Note: Pairwise correlation coefficients between the equity dividend-price ratio, housing rent-price ratio and the corporate bond spread. Underlying data are 3-year moving averages, to smooth over timing idiosyncracies across assets, and are demeaned at country level. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$

that bond valuations were low (and default rates, high) in the 19th century, peaked in the mid-20th century and have been falling over the recent decades.

Despite this long-run correspondence, the price of each asset displays large cyclical deviations from it ex post rational value. Both in the early 1900s and 1990s, the rational equity price was close to its sample average, while the actual equity price was far above it. Detrended equity prices reached their all-time low in the 1980s, when discounted cashflows were a little subdued but not far from their long-run mean. The post-1960 surge in global house prices (Knoll, Schularick, and Steger, 2017) seems to have occurred without any discernible increase in real rental growth. And corporate bond spreads have fluctuated substantially since the end of World War 2, despite the fact that the default rate has remained roughly flat. The trends in Figure 2.4.3 fit a pattern of over-reaction to news about fundamentals. When discounted cashflows increase or fall, asset prices move in the same direction, but the magnitude of the price change is much larger than the movement in fundamentals.

Fact 5: Asset-specific discount rates do not co-move. One reason why asset prices vary more than future cashflows is that the rate at which these cashflows are discounted changes over time. But this discount rate variation should induce co-movement in valuations of different asset classes. Consistent with Figure 2.1.1 in the introduction, Table 2.4.2 shows that this co-movement is absent in the data. It shows the pairwise sample average correlations between asset-specific discount rate proxies: the dividend-price ratio, rent-price ratio and the corporate bond spread. The correlation coefficients are all close to zero and mostly not statistically significant.

This low co-movement between asset valuations is puzzling and poses a challenge to discount rate based explanations of excess volatility. Like the excess volatility finding of Figure 2.4.3, it echoes earlier work by Shiller (1982), who argued that prices of equity, housing and corporate bonds in the US follow highly distinct time series patterns, and hence the time variation in these prices is likely attributable to changing expectations rather than discount rates. But Shiller (1982) admitted that his data lacked accuracy and only included prices, not cashflows, which stopped him from drawing firm conclusions based on this low co-movement. The next two sections of this paper use my new historical dataset to test these propositions of excess volatility and low co-movement carefully withing the framework of return predictability regressions.

2.5 Return predictability within asset classes

Are swings in risky asset prices driven by time varying fundamentals, or non-fundamental factors such as discount rates or expectations? If financial volatility is purely fundamental driven, the lack of co-movement between asset valuations in Figure 2.1.1 simply corresponds to idiosyncratic cashflows, and is not, in itself, puzzling. As discussed in Section 4.3, I first test whether asset valuations *dp* help predict future returns and cashflows to gauge the relative importance of discount rate and cashflow variation.

Table 2.5.1 presents the results of predictability regressions in equations (2.2.4)–(2.2.5) for each risky asset class i. The first two columns show the outcomes of regressing, respectively, real equity returns r_{t+1} and real dividend growth dg_{t+1} on the dividend-price ratio dp_t. The numbers correspond to the predictive coefficients β_2 in equation (2.2.4) and δ_2 in equation (2.2.5) respectively. Columns 3 and 4 present the same results for housing. Column 5 tests for predictability of real corporate bond returns, while column 6 instead uses spread growth as the corporate bond return proxy, and effectively tests for mean reversion in spreads. Because spreads move in the opposite direction to returns, the coefficient in column 6 is equivalent to $-\beta_2$ in equation (2.2.4). In each regression, all the variables are demeaned at the country level.

The formal analysis in Table 2.5.1 confirms the pattern shown in Figure 2.4.3: returns on all three risky asset classes are predictable and hence, the prices of these assets are excessively volatile relative to fundamentals. Starting with equities, column 1 shows that high dividend-price ratios tend to be followed by high returns. This means that when the equity discount factor m is high, expected returns are also high, and hence either the price of risk is high, or market sentiment is depressed. A similar relationship is found for housing (column 3) and corporate bonds (column 5). The results are not only statistically, but also economically significant. A 1 percentage point increase in the dividend-price ratio forecasts 1.5 percentage points higher real equity returns 1 year ahead.[18] The magnitude of the effect is similar for housing, and somewhat smaller for corporate bonds – but returns on both of these assets are, in general, much less volatile than those on equities (Table 2.4.1). Finally, high corporate bond spreads not only predict high future returns, but a reduction, or mean reversion in the spread itself (column 6). The size of the mean reversion is

18. A 1 percentage point increase in the dividend-to-price ratio means a 25% relative increase (4% to 5%), hence $0.25 * 0.055) \approx +1.4\%$ change in the gross real return. The expected return increases by $1.07 * 0.014 * 100 = 1.5$ percentage points, where 1.07 is the mean real equity return in the sample.

Table 2.5.1. Predictability of real returns and cashflows within asset classes

	(1)	(2)	(3)	(4)	(5)	(6)
	Equity		Housing		Corporate bonds	
	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}	r_{t+1}	$\Delta spread_{t+1}$
Dividend-price ratio	0.055***	-0.120***				
	(0.014)	(0.030)				
Rent-price ratio			0.066***	-0.023**		
			(0.010)	(0.009)		
Bond spread					0.029***	-0.259***
					(0.007)	(0.042)
R^2	0.016	0.049	0.052	0.009	0.033	0.121
Observations	2203	2200	1818	1816	1594	1600

Note: OLS regressions with country fixed effects. Predictor (x) variables in rows. Dividend-price and rent-price ratios are in logs. Bond spread is the percentage point yield difference between corporate and government bonds. Dependent (y) variables in columns. r is the log real total return, dg is log real dividend or rental growth. $\Delta spread$ is the percentage point change in the bond spread. All variables are demeaned at the country level. Country-clustered standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

substantial: a 1 percentage point higher spread in year t predicts 0.28 percentage point lower spreads at $t + 1$.

Asset valuations forecast not only future returns, but also future cashflows (Table 2.5.1, columns 2 and 4). Again, the impact is both significant and economically sizeable. A 1 percentage point higher dividend-price ratio implies 3 percentage points lower dividend growth 1 year ahead.[19] For housing, the importance of cashflows is somewhat smaller: 1 percentage point higher rent-price ratios forecast only 0.5 ppts lower rental growth 1 year ahead. For corporate bonds, Appendix Table 2.A.3 shows that high bond spreads also predict high future defaults in US data, but the relationship is rather weak, for example only holding in levels and not rates of change, consistent with the findings of Giesecke, Longstaff, Schaefer, and Strebulaev (2011). Consistent with the broader time series pattern of Figure 2.4.3, financial booms and busts do partly reflect time-varying fundamentals, but that alone is insufficient to explain financial volatility, leaving an important role for time-varying discount rates or sentiment.

These findings hold under a wide range of alternative regression specifications, and across a variety of subsamples, presented in the Appendix Table 2.A.4. The results for each risky asset class are robust to all standard tests undertaken in the literature. They hold under alternative variable definitions – for example, for nominal returns, excess returns and 5-year ahead returns. Consistent with the literature on US equities, predictability becomes more powerful at longer horizon: a 1 percent-

19. The sample mean real dividend growth is 1.003, so the calculation is $0.25 * (-0.12) * 1.003 * 100 \approx 3$ percentage points.

Table 2.5.2. The relative importance of time varying discount rates and fundamentals

	(1) Equity	(2) Housing	(3) Corporate bonds
The share of total variation in asset valuations explained by:			
Discount rate news	32	58	79
Cashflow news	68	42	21

Note: Ratios of discount rate and cashflow news variance to total dividend-price, rent-price and bond spread variance, per cent. Discount rate news capture both time-varying discount rates and expectation errors. Cashflow news capture changes in fundamentals. Equity and housing shares are derived from the long-run covariance matrix of the VAR in returns, dividend growth and valuations, estimated using present value moment constraints. Bond discount rate news share is the ratio of the variance in spreads to discounted 10-year ahead spread growth; cashflow share is the residual.

age point higher discount rate *dp* predicts 3–10 percentage point lower cumulative returns 5 years ahead, depending on the asset class (3 ppts for corporate bonds, 5 for equities and 10 for housing). Results remain the same if estimated using a VAR with present value moment constraints rather than OLS, or OLS with time fixed effects. Predictability becomes somewhat stronger if valuation ratios *dp* are adjusted for structural breaks. I find little difference in predictability across the full sample or after 1950. Unlike the findings for US equities, this long-run predictability is also stronger for cashflows, not just returns. Unlike much of the existing literature which argues that predictability is strongest in recessions, I find similar predictive power across economic recessions and expansions. Appendix Table 2.A.5 also shows the predictability results for individual countries. Return predictability for housing and corporate bonds, and dividend growth predictability, is pretty much ubiquitous. For equity returns, predictive coefficients are generally of similar size to baseline but are not statistically significant in some countries due to the relatively lower precision of the within-country annual data.

Predictability regressions also allow me to calculate the relative importance of fundamental and expected return variation, by estimating the discount rate and cashflow news variance as discussed in Section 4.3, with estimation details provided in Appendix 2.A.3.3. In brief, I estimate a VAR in returns, cashflow growth and valuations in order to calculate the infinite discounted sum of future expected returns $\mathbb{E}\sum_{j=0}^{\infty}\rho^j r_{t+1+j}$, and cashflow growth $\mathbb{E}\sum_{j=0}^{\infty}\rho^j dg_{t+1+j}$, and use their relative variance to assess the contribution of discount rates and cashflows to variation in the valuation ratios dp_t. For corporate bonds, I use 10-year ahead predictive OLS regressions instead of a VAR, because of the lack of corporate default (cashflow) data.

Table 2.5.2 summarises the relative importance of discount rate and cashflow news variation for each asset class. A discount rate share equal to 100 means that all of the variance in the valuation ratio *dp* is accounted for by changes in future returns

– i.e. movements in discount rates or sentiment – and none by fundamentals. A share of 0 means all *dp* variation is fundamental driven. For equities, roughly one-third of the variation in the dividend-price ratios is accounted for by future returns, and two-thirds – by future cashflows. The discount rate news share is higher for housing, at 60%, and highest for corporate bonds. Roughly four-fifths of the corporate bond spread variation can be attributed to future spread growth. The high excess volatility of corporate bonds mirrors the large R^2 and high mean reversion of the predictability regression in Table 2.5.1 column 6.

Appendix Section 2.A.3.4 assesses how the excess volatility for each asset class has varied over time. It turns out that the importance of discount rate news has, if anything, increased throughout the 20th century, despite a number of fiscal and regulatory policy measures, such as the establishment of the welfare state, which should have reduced the time variation in risk aversion and financial intermediary risk appetite. The equity discount rate news share also showed a sharp spike during the Great Depression, pointing to the importance of non-fundamental asset price variation during this economic downturn, in line with views of a number of contemporaries (Keynes, 1936).

When it comes to analysing the drivers of excess volatility, going beyond the US equity market really does matter. The excess volatility in equity markets beyond recent US data is much lower, simply because the cashflow variation is much greater: extending the sample to the 19th century and to other advanced economies greatly increases the number of cashflow shocks and disaster events, as shown by the high variability of real dividend growth in Table 2.4.1 and Figure 2.4.2. But going beyond the relatively liquid, well-informed and centralised equity market to the markets for housing and corporate bonds, again, makes expected return variation central to the excess volatility analysis. Time-varying discount rates and sentiment explain more than half of movements in housing valuations, and almost all the variation in credit spreads.

The analysis of excess volatility paints a somewhat different picture to the "unconditional" volatility, or standard deviation of raw returns in Table 2.4.1. Unconditionally, equity returns are twice as volatile as those on housing and corporate bonds. But because dividends vary much more than rents and corporate defaults, the conditional or excess volatility in equity markets is, in fact, the lowest out of all three asset classes. Rents tend to be sticky and slow to adjust, and corporate defaults are rare and rather stable, at least in the 20th century US data. Cashflow volatility does little to explain the high variance of housing and corporate bond returns. Instead, this variation has to be driven by time-varying expected returns, i.e. discount rates or sentiment. Extending the analysis beyond US equities to include other countries and the two other risky asset classes is essential for correctly documenting the extent of this excess volatility, and understanding what it is that ultimately drives it.

As well as estimating the variance contribution of discount rates and cashflows, I can use the same empirical framework to construct an annual time series for the discount rate and cashflow news in unexpected returns following the methodology of Campbell (1991). These two time series allow me to disentangle the contribution of time-varying expected returns and cashflows to the discount rate proxy *dp*, with the discount rate news effectively providing a "clean" measure of the discount rate variation in the *dp* series. For equity and housing, discount rate news correspond to the change in the present value of future predicted returns, with the prediction constructed based on valuations, cashflow growth and returns in the current year. For corporate bonds, due to the lack of cashflow data and because most of the variation in spreads is attributable to discount rate news, I simply estimate discount rate news as the change in the spread. The estimation details are provided in Appendix 2.A.3.3.

The analysis in this section has two important implications for the co-movement tests in the Section 2.6. First, returns on each asset class are predictable, therefore expected returns do vary over time, and their co-movement can tell us something about the relative importance of volatile discount rates and expectations. Second, the predictability regressions allow me to construct a cleaned discount rate news series for each asset class, which is a proxy for the asset-specific discount factor m. Analysing the co-movement in this series thus allows for a more direct test of hypotheses 2 and 3 than that of the raw valuation ratios *dp*, since *dp* is driven by both discount rates and cashflows.

2.6 Comovement of discount rates across asset classes

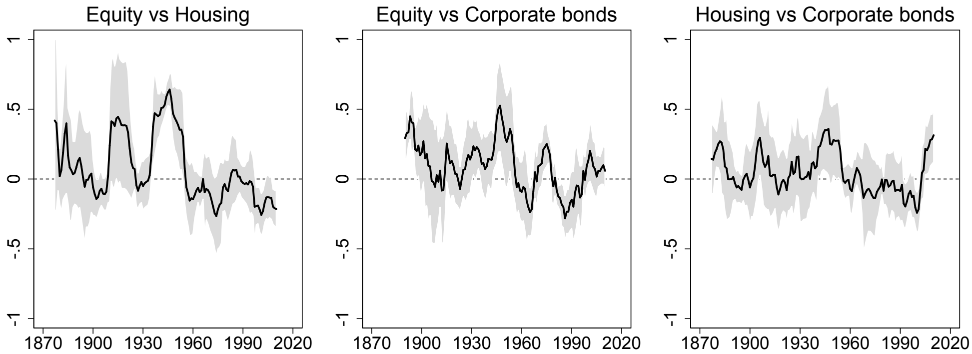
Are expected returns driven by a single factor – the discount rate – or asset-specific factors? As discussed in Section 4.3, the new cross-asset dataset allows me to disentangle these two forces by testing for co-movement and co-dependence in discount rates of different asset classes. If discount rates are an important determinant of asset price volatility, all the asset-specific discount rates should be driven by a single common factor.

2.6.1 Discount rate correlations

I start by assessing the raw correlations between discount rate and cashflow news of different asset classes. Figure 2.1.1 and Table 2.4.2 show that risky asset valuations show low co-movement. But Section 2.5 shows that these valuations are affected by both future cashflows and expected returns. It could be that idiosyncratic cashflow risk, rather than differences in expected returns are reponsible for the low co-movement between valuations of different asset classes.

Figure 2.6.1 shows the pairwise correlations of discount rate news on equity, housing and corporate bonds, with discount rate news on each asset i in year t

Figure 2.6.1. Co-movement of discount rate news across asset classes



Note: Pairwise correlation coefficients between the discount rate news on equity, bonds, and housing over rolling decadal windows (e.g. the value for 1875 if the correlation over the window 1870–1880). Shaded areas are 90% confidence intervals, using country-clustered standard errors. Underlying data are 3-year moving averages, to smooth over timing idiosyncracies across assets. Discount rate news correspond to changes in the present in future expected returns, for each asset class and year in the sample.

estimated as the change in the present value of expected returns in that given year (see Section 2.5 and Appendix 2.A.3.3). As in Figure 2.1.1, the correlations are computed over decadal rolling windows, with each underlying discount rate news series averaged over 3 years to make sure the results are not unduly affected by how quickly individual asset prices incorporate the changes in discount rates.

The correlations in Figure 2.6.1 paint an even starker picture than those for the raw valuation series in Figure 2.1.1. There is, give or take, no correlation between discount rates of different asset classes. The only robustly positive correlation appears during world war 2 – a unique historical episode when risk aversion and, hence, the price of risk is likely to have been high throughout the sample. If anything, the positive discount rate correlation during world war 2 suggests that the low co-movement during other periods is not a simple artefact of the data, and that periods when the price of risk varies substantially are rather rare.

Table 2.6.1 shows the pairwise correlation coefficients between discount rate news of different asset classes, computed over the full sample panel, with each underlying series demeaned at country level and averaged over 3 years. The results echo those of Figure 2.6.1: the discount rate correlations are close to zero and not statistically significant. Equity and housing cashflow news, on the contrary, do show a robustly positive correlation.[20] This stylised fact speaks against the fundamental based explanations for the low co-movement puzzle. These explanations would require cashflow news on different asset classes to show zero or negative co-movement which offsets the positive co-movement in discount rates. But the reality is the opposite: it is cashflow news that correlate strongly across asset classes, perhaps driven

20. Because of the lack of data on corporate bond default rates, I restrict the cashflow news analysis to the housing and equity asset classes.

Table 2.6.1. Correlations of discount rate and cashflow news across asset classes

	Discount rate news			Cashflow news	
	Equity	Housing	Corporate bonds	Equity	Housing
Equity	1			1	
Housing	0.16	1		0.32***	1
Corporate bonds	0.04	0.06	1		

Note: Pairwise correlation coefficients. Underlying data are 3-year moving averages, to smooth over timing idiosyncracies across assets. Discount rate and cashflow news for equities and housing are estimated as the innovations to present value of, respectively, future returns and cashflows for each asset class, within a VAR in returns, cashflow growth and valuations, and present value moment constraints. Discount rate news for bonds equal the present value of 10-year ahead spread growth. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$

Table 2.6.2. Discount rate and cashflow news correlations, alternative samples and definitions

	(1)	(2)	(3)	(4)	(5)
	Baseline	Housing 1 year ahead	10-year averages	Post 1950	Post 1980
Discount rate news:					
Equity, housing	0.16	0.13**	0.18*	−0.05	−0.11**
Equity, corporate bonds	0.04	0.04	−0.04	−0.04	0.00
Housing, corporate bonds	0.06	0.04	−0.03	−0.04	−0.13
Cashflow news:					
Equity, housing	0.32***	0.23***	0.40***	0.21***	0.26***

Note: Pairwise correlations. Unless otherwise stated, underlying data are three-year moving averages of discount rate news and cashflow news. All series are demeaned at country level. Housing 1 year ahead correlates the housing data at $t + 1$ and other asset data at t. 10-year averages smoothes each series using a 10-year rolling average. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$

by general macroeconomic risk, and discount rates which are an asset-specific phenomenon.

Table 2.6.2 shows that the low discount rate and high cashflow correlations hold across a variety of different variable definitions and subsamples. Column 1 shows the baseline correlations in Table 2.6.1, whereas each of the columns 2–5 changes the sample or variable definitions. The correlations are low and mostly insignificant when I forward housing returns by 1 year to account for delayed price transmission in this market (column 2), when looking over medium-term correlations using 10-year moving averages of the series (column 3), or when restricting the sample to the more recent decades (columns 4 and 5). After 1980, the correlation between equity and housing discount rate news is negative. Appendix Table 2.A.2 shows that the low discount rate correlation and high cashflow correlation are also the predominant feature of the data within individual countries, including the US. Taken together, these findings confirm that the low co-movement of asset-specific discount rates is, indeed, a robust feature of the data.

2.6.2 Cross-asset return predictability

The statistical power of cross-asset co-movement in expected returns can be tested more formally using the framework of return predictability regressions described in Section 4.3. For example, it could be the case that the dividend-price ratio helps predict housing or corporate bond returns, even if it is uncorrelated with the rent-price ratio or the bond spread. In this case, changes in the dividend-price ratio would correspond to the time-varying price of risk, with rent-price ratios and bond spreads measuring the asset-specific factors affecting housing and corporate bond return expectations.

Table 2.6.3 shows the results of cross-asset return predictability regressions. As in Table 2.5.1, each column corresponds to a different asset class, and each row – to a different predictor. Panel 1 tests for unconditional predictability across asset classes, and Panel 2 additionally includes the own-asset predictor (for example, the dividend-price ratio for equities), and tests whether the discount rates on asset classes carry

Table 2.6.3. Predictability of real returns across asset classes

	(1)	(2)	(3)
	Equity r_{t+1}	Housing r_{t+1}	Corporate bond $\Delta spread_{t+1}$
Panel 1: Unconditional predictability			
Dividend-price ratio		-0.002	-0.012
		(0.011)	(0.050)
Rent-price ratio	0.030		-0.079
	(0.023)		(0.057)
Bond spread	0.009	0.000	
	(0.010)	(0.003)	
R^2	0.003	0.000	0.001
Observations	1263	1227	1217
Panel 2: Conditional predictability			
Dividend-price ratio	0.089***	-0.009	0.024
	(0.024)	(0.011)	(0.049)
Rent-price ratio	0.007	0.073***	-0.018
	(0.022)	(0.014)	(0.094)
Bond spread	0.006	-0.002	-0.294***
	(0.009)	(0.002)	(0.059)
R^2	0.036	0.053	0.141
Observations	1226	1220	1217

Note: OLS regressions with country fixed effects. Dependent (y) variables in columns. Dividend-price and rent-price ratios are in logs. Bond spread is the percentage point yield difference between corporate and government bonds. Predictor (x) variables are the log real total return on each asset class. Country clustered standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

predictive power for future returns that goes above and beyond that of the own-asset predictor.

Consistent with the discount rate correlation evidence in Figure 2.6.1 and Table 2.6.1, the estimates in Table 2.6.3 show that there is, essentially no cross-asset return predictability in the data. Asset-specific valuations do not help predict returns on other asset classes, either conditionally or unconditionally. Knowing that the rent-price ratios are low, and housing valuations are high, tells us nothing about the expected returns on equity or corporate bonds.

Table 2.6.4 shows that the lack of cross-asset return predictability continues to hold under a variety of alternative return definitions, and across different time periods. Each row shows the predictive coefficient on the asset-specific return, when regressed on valuations of the other two risky assets. Column 1 corresponds to the unconditional predictability results in Table 2.6.3 Panel 1. Columns 2-6 look, respectively, at longer-horizon predictability, nominal returns, and restricting the sample to the post-1950 period, recessions or expansions (identified by applying the Bry and Boschan (1971) algorithm to real GDP per capita). There is no consistent evidence for cross-asset return predictability under any of these alternative specifications. When it comes to nominal returns, high dividend-price ratios predict low, rather than high future returns on housing. Similarly, high dividend-price ratios fore-

Table 2.6.4. Predictability of real returns across asset classes: alternative specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	5-year returns	Nominal returns	Post 1950	Recessions	Expansions
Equity r_{t+1}:						
Rent-price ratio	0.030	-0.016	0.025	0.044	-0.002	0.039
	(0.023)	(0.022)	(0.026)	(0.036)	(0.038)	(0.025)
Bond spread	0.009	0.001	0.002	0.011	0.049**	-0.002
	(0.010)	(0.006)	(0.008)	(0.013)	(0.021)	(0.010)
Housing r_{t+1}:						
Dividend-price ratio	-0.002	0.001	-0.019	-0.005	0.015	-0.006
	(0.011)	(0.008)	(0.016)	(0.018)	(0.020)	(0.011)
Bond spread	0.000	0.002	-0.006	0.003	0.002	0.002
	(0.003)	(0.004)	(0.006)	(0.004)	(0.013)	(0.004)
Corporate bond $\Delta spread_{t+1}$:						
Dividend-price ratio	-0.012	-0.031		0.036	0.190*	-0.035
	(0.050)	(0.037)		(0.076)	(0.111)	(0.048)
Rent-price ratio	-0.079	-0.022		-0.122**	-0.198	-0.087
	(0.057)	(0.027)		(0.052)	(0.289)	(0.059)

Note: Predictive coefficents of returns on one asset class at $t + 1$ regressed on the discount rate proxies of other asset classes at t. Dividend-price and rent-price ratios are in logs, bond spread is in percentage points. Predictors (x variables) in rows. Specifications in columns. r_{t+1} is log real total return; $\Delta spread$ is the percentage point change in the corporate bond spread. 5-year returns uses 5-year ahead average returns or spread growth as the dependent (y) variable. Expansions and recessions are dated using the Bry-Boschan algorithm. OLS regressions with country fixed effects. Country clustered standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

cast high spreads – and hence low, rather than high corporate bond returns – during recessions. During expansions, dividend-price ratios do help forecast bond spread growth in the right direction, but they do not forecast housing returns.

2.6.3 Predictive power of macro-financial risk factors

I now turn to examine the predictive power of various macro-financial risk factors. The factor selection mirrors existing theoretical and empirical asset pricing literature, subject to historical data availability constraints. Broadly speaking, I include those macroeconomic and financial variables which serve as proxies for the price of risk in theoretical asset pricing models, or have been shown to predict equity returns in the empirical literature. The list of factors can be broadly divided into the following three categories:

Consumpton-based factors. The notion of the price of risk is strongly linked to household consumption. When consumption is low, households should be unwilling to save or take risks, and hence demand high compensation for holding risky assets. Campbell and Cochrane (1999) show that rather than looking at raw consumption growth, the deviation of consumption from a slow-moving trend – or "habit" – does a better job in explaining the time variation in equity valuations. Lettau and Ludvigson (2002) show that rather than consumption itself, the "consumption-wealth ratio" *cay* – measured as the deviation of consumption relative to wealth from its long-run trend – has strong predictive power for stock returns in the US.

To proxy the consumption and habit factors, I use, respectively, the 3-year consumption growth, and the deviation from consumption from its 10-year moving average trend. For the consumption-wealth ratio, I proxy total financial wealth as the sum of stock market capitalization and housing wealth, and use real wages as a proxy for real income.[21] I then follow Lettau and Ludvigson (2002) and estimate the cointegrating relationship between consumption, financial wealth and labour income, and compute the consumption-wealth ratio as the deviation of consumption from this long-run cointegrating relationship. The consumption and wage data are sourced from the latest vintage of the Jordà-Schularick-Taylor macrohistory database (Jordà, Schularick, and Taylor, 2016), stock market capitalization data come from Kuvshinov and Zimmermann (2018), and housing wealth data are from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2017).

Financial intermediary factors. A series of recent papers suggest that high risk appetite by financial intermediaries affects prices of risky assets, both in theory and in the data (He and Krishnamurthy, 2013; Baron and Muir, 2018). Most measures of

21. Because of the lack of data on other classes of wealth, or total income for my historical sample, I restrict the analysis to stock market and housing wealth, and use wages rather than total labour income to proxy the flow of human wealth.

risk appetite either correspond to balance sheet strength, or balance sheet growth of financial intermediaries. Growing leverage or bank assets are, therefore, signs of high intermediary risk appetite. To this end, I add the 3-year growth in bank leverage, real bank assets and real credit to the predictor set. The leverage data are from Jordà, Richter, Schularick, and Taylor (2017), while bank asset and credit data come from (Jordà, Schularick, and Taylor, 2016).

Other factors. I add the stock market capitalization relative to GDP and term spread – two variables which have been shown to have considerable forecasting power for equity returns. Kuvshinov and Zimmermann (2018) show that the stock market cap to GDP ratio outperforms the dividend-price ratio as an equity return predictor, because it incorporates changes in quantities, or issuance, as well as prices. Campbell (1991), among others, finds that the term spread reliably forecasts US stock returns.

Table 2.6.5 reports the outcomes of forecasting regressions in equation (2.2.7). As before, each column corresponds to a different asset class, while each row corresponds to a different predictor. For each predictor, the table includes the expected sign of the coefficient on equity and housing returns. Because I use spread growth to proxy bond returns, the expected coefficient sign in Table 2.6.5 column 3 is the opposite of that in columns 1 and 2, e.g. positive for real consumption growth. Because most of my indicators are high when the price of risk is low, the expected coefficient sign is mostly negative. Expected returns should be low when consumption growth or surplus consumption is high (and hence, desire to save in risky assets – high, and price of risk – low), or when intermediary risk appetite is high. Expected returns should be high when consumption relative to wealth is high (and hence, the high quantity of wealth is suggestive of a low desire to save more in risky assets). Turning to other factors, high market capitalization is a sign of high equity valuations, a low price of risk and low expected returns, while high term spreads indicate that the price of term, or duration, risk is high, and hence expected returns should be high.

The results in Table 2.6.5 confirm those in the preceding sections. The cross-asset predictive power of macro-financial risk factors is limited. Some of these factors are important for individual asset classes. For example, high consumption growth or high stock market capitalization predict low subsequent equity returns. Growing bank leverage – and hence, high intermediary risk appetite forecasts low housing returns. But this predictive power is highly asset specific. No single factor predicts returns on all three asset classes, and some factors predict returns in the wrong direction. For example, high credit growth predicts high, rather than low housing returns, which suggests that these higher housing returns and prices are simply a result of increased credit availability, rather than intermediary-based asset pricing factors.

The only factor that predicts returns on more than one asset class in the right direction is the term spread, which measures time-varying term premia, rather than

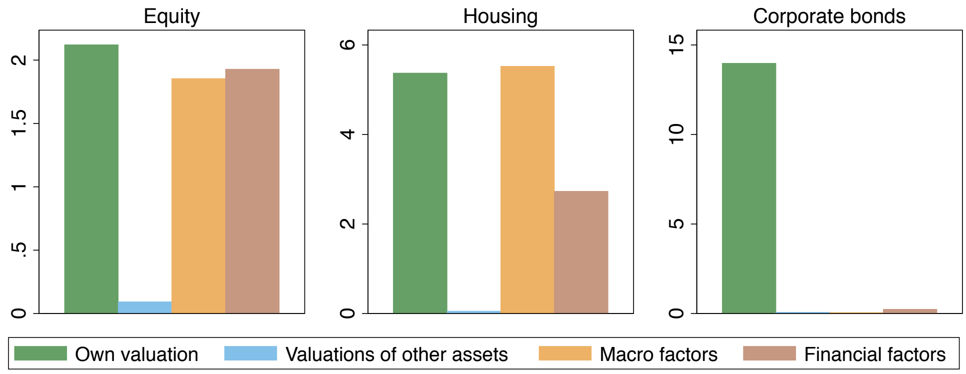
Table 2.6.5. Return predictability using macro-financial risk factors

	(1)	(2)	(3)
	Equity r_{t+1}	Housing r_{t+1}	Corporate bond $\Delta spread_{t+1}$
Consumption-based factors:			
$\Delta_3 Real\ Consumption_t(-)$	-0.271**	0.172*	0.432
	(0.136)	(0.088)	(0.390)
$Surplus\ Consumption_t(-)$	0.021	-0.040	-0.774
	(0.154)	(0.094)	(0.581)
$cay_t(+)$	0.018	0.037	0.140
	(0.041)	(0.027)	(0.103)
Financial intermediary factors:			
$\Delta_3 Bank\ Leverage_t(-)$	-0.060	-0.033**	-0.013
	(0.039)	(0.016)	(0.115)
$\Delta_3 Real\ Bank\ Assets_t(-)$	-0.029	0.019	0.242
	(0.055)	(0.024)	(0.259)
$\Delta_3 Real\ Credit_t(-)$	0.007	0.046***	0.257
	(0.043)	(0.016)	(0.170)
Other factors:			
$log(MCAP_t/GDP_t)(-)$	-0.045***	0.000	-0.021
	(0.008)	(0.004)	(0.034)
$Term\ Spread_t(+)$	0.012**	0.006***	0.041
	(0.006)	(0.002)	(0.034)
R^2	0.039	0.053	0.013
Observations	1471	1391	1103

Note: OLS regressions with country fixed effects. Dependent (y) variables in columns. $\Delta_3 Real\ Consumption_t$ is the log change in real consumption per capita from $t-3$ to t. $Surplus\ Consumption_t$ is the real consumption per capita at t relative to a backward-looking 10-year moving average trend, from t-10 to t. cay_t is a proxy for the consumption-wealth ratio, estimated as the deviations from the cointegrating relationship between real consumption, a proxy for financial wealth (the combined capitalization of the equity, housing and government bond markets) and real wages. $\Delta_3 Bank\ Leverage_t$ is the change in the log of bank leverage from $t-3$ to t. $\Delta_3 Real\ Bank\ Assets_t$ is the change in the log of real bank assets from $t-3$ to t. $\Delta_3 Real\ Credit_t$ is the change in the log of real credit to non-financials from $t-3$ to t. $log(MCAP_t/GDP_t)$ is the log of the market capitalization to GDP ratio. $Term\ Spread_t$ is the percentage point yield spread between the long and short term government debt. Predictor (x) variables are the log real total return on equity and housing, and the percentage point change in the corporate bond spread. Country clustered standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

the time-varying price of risk. Indeed, similarly to the high cashflow news correlation (Table 2.6.1) or high equity and housing discount rate co-movement during wars (Figure 2.6.1), the predictive power of the term premium works against discount rate based explanations for excess asset price volatility. Because both equities and housing are long duration assets, we would expect their returns to be predictable by a factor that reflects the price of duration risk, and indeed this is what we find. Because corporate bond spreads are taken over those of government bonds of similar maturity, the term premium has no explanatory power for corporate bond spreads. This shows that when there is a common risk factor which affects multiple asset

Figure 2.6.2. Explanatory power of different sets of predictors



Note: Comparison of R2 statistics from return predictability regressions. The dependent (y) variables are the log real total return on equity or housing, and the spread growth for corporate bonds, at $t + 1$. Predictor (x) variables change depending on the specification. The "own valuation" specification uses the asset-specific valuation ration dp_t only, "valuations of other assets" uses only the valuations of the other two asset classes. The macro factors are consumption growth, surplus consumption, and the consumption-wealth ratio. The financial factors are growth in bank leverage, real bank assets and real credit. All macro-financial growth rates are from year $t - 3$ to t. OLS regressions using country fixed effects. Consistent sample used across all specifications within an asset class: i.e. all the bars in the equity specification use the same sample, but a different sample may be used for housing.

classes, this factor shows up with significant predictive coefficients for the respective returns, with the right sign. The fact that other factors do not show up in such a manner speaks against single-factor explanations of excess asset price volatility.

The low cross-asset power of macro-financial factors holds under a variety of return definitions and sample specifications, with the results reported in the Appendix Table 2.A.9. Conditioning on own valuations, looking at 5-year ahead returns or limiting the sample to after 1950 does not materially change the results.

Figure 2.6.2 compares the cross-asset predictive power of own asset valuations, valuations of other assets, and those of macro and financial risk factors. More precisely, it contrasts the R^2 from return predictability regressions using only the own asset valuation (as in Table 2.5.1), to those using only the valuations of other asset classes (as in Table 2.6.3 panel 1), and those using only the consumption-related, or only the financial factors in Table 2.6.5. Own valuations are the most informative predictor of expected returns. Valuations of other assets have almost no predictive power. Macro-financial factors do have some predictive power for equity and housing returns, but as discussed above, this power corresponds to asset-specific rather than cross-asset factors. Consistent with the comparison of excess volatility across asset classes in Section 2.5 Table 2.5.2, the relative forecasting power as measured by the R^2 is highest for corporate bonds, moderate for housing and relatively low for equities.

Taken together, the analysis in this section makes clear that discount rates – or expected returns – on different assets do not co-move. Therefore, discount rate variation is not an important driver of asset price volatility. But as the analysis in Section 2.5 has shown, asset prices do move excessively relative to fundamentals. If not discount rates, what is it, then, that drives this excess financial volatility? The next section explores a range of potential theoretical explanation behind the asset-specific variation in expected returns.

2.7 The time varying risk puzzle

Sections 2.5 and 2.6 have documented the following stylised feature of the new cross-asset risk and return dataset: asset-specific discount rates vary over time, but do not correlate across asset classes. This stylised fact poses a challenge to most of the prominent models in macro-finance. Cochrane (2017) argues that most modern macro-finance theories rely on the time variation in the discount rate, or the price of risk, to generate excess volatility, and only differ in the specific ways in which this time variation in the price of risk is generated. But even though such models can match asset price volatility within an asset class, they are not able to generate a lack of co-movement in expected returns across asset classes, because the discount factor variation is common to all risky assets. The lack of co-movement in asset-specific discount rates, therefore, constitutes a new "time varying risk" puzzle for the asset pricing literature.

In this section, I evaluate in more detail whether standard asset pricing models are able to match this low co-movement fact, and discuss several modifications or classes of models – including investor heterogeneity and volatile expectations – that may be able to account for the low co-movement in discount rates observed in the data. Asset pricing theories which rely on time variation in the price of risk can be divided into five broad categories. Each of these classes of models relies on time variation in the discount rate m to explain excess volatility, but gives a different reason as to why m should vary over time.

Representative agent consumption CAPM. The simplest macro asset pricing model simply attributes time variation in the price of risk to changes in the marginal utility of consumption:

$$m_t = u'(c_{t+1})/u'(c_t)$$

It is generally accepted that this formulation of the pricing kernel does not generate sufficient volatility in m, hence a number of modifications have been proposed to make the discount factor and hence asset prices more volatile.

Time-varying risk aversion. These theories stipulate that the price of risk is determined by some additional parameter H, which affects investors' marginal utility and thus makes their "effective" risk aversion volatile, such that the marginal utility of consumption varies by more than consumption itself:

$$m_t = f(H_{t+1}/H_t)$$

The most most popular candidate for H is time variation in the consumption habit or surplus consumption proposed by Campbell and Cochrane (1999). Piazzesi, Schneider, and Tuzel (2007) argue that time variation in the housing consumption share can also drive such movements in risk preferences.

Long-run risk. models, such as Bansal and Yaron (2004), rely on time variation in consumption volatility to generate movements in the discount factor:

$$m_t = f(\sigma_{C,t})$$

Rare disasters. In this class of models, time variation in m reflects the time-varying disaster risk, determined by the probability of disasters p, disaster consumption losses B and disaster-specific asset returns R_{t+1}^{dis} (Barro, 2006; Barro and Ursua, 2008; Gabaix, 2012):

$$m_t = p_t f(B_{t+1} R_{i,t+1}^{dis})$$

A high disaster risk would elicit high expected returns on on all risky asset classes, particularly for those assets which experience low disaster returns.

Intermediary asset pricing. This set of theories argues that instead of being priced by a representative households, assets are priced by financial intermediaries or market makers, with the pricing kernel m corresponding to financial intermediary risk appetite, usually tied to their balance sheet characteristics bs (He and Krishnamurthy, 2013):

$$m_t = f(bs_t)$$

The measures bs_t include intermediary leverage, balance sheet strength and asset growth, which act as proxies for the ability and willingness to bear financial risk.

Every single one of these models relies on time variation in the price of risk to generate the volatility in m necessary to explain the variation in expected returns documented in Section 2.5. This discount rate variation, however, induces a strong co-movement in expected returns on different asset classes, which is something that we do not observe in the data, as documented in Section 2.6. This lack of co-movement is directly stipulated by the time-varying risk aversion, long-run risk and financial intermediary theories. It is somewhat more subtle in the case of disaster risk models: if returns and risk premiums during disasters differ markedly across the risky asset classes, time-varying disaster risk loadings may generate some heterogeneity in cross-asset expected returns. But in the data, all three assets show low returns during consumption disaster period, and expected returns across the three asset classes show high co-movement during consumption disasters as illustrated, for example, by the high equity and housing discount rate correlations during the two world wars in Figure 2.6.1. Therefore, the heterogeneity in disaster returns is unlikely to be important in explaining the low co-movement puzzle.

There are three different ways to reconcile asset pricing models with the low co-movement and resolve the time varying risk puzzle. The first approach introduces additional pricing factors or frictions which delink asset-specific returns from the discount factor. The second approach seeks to generate asset-specific m_is which are rationalised via some form of market and investor heterogeneity. The third approach uses an altogether different source of variation in expected returns, namely volatile expectations, to generate excess financial volatility without a need for high cross-asset expected return co-movement. I consider each of these solutions in turn.

2.7.1 Asset-specific risk and non-monetary payoffs

Expected return on asset class i does not depend on the discount factor m alone. It also depends on the riskiness of the asset class and any non-cash utility payoffs accruing to the asset holder, such as liquidity services. Asset-specific variation in these risks or payoffs could therefore delink expected returns from discount rate variation, and generate the low co-movement observed in the data.

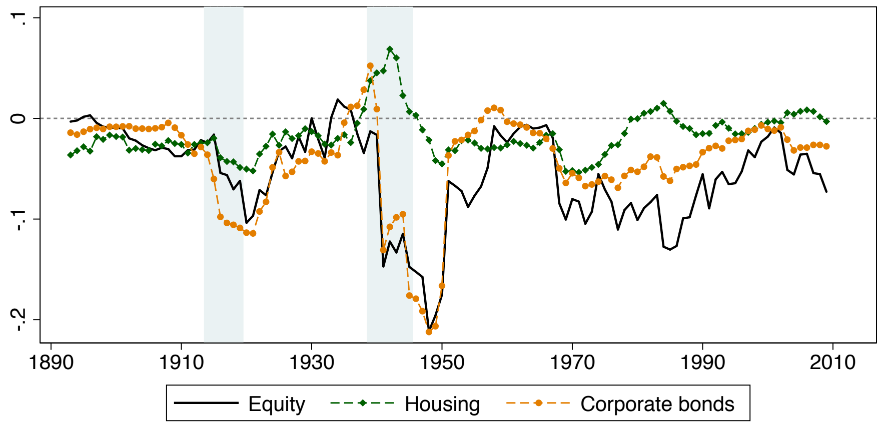
Asset-specific risk. Expected returns should be higher if an asset is riskier. In the CAPM, the excess return premium is captured by the asset "beta": the higher the co-movement of an asset with the market return, the higher the required compensation for that extra risk. In macro asset pricing models, asset "beta" equals the covariance of the return with the discount factor: the lower the asset's payoffs in bad states, the higher the expected return. Decomposing the joint expectation $\mathbb{E}(mR)$ into variation in expected return $\mathbb{E}(R)$ and risk $cov_t(R_{i,t+1}, m_{t+1})$ yields:

$$\mathbb{E}_t(R_{i,t+1}) = \frac{1 - cov_t(R_{i,t+1}, m_{t+1})}{\mathbb{E}_t(m_{t+1})} \tag{2.7.1}$$

Most asset pricing models implicitly assume that the beta, or asset risk, stays constant over time. But this need not be so. For example, Lettau, Maggiori, and Weber (2014) augment the standard CAPM with an additional downside-risk beta, with the asset riskiness at time t effectively an average of the exposure to downside risk and the riskiness during normal times. An increase in downside risk would then increase the riskiness of those asset classes which are sensitive to this factor, such as fixed-income securities.

The question, then, becomes: how much does asset-specific risk or beta fluctuate in practice? This is inherently difficult to assess because the discount factor m, which affects asset riskiness through covariance $cov_t(R_{i,t+1}, m_{t+1})$ is itself not directly measurable, and I argue that the asset-specific discount rate proxies are a poor measure of the aggregate pricing kernel. But we can proxy the aggregate m as the simple average of the asset-specific ms: the dividend-price ratio, rent-price ratio, and

Figure 2.7.1. Time variation in asset-specific risk, or “beta”



Note: Covariance between equity, housing and corporate bond returns, and a simple proxy for the discount factor *m*. Centered 10-year rolling windows. The *m* proxy equals the average of three asset-specific discount factors (the log inverse dividend-price ratio, rent-price ratio and corporate bond spread, each normalised to mean 0 and standard deviation of 1 over the full sample).

the corporate bond yield (after normalising each series to mean zero and standard deviation of 1). Figure 2.7.1 then plots a proxy for the time-varying asset-specific beta: the covariance between the realised return and the proxy for m computed over rolling 10-year windows.

Figure 2.7.1 suggests that asset-specific riskiness or beta does vary over time: for example, asset risk increased during the two world wars, and fell during the Great Moderation. But the time variation in betas is both small – with a covariance of around -0.1 during most of the peacetime – and similar across the different risky asset classes. The only instance where the risk of each asset class diverges is World War 2, when housing returns exhibited substantally lower risk than those on equity and corporate bonds. Notably, world war 2 happens to be the only time period during which the discount rates on different asset classes actually do co-move (Figure 2.6.1).[22] Appendix Figure 2.A.5 uses a more sophisticated proxy for m which utilises data on all cross-asset expected returns and macro-financial risk factors, but finds that the time variation in riskiness becomes smaller and more homogenous across asset classes.

This suggests that time-variation in betas is small and similar across the three risky asset classes, and hence not capable of accounting for the zero expected return correlation in the data. The low importance of asset-specific risk is a direct consequence of the dataset used in this paper: all the assets included in this study are risky and exclude safe securities such as bills or government bonds and further-

22. In addition to this, the riskiness of housing investments during world war 2 is likely to be understated in the data, for two reasons. First, may countries introduced rent or price controls which means that returns do not necessarily reflect true market prices. Second, my measure of housing return does not account for destruction of buildings during the war (see Jordà, Knoll, et al., 2017).

more, each of these assets is actually a major asset class and a sizeable component of the return on risky wealth. This assures that the asset classes remain risky throughout the sample, and their returns do not deviate markedly from those on the market portfolio.

Non-monetary payoffs. The monetary cashflows in the data may not capture the full value of utility services provided by an asset class. It the expected value of these services varies over time and across asset classes, it could make specific asset classes relatively more or less attractive, and reduce the co-movement in expected returns. For example, housing is relatively illiquid, and if liquidity premia rise, expected returns on housing should fall relative to the other two asset classes.

The two primary candidates for these non-monetary services are housing utility and liquidity services. Housing utility variation should, however, be captured in my rental cashflow measure, which includes imputed rents of owner occupiers. Still, the imputed rent measure is noisy and may not fully capture the time variation in preferences. As further evidence, Appendix Figure 2.A.6 shows that the share of rents in GDP – a proxy for housing consumption share, which in turn drives the marginal utility of housing services (Piazzesi, Schneider, and Tuzel, 2007) – displays little cyclical variation, especially during peacetime. Finally, the variation in housing preference would be unable to explain the low co-movement between equities and corporate bonds.

Turning to liquidity, because housing transactions are costly and take time to execute, a high liquidity preference should reduce housing valuations relative to those of the other two asset classes. But corporate bond markets are also illiquid (Bao, Pan, and Wang, 2011). This means that time variation in liquidity premiums should induce some co-movement between housing and corporate bond discount rates. In the data however, the co-movement between housing and corporate bond discount rates is even lower than that between equities and housing. Further to this, liquidity tends to be pro-cyclical and co-move strongly with risk appetite measures such as intermediary leverage (Brunnermeier and Pedersen, 2009; Adrian and Shin, 2010).

Taken together, the time varying risk puzzle cannot be fully attributed to time-varying betas or asset-specific frictions which generate heterogeneity in asset-specific returns even in the presence of an aggregate discount factor m. But asset-specific discount factors could themselves be heterogeneous because of differences among investors and market segmentation – a set of explanations for the puzzle that I consider next.

2.7.2 Heterogenous investors and segmented markets

Investor heterogeneity or market segmentation alone cannot account for the time varying risk puzzle. If investors are heterogenous but markets are not segmented, there will be a common discount factor which drives expected returns on all risky asset classes, with the time variation in that discount factor driven by some average of individual investor discount rates (Constantinides and Duffie, 1996; Gârleanu and Panageas, 2015).[23] If markets are segmented but investors in each market are the same, the discount factors across different asset classes will also be equal. Put differently, heterogeneity and segmentation need to be large enough to, effectively, allow for arbitrage across expected returns on different asset classes. Investor heterogeneity allows these return differentials to arise in the first place, and market segmentation ensures that they are not arbitraged away. Existing evidence, however, suggests that neither of these two heterogeneities is sufficiently large to generate the observed absence of discount rate co-movement.

Starting with market segmentation and limits to arbitrage, several recent studies have documented an existence of arbitrage opportunities across different assets. But outside of crises, the size of return differentials generated by these arbitrage strategies is an order of magnitude smaller than those implied by the cross-asset discount rate differentials in my data. For example, Fleckenstein, Longstaff, and Lustig (2010) document a return differential of around 20 bps between inflation-linked bonds and a combination of a nominal bond and an inflation swap outside of crisis years, and Hu, Pan, and Wang (2013) document differentials of around 6 bps between different government securities. In my data, a one standard deviation higher rent-price ratio is not informative of future equity or corporate bond returns, but implies a 150 bps higher return on housing 1 year ahead (Table 2.5.1), and 800 bps higher returns over 5 years (Table 2.A.4).

These return differentials remain large if we narrow the focus to equity and corporate bond securities, which have low transaction costs and are typically traded on the same stock exchange by the same investors. A 1 standard deviation higher dividend-price ratio is not informative about future bond returns, but signals 150 bps higher equity returns 1 year ahead, and 500 bps higher cumulative return 5 years ahead. All these results hold in normal times or during economic expansions, as well as during recessions or financial crisis.

Arbitrage opportunities aside, the degree of market segmentation and investor heterogeneity across the three risky asset classes also does not appear to be quantitatively large. A large share of the equity and corporate bond market is held by large institutional investors, especially over recent decades (Gompers and Metrick, 2001; Biais and Green, 2007). Even though housing is primarily held by the middle class, and equity – by high-income individuals, there is a large degree of cross-asset holdings even among these groups: high income earners still own their house, and middle income earners have an exposure to the equity market through their pensions (Garbinti, Goupille-Lebret, and Piketty, 2017; Kuhn, Schularick, and Steins,

23. In Constantinides and Duffie (1996), this discount factor can be linked to the cross-sectional variance of consumption growth, and in Gârleanu and Panageas (2015) – to a weighted average of individual investors' risk aversion.

Table 2.7.1. Correlations of real post-tax income growth across the income distribution

	Average growth	Income growth correlations		
		Bottom 50%	Middle 40%	Top 10%
Bottom 50%	1.23	1		
Middle 40%	1.34	0.76	1	
Top 10%	1.95	0.43	0.76	1

Note: Mean real post-tax income growth, and pairwise correlation coefficients of mean real post-tax income growth among the bottom 50%, percentiles 50–90, and the top 10% of the income distribution. Data are from Piketty, Saez, and Zucman (2018), and cover US 1962–2014 only. All correlations are significant at the 1% level.

2017; Martıúnez-Toledano, 2018). The income shocks – and hence discount factors – of these two groups are also highly positively correlated.

Table 2.7.1 reports the average income growth, and income growth correlation of the bottom of the income distribution – who can be thought of as holding no risky assets – the middle – the representative housing owner – and the top, the representative equity owner, using the post-1962 US data from Piketty, Saez, and Zucman (2018). While these income groups have experienced substantially different trend income growth (Table 2.7.1, column 1), the year-to-year growth variation which drives the stochastic discount factor is strongly positively correlated. The correlation in income growth between a representative housing and equity owner – the middle 40% and the top 10% – is around 0.76.

These facts suggest that while investor heterogeneity and market segmentation may reduce the co-movement in expected returns, they are unlikely to explain the zero correlation observed in the data – either alone or in combination with asset-specific risk or non-monetary payoffs (Section 2.7.1). To reconcile the low co-movement fact with theory, it may be more fruitful to focus on a different source of time variation in expected returns altogether: that of volatile expectations.

2.7.3 Volatile expectations

Volatility in expectations means that investors form their forecasts of future cashflows using an operator $\mathbb{E}_i^*$ that differs from the ex-post rational forecast $\mathbb{E}$. These differences can vary over time and across asset classes, generating the patterns of asset-specific excess asset price volatility documented in Sections 2.5 and 2.6.

Behavioural theories provide a number of possible forms for such variation in expectations. Most of these rely on some form of extrapolation – of past returns, cashflows, or other features of observable data. The simplest form of extrapolation simply makes return forecasts on the basis of past realised returns on the specific asset class i:

$$\mathbb{E}_{i,t}^*(R_{i,t+1}) \;=\; f(R_{i,t-1}) \tag{2.7.2}$$

For example, if returns are high during one period, investors expect them to also be high in the future, even if these expectations are not justified from a rational ex post perspective. Barberis, Greenwood, Jin, and Shleifer (2015) argue that a model where some investors form beliefs from extrapolating past price changes, and other investors hold rational beliefs provides a good match for asset price data and survey evidence. Adam and Merkel (2018) show that a general equilibrium model with extrapolative expectations can reconcile the variation in equity prices with that in macroeconomic aggregates.

A somewhat more sophisticated form of extrapolation builds on the representativeness heuristic through "diagnostic expectations" (Bordalo, Gennaioli, La Porta, and Shleifer, 2017). It assumes that investor expectations overweight surprising return outcomes, and expect assets which have shown higher than expected returns in the past to also do so in the future:

$$\mathbb{E}^*_{i,t} = f(R_{i,t-1} - \mathbb{E}_{t-1}R_{i,t-1}) \tag{2.7.3}$$

Another class of models helps explain why expected return co-movement across asset classes may be negative, rather than positive. The "style investing" model of Barberis and Shleifer (2003) assumes that if one asset class performs relatively well to others, investors expect continuied overperformance:

$$\mathbb{E}^*_{i,t} = f(R_{i,t-1} - R_{k\neq i,t-1}) \tag{2.7.4}$$

For example, if housing returns outperform those of equities over a certain time period, investors would expect this to be the case in the future, which means that the housing investment class will become popular and housing sentiment – elevated.

All of these theories are consistent with a low or negative co-movement in expected returns across asset classes. Expected returns are simply a function of past asset-specific factors, and there is no reason for them to co-move positively across asset classes. Extrapolation of non-fundamental factors – such as return surprises – or style investing can also generate a correlation in expected returns that is lower that that of realised returns or fundamentals, consistent with the stylised fact in Table 2.6.1.

Extrapolative expectation formation can not only match the low co-movement fact in theory, but also matches a number of other salient features of the new cross-asset data. To test for this, I first construct a refined expected return measure for each asset class – in line with the literature, labelled as "sentiment" – and then test how this measure co-moves with past returns, return surprises and across asset classes.

A measure of asset-specific sentiment. I follow López-Salido, Stein, and Zakrajšek (2017) and measure asset-specific sentiment as the inverse of the expected return on asset class i. To do this, I first estimate the expected return on each asset class i using the familiar toolkit of return predictability regressions. I predict returns on each asset class using their own valuation and the macro-financial risk factors F, which all have considerable predictive power for returns on individual asset classes (Figure 2.6.2). The predictive regression takes the following form:

$$r_{i,j,t+1} = \beta_i dp_{i,j,t} + \sum_{f=1}^{F} \beta_{f,i} F_{j,t} + u_{i,j,t} \tag{2.7.5}$$

As in López-Salido, Stein, and Zakrajšek (2017), the corporate bond return is approximated as negative spread growth, $\hat{r}_{bond,j,t+1} \approx -\widehat{\Delta spread_{j,t+1}}$. Expected return is the year-ahead prediction from regression (2.7.5):

$$\hat{r}_{i,j,t+1} = \hat{\beta}_i dp_{i,j,t} + \sum_{f=1}^{F} \hat{\beta}_{f,i} F_{j,t} \tag{2.7.6}$$

When expected returns are high, risk premiums are also high, and hence sentiment – low or depressed. Sentiment is then simply the negative expected return from equation (2.7.6):

$$\widehat{sent}_{i,j,t} = -\hat{r}_{i,j,t+1} \tag{2.7.7}$$

Reformulated in terms of sentiment, the empirical predictions of the different extrapolation theories above can be summarised as follows.

Prediction 1. Extrapolation of returns: High returns at t increase investor demand for asset class i, and hence predict high sentiment at $t+1$.

Prediction 2. Diagnostic expectations: Higher than expected returns at t increase investor demand for asset class i, and hence predict high sentiment at $t+1$.

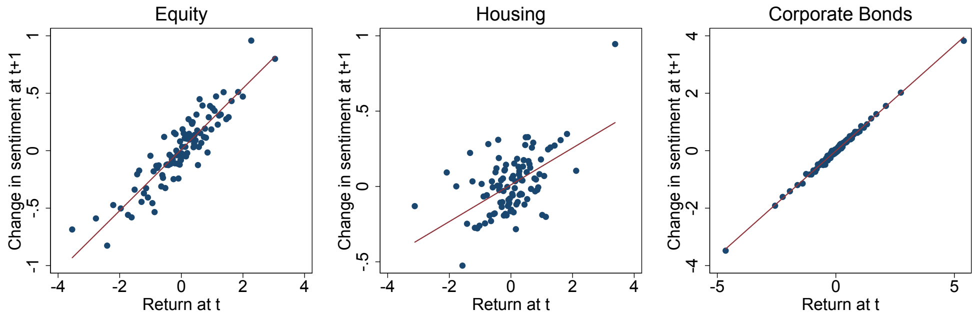
Prediction 3. Style investing: elevated sentiment on asset i is a sign of high expected returns relative to other asset classes, and hence predicts depressed sentiment on asset j.

Evidence for return extrapolation. Figure 2.7.2a tests prediction 1 by documenting the correlation between past returns and future sentiment. Importantly, I look at correlation with changes rather than levels of sentiment (results do, however, also hold in levels). High past returns at t would mechanically lower the asset valuation at t, and because valuations are highly autocorrelated, also the valuation at $t+1$, which would mechanically lead to elevated levels of sentiment at $t+1$. To avoid this from biasing my results, I instead assess the correlation between the return from $t-1$ to t, and the change in sentiment from t to $t+1$. If there is no extrapolative behaviour, these two variables should be uncorrelated. But Figure 2.7.2a shows that they, in fact, do show a strong positive co-movement in the data.

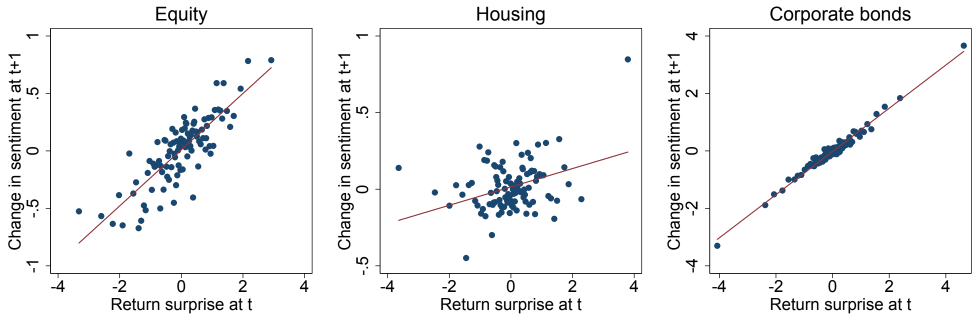
Figure 2.7.2a shows the binned scatter plot of returns at t and changes in sentiment at $t+1$, for each of the three asset classes. To ease exposition, data are grouped

Figure 2.7.2. Behavioural drivers of asset-specific sentiment

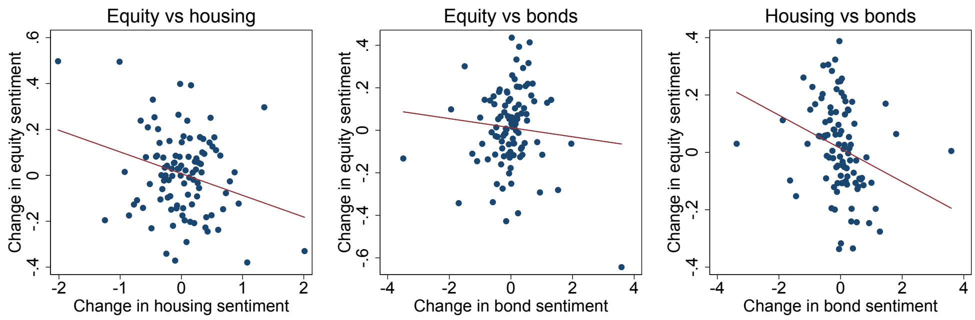
(a) Changes in sentiment and past returns (extrapolation)



(b) Changes in sentiment and past return surprises (diagnostic expectations)



(c) Co-movement of sentiment across asset classes (style investing)



Note: Binned scatter plots, number of bins equals 100. High asset-specific sentiment at t means a low expected return on asset i at $t + 1$, estimated by regressing returns at $t + 1$ on asset-specific valuations, macro-financial and other risk factors at t.

into 100 bins and each series is normalised such that +1 on the x or y axis corresponds to a 1 standard deviation increase in returns or sentiment. As predicted by equation (2.7.2), higher returns at t forecast elevated sentiment at $t+1$. The relationship is both statistically and economically significant for each asset class. A one standard deviation increase in the asset-specific return increases sentiment by 0.3 standard deviations for equities or housing, and by 1 standard deviation for corporate bonds. Echoing the results of Table 2.5.2 in Section 2.5 , the degree of extrapolation is highest for corporate bonds. This suggests that high excess volatility and mean reversion in corporate bond spreads are largely driven by extrapolative behaviour – a finding consistent with models and US data on investor sentiment in the bond market (Greenwood and Hanson, 2013; López-Salido, Stein, and Zakrajšek, 2017).

Evidence for diagnostic expectations. Figure 2.7.2b shows the correlation between past return surprises and sentiment for each asset i. The "surprise" return is simply the residual from the predictive regression in (2.7.5):

$$\hat{r}^{surprise}_{i,j,t+1} = r_{i,j,t+1} - \hat{\beta}_i dp_{i,j,t} - \sum_{f=1}^{F} \hat{\beta}_{f,i} F_{j,t} \tag{2.7.8}$$

Figure 2.7.2b shows that these unexpected returns are positively correlated with future changes in sentiment, again using scatter plots of normalised data divided into 100 bins. A one standard deviation increase in surprise return at t is associated with 0.3 standard deviations elevated equity sentiment, 0.2 standard deviations elevated housing sentiment and 1 standard deviation elevated corporate bond sentiment. with all the relationships statistically significant. This strong positive correlation is in line with prediction 2. The size and significance of the relationship suggests that diagnostic expectation formation can explain a large part of the excess asset price volatility in my data.

Evidence for style investing. Figure 2.7.2c considers evidence for style investing by correlating changes in sentiment across asset classes. It compares the changes in sentiment on equity and housing (left panel), equity and corporate bonds (middle panel), and housing and corporate bonds (right-hand panel), again using normalised data divided into 100 bins. Consistent with prediction 3, the contemporaneous correlation in sentiment across asset classes is negative: for example, a one standard deviation increase in housing sentiment predicts a one-quarter standard deviation fall in equity market sentiment, with each negative correlation, again, statistically significant. This negative co-movement in sentiment speaks against discount rate based theories of excess volatility, which would, instead, predict a strongly positive co-movement in expected returns across asset classes. The strength of the correlation between sentiment of different asset classes is, however, weaker than that between past realised or surprise returns in Figures 2.7.2a and 2.7.2b. Moreover, low past returns on one asset class do not, generally, predict high future sentiment on other asset classes, which is one feature of the style investing model in Barberis and Shleifer (2003).[24] This leads me to conclude that the evidence for style investing is somewhat weaker than that for extrapolative and diagnostic expectations – but these types of cross-asset linkages are also more difficult to test for in my broad macro-historical data.

Taken together, the analysis in this section suggests that time variation in expected returns is not driven by volatile discount rates. The prominent asset pricing theories that rely on such discount rate variation will struggle to match the low co-movement of discount rates, even allowing for some variation in asset-specific riskiness and liquidity premiums. Models with investor heterogeneity and segmented markets also only play a limited role in explaining the puzzle. Instead, the explanation lies in embracing volatile expectations, not volatile discount rates, as the primary driver of financial volatility. I next consider whether these volatile expectations have an effect not only on asset prices, but also on real activity.

2.8 Real effects of time-varying sentiment

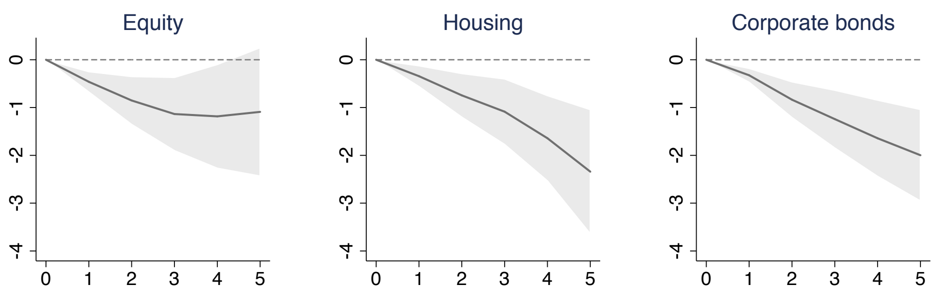
Volatile expectations – or time-varying sentiment – is an important driver of asset price volatility. These swings in asset prices, by themselves, directly affect the structure and distribution of wealth (Piketty, 2014; Benhabib and Bisin, 2016; Garbinti, Goupille-Lebret, and Piketty, 2017; Kuhn, Schularick, and Steins, 2017), and carry implications for financial market efficiency (Shiller, 2000). The work of Minsky (1977), however, puts these changes in expectations and sentiment not only at the center of asset price fluctuations, but also those in real activity. Under Minsky's financial instability hypothesis, elevated sentiment and optimistic expectations are accompanied by booming economic activity, but when these expectations are eventually reversed, the economy enters a crisis phase, with both real activity and health of the financial system deteriorating markedly. To test whether Minsky's hypothesis is supported by the new macro-financial data in this paper, I test whether elevated levels of sentiment are, indeed, associated with future recessions and financial crises.

2.8.1 Sentiment and GDP growth

To assess how the influence of sentiment on future real activity, I follow López-Salido, Stein, and Zakrajšek (2017) and regress GDP growth on the past level of asset-specific sentiment for each of the three risky asset classes. Unlike López-Salido, Stein, and Zakrajšek (2017), I estimate the full dynamic path of future GDP 1 to 5 years ahead to enable a more parsimonious assessment of the effects. To do this, I estimate a Jordà (2005) local projection response of cumulative GDP growth for

24. Results available upon request.

Figure 2.8.1. GDP response to elevated asset-specific sentiment



Note: Response of cumulative real GDP growth to a one standard deviation increase in asset-specific sentiment at $t = 1$. Estimated using local projections. Shaded areas are 90% confidence intervals. High sentiment at $t = 1$ means a low asset-specific discount rate at $t = 0$.

years $t+1$ to $t+5$ as a function of sentiment at t:

$$\Delta\text{GDP}_{j,t+h} \;=\; \beta_1^h + \beta_{gdp}^h \Delta\text{GDP}_t + \sum_{i=eq,hous,bond} \beta_i^h \widehat{sent}_{i,j,t} + \epsilon_{j,t}$$

Figure 2.8.1 shows the response of real GDP to a one standard deviation increase in sentiment $\widehat{sent}_{i,j,t}$ on each of the three risky asset classes, over a horizon of up to five years. Elevated sentiment – or low risk premiums – in each asset market tends to be followed by low, rather than high real GDP growth. The effect of time-varying equity sentiment on future GDP is small and relatively short-lived, peaking at less than 1% GDP. Effects of time varying housing and corporate bond market sentiment are, however, much larger and more persistent. A one standard deviation increase in housing or corporate bond market sentiment forecasts around two percentage points lower real GDP over the horizon of 5 years.

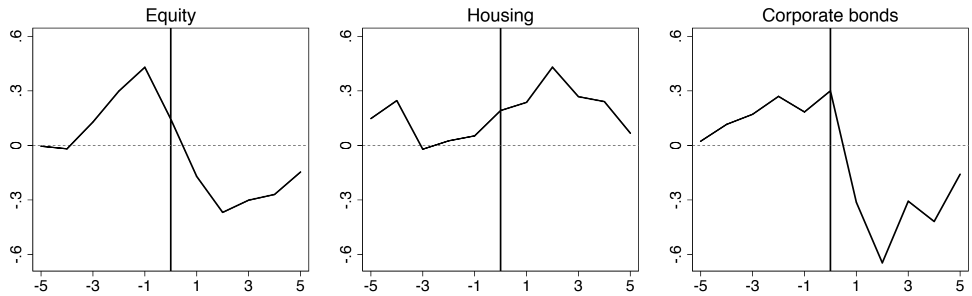
The impulse responses in Figure 2.8.1 suggest that low risk premiums and elevated market sentiment forecast poor, rather than favourable economic performance. My results cannot be interpreted causally but again, this evidence is consistent with behavioural explanations for the time varying risk puzzle: low risk premiums are precursors of falling asset prices and GDP, rather than precursors of favourable economic and financial fundamentals It turns out that the two markets for which excess volatility has, up to this point, been relatively sparsely documented and less well explored – those of corporate bonds and housing – are also the ones that matter most for the macroeconomy.

2.8.2 Sentiment and financial crises

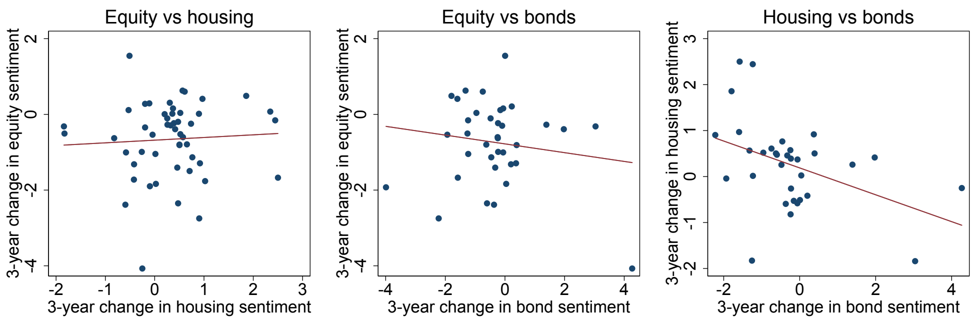
Muir (2017) documents that equity risk premia tend to increase sharply around financial crises, and Krishnamurthy and Muir (2017) document a similar pattern for corporate bond spreads. Muir (2017) also shows that risk premia do not show similar spikes during consumption disasters, and interprets the high financial-crisis

Figure 2.8.2. Sentiment and financial crises

(a) Average level of sentiment around financial crisis events



(b) Co-movement of sentiment during financial crises



Note: Financial crisis at $t = 0$. All sentiment variables are standardised to mean 0 and standard deviation of 1 in the full sample. Correlations are for the 3-year change around the crisis year (from $t = -1$ to $t = +1$). Each point on the scatter is an individual crisis event. High asset-specific sentiment at t means a low expected return on asset i at $t + 1$, estimated by regressing returns at $t + 1$ on asset-specific valuations, macro-financial and other risk factors at t.

premia as evidence for intermediary asset pricing theories, and a fact difficult to reconcile with behavioural theories. Here I re-examine these findings using the new dataset in this paper, which offers a more extensive cross-country coverage, and a more refined measure of risk premiums – asset-specific sentiment described in Section 2.7.3, which covers housing as well as equities and corporate bonds. I show that while sentiment does tend to fall around financial crises – or, in other words, risk premiums tend to increase – it is volatile expectations rather than intermediary balance sheets that are likely to underly these trends.

Figure 2.8.2a plots the evolution of sentiment around financial crises. It depicts the average level of sentiment for each asset class over the time window of 5 years before to 5 years after the crisis ($t = 0$), with crisis dates taken from Jordà, Schularick, and Taylor (2016). As described in Section 2.7.3, sentiment measures the expected return on each asset class, predicted using own valuation and a broad range of macro-financial factors. High levels of sentiment correspond to low expected returns and low discount rates or risk premia, with the sentiment measure standardised to mean 0 and standard deviation of 1 across the full sample.

The behaviour of sentiment across all three risky asset classes shares a number of characteristics with a Minsky (1977) boom-bust cycle. There is “froth” in the build-up, with sentiment elevated above the sample average, followed by a sharp reversal in expectations and depressed levels of sentiment over subsequent years. The froth is most strongly manifested in equity markets, while the expectation reversal is strongest for corporate bonds. This is consistent with the notion of equity markets as best incorporating forward-looking expectations, and corporate bonds being most responsive to extreme macroeconomic events. The delayed slow rise and gradual reversal in housing sentiment indicates a build-up of vulnerabilities during the early crisis phase, which are then slow to adjust during the crash. The findings are consistent with those of Muir (2017) and Krishnamurthy and Muir (2017), who find that equity and corporate bond risk premiums spike around the crisis using simpler dividend-price ratio measures. Appendix 2.A.6 shows that the simpler valuation and discount rate news measures behave in a similar ways to the sentiment measure in Figure 2.8.2 around crisis events.

But are these changes in sentiment best interpreted as effects of financial frictions, or time-varying expectations. In his analysis, Muir (2017) favours the financial-friction based interpretation of these events, while this paper argues that most of such asset price swings can be traced back to expectation volatility. To distinguish between these two hypotheses, Figure 2.8.2b plots the correlation in changes in sentiment during the 3 years surrounding the start of the crisis. This correlation, just like the broader cross-asset discount rate correlations in the full sample, turns out to be roughly zero, and sometimes negative. The lack of correlation carries over to simpler measures of asset-specific discount rates such as valuation ratios and discount rate news, presented in Appendix 2.A.6. Some of this low correlation can be attributed to the slightly different timing in sentiment reversals across asset classes. But especially when it comes to corporate bonds and equities, the lack of co-movement during financial crisis largely reflects the fact that crisis-related build-ups in sentiment tend to be asset-specific phenomena, with some crises resulting in high equity or corporate bond risk premiums, but relatively few displaying high risk premiums on all three risky asset classes.

The zero sentiment correlation strongly hints that financial crises, and the associated asset price booms and busts can be traced back to time variation in expectations, rather than financial frictions. The fact that this variation is especially large is then a factor contributing to the crisis, rather than a crisis outcome. The reversal in expectations during the crisis is so large that it generates distress and bankruptcies in the banking system, and carries a large economic cost. It is highly likely that financial frictions play an important role in generating this cost (Eggertsson and Krugman, 2012; Mian and Sufi, 2014). This paper merely argues that these frictions are relatively unimportant for the asset price swings around the crisis, which are driven largely by agents' expectations.

Taken together, the real consequences of sentiment variation tend to be persistent and sizeable, and manifest themselves both during general business cycle fluctuations and more severe crisis events. Elevated sentiment does not entail the anticipation of an economic boom, but is instead a forebearer of asset price declines, financial crises and economic downturns. The associated boom-bust cycle echoes Minsky (1977)'s financial instability hypothesis, with elevated sentiment fostering an economic boom while sowing the seeds of a future financial crisis, which materialises when once the overoptimistic expectations are reversed. This shows that volatile expectations on all three risky asset classes not only affect the prices of these assets, but also broader economic outcomes. Incorporating this expectation volatility into business cycle models could, therefore, foster a deeper understanding of the underlying drivers behind macroeconomic fluctuations and their links with financial market outcomes.

2.9 Conclusion

This paper has introduced a new historical cross-country dataset of returns, cashflows and valuations of three major risky asset classes: equity, housing and corporate bonds, and used it to study the determinants of booms and busts in prices of these assets. These data have both cast an old puzzle in new light, and put forward a novel asset pricing anomaly.

The old excess volatility puzzle of Shiller (1981) stands firm in these new data. In some ways, excess volatility is not as bad as previously thought: time-varying discount rates only explain one-third of the variation in equity dividend-price ratios, rather than all of it. In other ways, it is worse than previously thought: the relatively unexplored housing and corporate bond markets display far higher excess volatility than equities, with valuations determined by expectations of returns much more than fundamentals.

The new puzzle that emerges from these data is that discount rates on different asset classes do not co-move. This lack of co-movement cannot be explained by idiosyncratic fundamentals, and does not disappear once conditioned on a broad range of macro-financial risk factors. It is at odds with most asset pricing theories, which tend to rely on the discount rate – or time variation in the price of risk – to explain why excess volatility arises. Model modifications which increase asset-specific heterogeneity but do not modify agents expectations are unlikely to fully resolve the puzzle.

Instead, my analysis suggests that behavioural theories, centered around extrapolative biases in investor expectations, provide an explanation for both the old volatility puzzle and the new time varying risk puzzle. Extrapolation of asset-specific returns or return surprises can generate the low co-movement in theory, and is consistent with a number of stylised features of the data. Such time variation in asset-specific sentiment contributes to volatility not only in financial markets, but also the real economy. A better understanding of the underlying drivers of these sentiment swings, much of which is left to future research, should therefore help shed light on the underlying determinants of macroeconomic, as well as financial, booms and busts.

Appendix 2.A Appendix

2.A.1 Accuracy of corporate bond data

Constructing a data series that captures the evolution of corporate bond credit risk premiums over time and across countries faces three main challenges. First, corporate bond data are subject to sample selection issues. Not all companies have access to the corporate bond market, and the type of company that has access may vary over time. Much of the early corporate bond market was dominated by railway companies. Later, railway bonds became less and bank bonds – more important. A second, related, bias is that outside of the US and the recent sample period, I do not have data on bond credit ratings. Therefore, the credit quality of the representative bond in the sample may change over time and across countries.

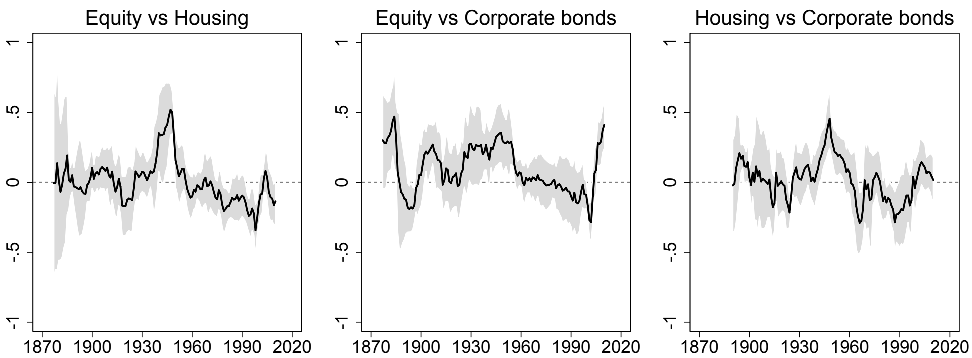
I guard against these biases in several ways. First, I utilise a wide array of new and previously unused sources that document bond prices and yields on domestic stock exchanges, in over the counter transactions, and for primary market issuance, complemented by data from international or foreign exchange listings. This gives me a comprehensively broad coverage of historical corporate bond transactions. Within this broad coverage, the scope for selection bias is less apparent. For example, the series that were constructed from microdata include non-financial non-railway bonds for every country and data period, and in general, the trends in the yields of non-railway bonds are very similar to the overall index. By excluding mortgage bonds and government-guaranteed bonds, I ensure that my sample always measures the credit risk faced by private sector bond issuers, regardless of selection. Additionally, the corporate bond risk premium shows no clear time trend – in fact, the long-run risk premium on this asset class is more stable than those for equities and housing.

To the extent that selection biases do exist, they seem to mainly affect cross-country rather than cross-time differences. Some countries, such as UK, US and Germany, have had active and diverse corporate bond markets throughout most of the sample. In other countries, such as Sweden and Australia, market participation was generally tilted towards larger, safer companies. Several countries, such as Portugal and Spain, had comparatively small but diverse corporate bond markets, which included a wide variety of credit risks. To guard against remaining selection biases, I always demean the series within country, and as a robustness check, also across country-specific time periods, using the algorithm of Bai and Perron (2003) to detect structural breaks. I also test the validity of my results across a variety of subsamples and historical time periods. The focus throughout the paper is on variation across time rather than across countries: most statistics shown are cross-country averages, and regressions include country fixed effects, with various robustness checks for stability across time periods. Even though it is not possible to eliminate selection bias completely, the impact of this bias on my findings is likely to be small.

A third potential difficulty relates to calculating yields to maturity. Accurate maturity data are difficult to come by for some historical sources, and often do not account for embedded options and bond conversions. In one sense, the long time dimension of my data helps guard against such biases. For countries where I have the microdata, I can observe the first and last trade for each bond, and hence when the bonds were effectively matured or a redemption option exercised. These allow me to obtain additional, and improve existing bond maturity proxies. Some data series also contain information on options, in which case I follow the usual practice of taking the option date as the maturity date if the bond is trading above par (as in, for example, Klovland, 2004). A number of publications also include option-adjusted effective yield estimates, even for historical data, and even for individual bonds (see, for example Mediobanca, Various years). For the early historical period, relatively few bonds had embedded conversion options. That being said, I do sometimes have to rely on current yields, and some of the secondary sources do not specify whether the yield is calculated as a yield to maturity, or a simple current yield. Over the whole historical sample, the biases arising from uncertainty around maturity dates, and the use of current yield data, are likely to be small.

2.A.2 Discount rate co-movement: additional results

Figure 2.A.1. Co-movement of asset-specific discount rate proxies, first differences



Note: Pairwise correlation coefficients between 3-year changes in the dividend-price ratio, rent-price ratio and corporate bond spread over rolling decadal windows (e.g. the value for 1875 if the correlation over the window 1870–1880). Shaded areas are 90% confidence intervals, using country-clustered standard errors.

Table 2.A.1. Discount rate and cashflow news correlations within each country

	Discount rate news			Cashflow news
	Equity & housing	Equity & corporate bonds	Housing & corporate bonds	Equity & housing
Australia	0.14	0.22**	0.08	0.15
Belgium	0.61***	0.10	0.15	0.57***
Canada		0.21***		
Denmark	−0.01			0.09
Finland	0.16	−0.23	−0.48***	0.29***
France	0.22*	0.06	−0.16	0.40***
Germany	0.09	0.08	−0.09	−0.01
Italy	0.43**	0.28**	0.34**	0.29*
Japan	0.28***	−0.06	−0.02	0.19*
Netherlands	−0.11	0.30	−0.26	0.35***
Norway	−0.05	0.13	−0.18	0.19*
Portugal	−0.34**	−0.04	−0.14	0.47***
Spain	−0.06	−0.24*	0.28***	0.28***
Sweden	0.05	0.15	0.24**	0.16
Switzerland	−0.14	−0.20	0.28**	0.16
UK	0.17	0.00	−0.07	0.40***
USA	0.07	0.16*	0.21***	0.31***
Sig. > 0 / Total	4/16	4/16	5/15	11/16
Sig. < 0 / Total	1/16	1/16	1/15	0/16
Not sig. / Total	11/16	11/16	9/15	5/16

Note: Pairwise correlation coefficients. Underlying data are 3-year moving averages, to smooth over timing idiosyncracies across assets. Discount rate and cashflow news for equities and housing are estimated as the innovations to present value of future returns and cashflows, respectively, for each asset, using a VAR in returns, cashflow growth and valuations, and present value moment constraints. Discount rate news for bonds is the change in the spread. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$

Table 2.A.2. Discount rate and cashflow news correlations within each country, after 1950

	Discount rate news			Cashflow news
	Equity & housing	Equity & corporate bonds	Housing & corporate bonds	Equity & housing
Australia	0.15	0.25*	0.02	0.06
Belgium	0.50***	−0.24*	0.02	0.14
Canada		0.06		
Denmark	−0.08			0.27**
Finland	0.10	−0.23	−0.48***	0.31***
France	−0.24*	−0.12	−0.23**	0.02
Germany	−0.29**	0.01	0.11	−0.01
Italy	0.30*	0.11	0.30**	0.28**
Japan	0.51***	0.07	−0.01	0.13
Netherlands	−0.09	0.30	−0.26	0.31***
Norway	−0.26**	0.10	−0.39**	0.32***
Portugal	−0.34**	−0.02	−0.14	0.47***
Spain	−0.10	0.01	0.16	0.26*
Sweden	−0.13	0.12	0.02	0.04
Switzerland	−0.30**	−0.20	0.33***	−0.17
UK	−0.07	0.00	−0.09	0.14
USA	0.07	0.09	0.21	0.34**
Sig. > 0 / Total	3/16	1/16	2/15	8/16
Sig. < 0 / Total	5/16	1/16	3/15	0/16
Not sig. / Total	8/16	14/16	10/15	8/16

Note: Pairwise correlation coefficients in the post-1950 sample. Underlying data are 3-year moving averages, to smooth over timing idiosyncracies across assets. Discount rate and cashflow news for equities and housing are estimated as the innovations to present value of future returns and cashflows, respectively, for each asset, using a VAR in returns, cashflow growth and valuations, and present value moment constraints. Discount rate news for bonds is the change in the spread. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$

2.A.3 Return predictability: additional results

2.A.3.1 Corporate bond cashflow predictability

Table 2.A.3. Predicting corporate default rates with yield spreads, US data

	(1)	(2)
	Default rate, t+1	Δ Default rate, t+1
spread_t	1.485***	
	(0.274)	
$\Delta\ \text{spread}_t$		0.195
		(0.331)
R^2	0.430	0.002
Observations	142	141

Note: Dependent (y) variables are the one-year ahead level and absolute change in the corporate bond default rate. The default rate is calculate as the par value of bonds in default relative to total outstanding. Data are for US only. Predictor (x) variables are the level and the change in the corporate bond spread. OLS regressions with heteroskedasticity-robust standard errors in parentheses. *: $p <$ 0.1 **: $p < 0.05$ ***: $p < 0.01$.

2.A.3.2 Robustness to alternative specifications

Table 2.A.4 assess whether the baseline predictability result holds under different regression specifications. Column 1 shows the baseline estimates in Table 2.5.1, and each of the columns 2-10 changes the regression specification or variable definition to assess whether the baseline results still hold. Table 2.A.4 column 2 tests for predictability in nominal, rather than real returns and cashflows, and finds that regression coefficients are essentially unchanged. A similar result comes out of using excess returns over government bills, rather than returns in excess of inflation, as a dependent variable (Table 2.A.4 column 3). This suggests that changes in expected returns are best interpreted as movements in asset-specific risk premiums, rather than shifts in the safe interest rate.[25]

Column 4 of Table 2.A.4 shows that risky asset valuations also predict returns and cashflows at longer horizons. It regresses 5-year ahead average real return, cashflow growth or spread growth on the valuation dp_t. Over the longer horizon, the variation in excess returns and cashflows becomes even more powerful. A 1 percentage point higher dividend-price ratio forecasts 5 percentage points higher cumulative returns, and 10 percentage points lower cumulative real dividend growth 5 years ahead. 1 percentage point higher rent-price ratios forecast 8 percentage point higher cumulative returns, and 5 percentage point lower real rental growth. Elevated corporate

25. In the present value equation (2.2.1), the discount rate DR is the sum of the safe rate R^S and a risk premium RP_i.

Table 2.A.4. Return and cashflow predictability: alternative specifications

	(1)	(2)	(3)	(4)	(5)
Panel 1: Different definitions and specifications					
	Baseline	Nominal returns & cashflows	Excess returns	5-year ahead average growth	Constrained VAR
Equity:					
r_{t+1}	0.055***	0.037***	0.038***	0.039***	0.058***
	(0.014)	(0.013)	(0.011)	(0.012)	(0.012)
dg_{t+1}	-0.120***	-0.138***		-0.081***	-0.133***
	(0.030)	(0.029)		(0.015)	(0.020)
Housing:					
r_{t+1}	0.066***	0.067***	0.068***	0.064***	0.062***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.007)
dg_{t+1}	-0.023**	-0.022***		-0.042***	-0.029***
	(0.009)	(0.008)		(0.011)	(0.008)
Corporate bonds:					
r_{t+1}	0.029***	0.015***	0.021***	0.023***	
	(0.007)	(0.003)	(0.004)	(0.007)	
$\Delta spread_{t+1}$	-0.259***			-0.139***	
	(0.042)			(0.012)	
Panel 2: Different time periods					
	(6)	(7)	(8)	(9)	(10)
	Year fixed effects	Structural breaks	Post-1950	Expansions	Recessions
Equity:					
r_{t+1}	0.061***	0.110***	0.077***	0.057***	0.031
	(0.016)	(0.021)	(0.013)	(0.015)	(0.034)
dg_{t+1}	-0.153***	-0.221***	-0.111***	-0.125***	-0.094*
	(0.043)	(0.038)	(0.042)	(0.031)	(0.052)
Housing:					
r_{t+1}	0.069***	0.118***	0.066***	0.061***	0.083***
	(0.013)	(0.025)	(0.010)	(0.010)	(0.022)
dg_{t+1}	-0.023***	-0.024	-0.022	-0.027***	-0.017
	(0.007)	(0.028)	(0.015)	(0.009)	(0.014)
Corporate bonds:					
r_{t+1}	0.023***	0.029***	0.020***	0.024***	0.039***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.014)
$\Delta spread_{t+1}$	-0.264***	-0.298***	-0.327***	-0.233***	-0.345***
	(0.043)	(0.045)	(0.064)	(0.048)	(0.074)

Note: Predictive coefficents on the log dividend-price ratio for equity, log rent-price ratio for housing, and percentage point spread for corporate bonds. Dependent variables in rows. Specifications in columns. r_{t+1} is log real total return; dg_{t+1} is log real dividend or rent growth, Δspread$_t$ is the change in corporate bond credit spread. Baseline is OLS with fixed effects. Excess returns are net of short-term government bill rates. Constrained VAR estimates the VAR in three variables – r, dg and dp, using GMM subject to present value moment constraints. Structural breaks adjust not only for country-specific, but also time-specific movements in the mean of each variable, with break dates identified using the Bai-Perron procedure. Expansions and recessions are dated using the Bry-Boschan algorithm. Country clustered standard errors in parentheses. VAR estimation also allows for clustering across time. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

bond spreads more or less mean revert within 5 years: a 1 percentage point higher bond spread at t forecasts 0.7 percentage points lower spreads at $t+5$. Finally, Table 2.A.4 column 5 assesses the equity and housing return predictability using a VAR in dp, r and dg, which respects the present value moment constraints implied by equation (2.2.3). The details of the VAR estimation are provided in Appendix 2.A.3.3. The VAR allows for a richer dependence structure across variables (dp, r and dg for each asset class) and over time, but the resulting coefficient estimates are the same as the OLS in Column 1.

Panel 2 of Table 2.A.4 tests for coefficient stability across different time periods. Column 6 starts by taking out the common cross-country time variation through year fixed effects, and finds the results unchanged. Column 7 adjusts each series for structural breaks in the valuation ratio dp_t, following the methodology established by Lettau and Van Nieuwerburgh (2008) for the US equity market. To do this, I first identify the break dates for each country using the Bai and Perron (2003) procedure, allowing for a maximum of 3 country-specific breaks. I then demean the variables dp_t, r_t and dg_t within each country and structural break specific time period, before running the usual regressions (2.2.4) and (2.2.5) on the adjusted data.[26] The results for adjusted data are similar to baseline, and if anything find somewhat stronger return predictability, with a 1 percentage point increase in the dividend-price or rent-price ratio forecasting 3 percentage point lower returns, and a 1 percentage point elevated bond spread forecasting a 0.33 percentage point spread decline in the following year.

Table 2.A.4 columns 8–10 test for the stability of results across different time periods, to check whether the baseline results are applicable to the more recent data, and across the different stages of the business cycle. This is generally the case, with return predictability somewhat stronger after 1950, and relatively stable across expansions and recessions.[27] If anything, predictability is somewhat more robust during expansions than recessions, which favours sentiment and expectation-based explanations of excess financial volatility: whereas we would expect discount rate effects to be strongest in recessions when consumption is hit hard, the overoptimism and irrational expectations are likely to be more prevalent during expansions.

Table 2.A.5 tests the predictability of returns on each asset class within individual countries. The coefficients on all return, dividend growth and spread growth variables in columns 1–6 are of similar magnitude to the baseline panel regression in Table 2.5.1. Because of the annual data frequency and the lower number of obser-

26. For example, similarly to the findings of Lettau and Van Nieuwerburgh (2008), the US dividend-price ratio displays two structural breaks, in 1954 and 1995. The structural break adjusted regression demeans the dp_t variable in the 1870–1954 period using the 1870–1954 mean, the 1955–1995 dp_t using the 1955-1995 series mean, and the remainder using the post-1995 mean.

27. The recession and expansion periods are identified by applying the Bry and Boschan (1971) algorithm to real GDP per capita (see Jordà, Schularick, and Taylor, 2013, for further details on the application to this dataset).

Table 2.A.5. Return and cashflow predictability in individual countries

	(1)	(2)	(3)	(4)	(5)	(6)
	Equity		Housing		Corporate bonds	
	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}	r_{t+1}	$\Delta spread_{t+1}$
Australia	0.202***	-0.012	0.035	-0.030	0.023**	-0.328***
	(0.061)	(0.065)	(0.027)	(0.027)	(0.011)	(0.082)
Belgium	0.068	-0.123	-0.045	-0.152***	-0.015	-0.356***
	(0.056)	(0.142)	(0.043)	(0.054)	(0.058)	(0.088)
Canada	0.054	-0.220***			0.042***	-0.187
	(0.038)	(0.047)			(0.013)	(0.134)
Denmark	0.002	-0.132***	0.062***	-0.002		
	(0.031)	(0.046)	(0.015)	(0.007)		
Finland	0.066	-0.218**	0.104***	-0.034	0.003	-0.259
	(0.068)	(0.110)	(0.028)	(0.029)	(0.010)	(0.260)
France	0.210***	0.063	0.157***	0.085***	-0.002	-0.129**
	(0.042)	(0.066)	(0.036)	(0.027)	(0.020)	(0.050)
Germany	-0.048	-0.350***	0.095***	0.008	0.038***	-0.166**
	(0.052)	(0.077)	(0.022)	(0.009)	(0.012)	(0.075)
Italy	0.053	-0.158	0.043***	-0.031	0.058***	-0.566***
	(0.087)	(0.115)	(0.011)	(0.024)	(0.017)	(0.115)
Japan	0.049**	-0.046*	0.072***	-0.026	0.024	-0.139*
	(0.024)	(0.025)	(0.027)	(0.017)	(0.016)	(0.073)
Netherlands	0.043	-0.348***	0.073***	-0.024		
	(0.057)	(0.070)	(0.021)	(0.017)		
Norway	0.027	-0.131**	0.119***	0.019	0.087***	-0.317***
	(0.053)	(0.058)	(0.038)	(0.027)	(0.017)	(0.081)
Portugal	0.039	0.003	0.010	-0.080**	0.045***	-0.307***
	(0.051)	(0.111)	(0.037)	(0.032)	(0.017)	(0.106)
Spain	0.116***	-0.112*	0.067**	-0.015	-0.015	-0.032
	(0.044)	(0.065)	(0.028)	(0.015)	(0.013)	(0.118)
Sweden	0.005	-0.299***	0.055**	-0.015	0.041***	-0.125*
	(0.057)	(0.052)	(0.027)	(0.015)	(0.014)	(0.071)
Switzerland	0.003	-0.105	0.041	-0.042*	0.026***	-0.283***
	(0.058)	(0.075)	(0.029)	(0.022)	(0.007)	(0.103)
UK	0.210***	-0.202***	0.099***	-0.001	0.013	-0.357***
	(0.067)	(0.055)	(0.029)	(0.024)	(0.018)	(0.119)
USA	0.051	-0.090***	0.158***	-0.044	0.033***	-0.155***
	(0.035)	(0.028)	(0.059)	(0.032)	(0.005)	(0.057)
Significant/Total	5/17	11/17	12/16	4/16	9/15	12/15

Note: OLS regressions at country level. Predictor (x) variables are the log dividend-price ratio, log rent-price ratio and percentage point corporate bond spread. Dependent (y) variables in columns. r is the log real total return, dg is log real dividend or rental growth. Δ spread is the percentage point change in the bond spread. Heteroskedasticity-robust standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

Table 2.A.6. Return and cashflow predictability in individual countries, post-1950

	(1)	(2)	(3)	(4)	(5)	(6)
	Equity		Housing		Corporate bonds	
	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}	r_{t+1}	$\Delta spread_{t+1}$
Australia	0.374***	-0.002	0.023	-0.032	0.017	-0.269***
	(0.130)	(0.144)	(0.022)	(0.025)	(0.012)	(0.090)
Belgium	0.079	-0.036	0.064***	0.019**	0.036**	-0.361***
	(0.068)	(0.061)	(0.019)	(0.009)	(0.016)	(0.126)
Canada	0.069	-0.234***			0.050**	-0.156**
	(0.058)	(0.079)			(0.024)	(0.076)
Denmark	0.029	-0.156**	0.092***	0.003		
	(0.040)	(0.062)	(0.024)	(0.008)		
Finland	0.018	-0.300**	0.028	-0.139***	0.003	-0.259
	(0.084)	(0.141)	(0.065)	(0.050)	(0.010)	(0.260)
France	0.134**	-0.048	0.138***	0.073***	-0.009	-0.123**
	(0.066)	(0.044)	(0.029)	(0.025)	(0.009)	(0.052)
Germany	-0.049	-0.420***	0.046	-0.023*	0.030**	-0.136*
	(0.102)	(0.125)	(0.029)	(0.013)	(0.013)	(0.082)
Italy	0.119	-0.143*	0.087**	-0.114*	0.030**	-0.722***
	(0.081)	(0.074)	(0.037)	(0.061)	(0.014)	(0.142)
Japan	0.095***	0.025	0.121***	0.047*	0.003	-0.099
	(0.031)	(0.023)	(0.040)	(0.024)	(0.013)	(0.095)
Netherlands	0.074	-0.376***	0.111***	0.023**		
	(0.068)	(0.078)	(0.025)	(0.010)		
Norway	0.112	-0.097	0.068*	-0.002	0.101***	-0.278**
	(0.099)	(0.077)	(0.035)	(0.011)	(0.017)	(0.120)
Portugal	0.063	0.039	0.012	-0.077**	0.039***	-0.392***
	(0.082)	(0.185)	(0.038)	(0.033)	(0.012)	(0.144)
Spain	0.090*	-0.111***	0.055*	-0.016	-0.028***	-0.107
	(0.054)	(0.042)	(0.029)	(0.025)	(0.010)	(0.098)
Sweden	0.066	-0.279***	0.044	-0.008	-0.003	-0.395**
	(0.093)	(0.065)	(0.033)	(0.013)	(0.044)	(0.191)
Switzerland	0.072	-0.064	0.042	-0.040***	0.024***	-0.259**
	(0.090)	(0.146)	(0.029)	(0.009)	(0.007)	(0.109)
UK	0.284***	-0.031	0.073**	0.003	0.008	-0.379***
	(0.093)	(0.027)	(0.029)	(0.023)	(0.019)	(0.137)
USA	0.098**	-0.041**	0.096*	-0.030	0.041***	-0.388***
	(0.045)	(0.019)	(0.057)	(0.026)	(0.009)	(0.140)
Significant/Total	6/17	9/17	10/16	9/16	9/15	12/15

Note: OLS regressions at country level, post-1950 period. Predictor (x) variables are the log dividend-price ratio, log rent-price ratio and percentage point corporate bond spread. Dependent (y) variables in columns. r is the log real total return, dg is log real dividend or rental growth. Δ spread is the percentage point change in the bond spread. Heteroskedasticity-robust standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

vations, the statistical power of the estimates is somewhat smaller than that in the panel regression. Equity cashflows, housing returns, and bond return and spread growth are all robustly predictable. The coefficients in Table 2.A.5 columns 2, 3, 5 and 6 are significant in almost every country, with signs and magnitudes consistent with the panel. The same pattern holds if we limit the sample to the post 1950 period, with results shown in Table 2.A.6.

The evidence for within-country predictability of equity returns and housing cashflows is somewhat more mixed. The equity return coefficients in column 1 display the right sign and magnitude, but are only significant in one in every 3 countries. The coefficient on equity returns for the US is not significant, in contrast to the findings of Cochrane (2008). This difference mainly comes about from the data sample. Return predictability in the US becomes much stronger, and cashflow predictability – much weaker in the later parts of the sample. Most studies of the US data use the CRSP dataset which starts in 1925, and find strong evidence of return predictability, and only weak evidence for predictable dividend growth. My data are sourced from the S & P 500 index in Shiller (2000) which goes back to 1870. If I limit the estimation to the post-1925 sample, the β coefficient on US equity returns becomes significant, although it does not change much in size. For the longer post-1870 sample, the evidence for the US equity return predictability is weaker, and for dividend growth predictability – stronger, consistent with the findings of Golez and Koudijs (2018). Housing cashflows are only significantly predictable for one in every 4 countries.

The extent of excess volatility can be roughly approximated by comparing the size of the *r* and *dg* coefficients for housing and equities, or by looking at the degree of mean reversion for bonds (column 6). For stocks, French and Australian equity returns seem to vary most in excess of fundamentals, with coefficients on r_{t+1} 3 times the size of the panel estimate, and roughly zero *dg* coefficients. Denmark, Norway and Sweden, on the contrary, show strong cashflow predictability, but only a very weak correlation between equity valuations and year-ahead returns. Jordà, Knoll, et al. (2017) show that risky returns in Scandinavian countries have been relatively high and stable throughout history, which goes in line with the low levels of excess volatility for these countries. The US market has displayed the largest excess volatility when it comes to housing, with the booms and busts of the 1920s, 1990s and 2000s all occurring without substantial changes in future rental growth. The Italian corporate bond spreads are more excessively volatile than those in other countries, with more than half of any elevated spread level, on average, expected to mean revert during the following year. The Italian bond market has been dominated by relatively mature companies and banks throughout my data sample. Since this type of company is relatively unlikely to default, most of the variation in Italian bond spreads bears little relation to the underlying riskiness of the bond portfolio.

2.A.3.3 Discount rate and cashflow news, VAR estimation

Equation (2.2.3) decomposes risky asset valuation into a discount rate and cashflow component. Similarly, the variance in valuation ratios dp can be attributed to discount rate and cashflow news, as shown below:

$$\begin{aligned}\mathrm{Var}(dp_{i,t}) = \underbrace{\mathrm{Var}(\mathbb{E}\sum_{s=0}^{\infty}\rho_i^s r_{i,t+1+s})}_{\text{DR news}} + \underbrace{\mathrm{Var}(\mathbb{E}\sum_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s})}_{\text{CF news}} \\ - 2\mathrm{Cov}(\mathbb{E}\sum_{s=0}^{\infty}\rho_i^s r_{i,t+1+s}, \mathbb{E}\sum_{j=0}^{\infty}\rho_i^s dg_{i,t+1+s})\end{aligned} \quad (2.A.1)$$

To estimate the discount rate and cashflow news components of equity and housing valuations, I follow Golez and Koudijs (2018) and estimate a VAR in three variables $[r_{i,t}, dg_{i,t}, dp_{i,t}] \equiv z_{i,t}$:

$$z_{i,t} = Az_{i,t-1} + u_{i,t} \quad (2.A.2)$$
$$z_{i,t} = [r_{i,t}, dg_{i,t}, dp_{i,t}]' \quad (2.A.3)$$
$$\mathbb{E}(zz') = \Gamma; \quad \mathbb{E}(uu') = \Sigma; I = (e1, e2, e3) \quad (2.A.4)$$

The VAR is estimated using GMM, with the following 9 moment conditions:

$$E[(z_{i,t+1} - z_{i,t}) \otimes z_{i,t}] = 0 \quad (2.A.5)$$

The present value relation in (2.2.3) imposes three additional moment restrictions:

$$(e1' - e2' + \rho e3')A = e3' \quad (2.A.6)$$

I estimate the VAR using 6-equation GMM subject to the constraints in (2.A.6), and accounting for time and cross-sectional dependence in standard errors. The resulting estimates allow me to do two things. First, I can estimate the relative contribution of discount rate and cashflow news to the variance of the dividend or rent to price ratios in (2.A.1):

$$\mathrm{Var}(dp_{i,t}) = e3\Gamma e3 = \underbrace{e1'A(I-\rho A)^{-1}\Gamma e3}_{\text{DR news}} - \underbrace{e2'A(I-\rho A)^{-1}\Gamma e3}_{\text{CF news}} \quad (2.A.7)$$

Here, A is the VAR coefficient matrix, and Γ is the covariance matrix of the regressors in (2.A.4).

Second, following Campbell (1991), I can derive a "clean" series of the time-varying risk premium and fundamental component of returns on each asset – more precisely, the discount rate and cashflow news in the unexpected asset returns

$r_{i,t+1} - \mathbb{E}r_{i,t+1}$:

$$r_{i,t+1} - \mathbb{E}r_{i,t+1} = \underbrace{-e1'\rho_i A(I - \rho_i A)^{-1} u_{i,t+1}}_{\text{DR news}} + \underbrace{(e1 + e1'\rho_i A(I - \rho_i A)^{-1}) u_{i,t+1}}_{\text{CF news}} \tag{2.A.8}$$

As a robustness check, I also use an excess return VAR specification $z_{i,t} = [r^e_{i,t}, r^f_t dg_{i,t}, dp_{i,t}]'$ to check if changes in discount rates are driven by the risk-free rate r^f_t, rather than risk premiums. Here, $r^e_{i,t}$ is the excess return over the risk-free bill rate, $r^e_{i,t} = log(1 + R_{i,t} - R^{\text{bill}}_{i,t})$, and r^f_t is the the real short-term risk free rate $r^f_t = log[R^{\text{bill}}_t - \pi_t]$. It turns out that the risk-free rate is not forecastable by the valuation ratios, which means that we can think of the "discount rate news" term above as risk premium news. Results are available upon request.

For corporate bonds, Nozawa (2017) shows that the credit spread can be expressed as a sum of expected returns, and expected default risk over the remaining maturity of the bond – taken to be 10 years, the maturity I target in the data. Since I do not observe default risk, I simply compare the variance of the corporate bond spread with the variance of future spread growth, discounted at factor ρ, and take the residual to be attributable to cashflow news:

$$\text{Var}(\text{spread}_t) = \text{Var}(\sum_{s=1}^{s=9} \rho^s_{bond} \Delta\text{spread}_{t+s}) + \text{Cashflow news} \tag{2.A.9}$$

An alternative way of assessing the importance of expected returns is simply to regress the ten-year ahead credit spread growth on the current credit spread – an exercise that, similarly to the long-horizon regressions in Table 2.A.4, show that almost all of the variation in credit spreads can be accounted for by expected returns over the lifetime of the bond. Due to the lack of corporate default data outside of the US, and the fact that most spread variation reflects future expected returns, I simply use the change in the bond spread as a proxy for the discount rate news on this asset class.

Table 2.A.7 shows the estimated VAR for housing and equity, using GMM specified in equations (2.A.5) and (2.A.6), and the corresponding variance decomposition rations in equation (2.A.7). Dependent variables $z_{i,t+1}$ are in columns, and predictor variables $z_{i,t}$ are in rows. As in Table 2.5.1, the first two columns present the predictive coefficients on next period's real equity returns $r_{i,t+1}$ and real dividend growth $dg_{i,t+1}$. The bottom row uses the dividend-to-price ratio $dp_{i,t}$ as a predictor, so the estimates in the first two columns row 3 correspond directly to the OLS estimates in the first two columns row 1 in Table 2.5.1.

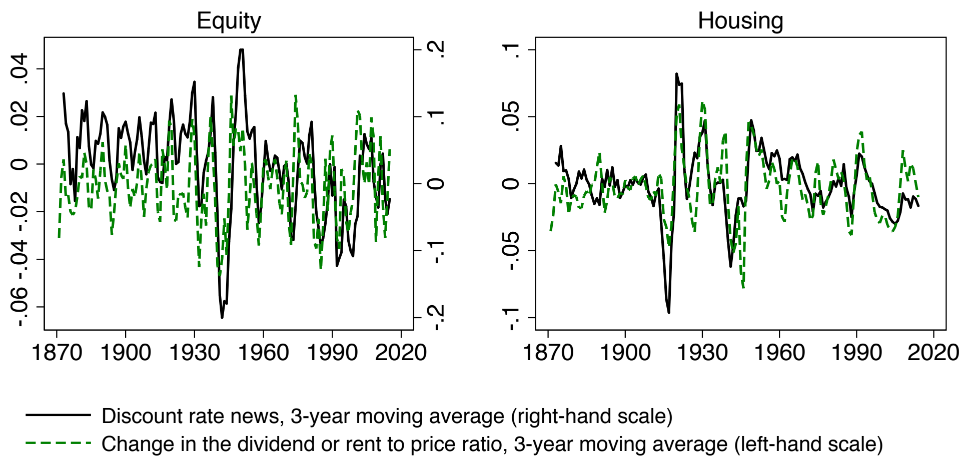
The coefficients on returns and dividend growth in row 3 are more or less unchanged when compared to the OLS estimation in Table 2.5.1. Compared to OLS, the VAR estimates a richer dynamic structure, also allowing current returns and

Table 2.A.7. Return predictability in a VAR

	(1)	(2)	(3)	(4)	(5)	(6)
		Equity			Housing	
	r_{t+1}	dg_{t+1}	dp_{t+1}	r_{t+1}	dg_{t+1}	dp_{t+1}
Estimated coefficients:						
r_t	0.094***	0.034	-0.063*	0.217***	-0.009	-0.236***
	(0.034)	(0.044)	(0.036)	(0.047)	(0.033)	(0.047)
dg_t	0.039	-0.049	-0.091**	0.155***	0.446***	0.306***
	(0.027)	(0.048)	(0.044)	(0.042)	(0.049)	(0.045)
dp_t	0.058***	-0.133***	0.840***	0.062***	-0.029***	0.954***
	(0.012)	(0.020)	(0.019)	(0.007)	(0.008)	(0.008)
Variance decomposition of dp_t:						
DR share			32			58
CF share			68			42
Observations			2177			1793

Note: VAR subject to present value moment constraints. Estimated using GMM, accounting for cross-sectional and time dependence in standard errors. Variables are log real total return r, log real dividend or rent growth dg, and log of dividend-price or rent-price ratio dp. DR share is the proportion of variation in dp_t that is due to discount rate news. CF share is the proportion of variation in dp_t that is due to expected cashflow movements. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

Figure 2.A.2. Discount rate news and raw valuation measures



Note: Discount rate news are the change in the present value of expected future returns in any given year.

dividend growth $r_{i,t}$ and $dg_{i,t}$ to predict future dividend and rental growth. The corresponding coefficients are in the top two rows of the table. Unlike the results in the literature on the US, current returns actually help predict future returns with a positive sign, even conditional on current valuation ratios. This is suggestive of the existence of aggregate momentum in the housing and equity markets. The coefficients are small – on average only one-tenth of high equity returns, and one-fifth of high housing returns is sustained into next period, but they do paint a more nuanced picture of the expected return variation. Even though high returns this year mean that the return is likely to be somewhat above mean next year, a prolonged period of high returns raises valuations *dp*, and eventually is followed by low returns, and mean reversion, which persists for several years. The estimates in the bottom row also show that valuation ratios are very persistent, with autocorrelation coefficients of 0.84 for equities and 0.95 for housing. The variance decomposition shares show that most variance in equity valuations comes about from future cashflows, and in housing valuations – from expected returns.

The annual time series for discount rate news on equity and housing, estimated using equation (2.A.8) are shown in Figure 2.A.2, alongside changes in the "raw" valuation metric *dp*. Consistent with Figure 2.1.1, both series are plotted as three-year moving averages. Changes in asset valuations *dp* closely follow those in discount rate news, and hence serve as useful proxies for variation in discount rates or systematic expectation errors.

2.A.3.4 The importance of discount rate news through time

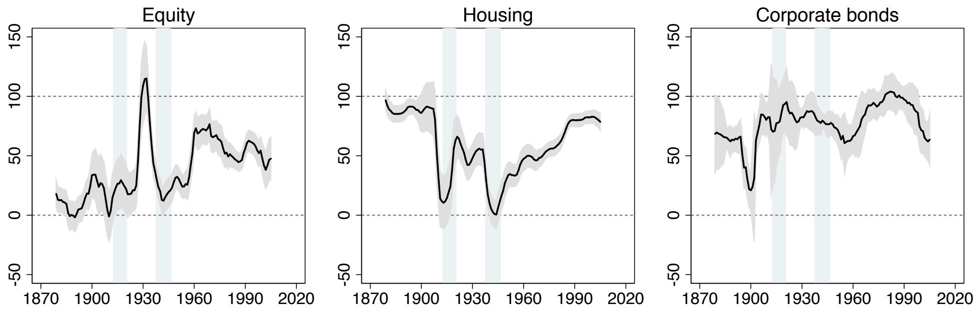
Figure 2.A.3 assesses how the excess volatility for each asset class has varied over time. It plots the relative importance of discount rate news over centered 20-year rolling time windows. For each time window, the discount rate news share is computed as the ratio of the predictive coefficient β in the return regression (2.2.4) to the sum of β and the coefficient γ from the cashflow growth regression (2.2.5):

$$\text{DR share}_{i,t} = \beta_{i,2,t-9,t+10}/(\beta_{i,2,t-9,t+10} + \gamma_{i,2,t-9,t+10}) \tag{2.A.10}$$

For corporate bonds, the discount rate share is computed by comparing the variance of 10-year ahead spread growth (discount rates), and the spread (bond valuations), within the 20-year window. As in Table 2.5.2, the value of 100 indicates that all the variance is accounted for by discount rate news, and none – by fundamentals.

Starting with the left-hand panel of Figure 2.A.3, the time trend suggests that if anything, equity markets have become more excessively volatile over time. This is consistent with the fact that the underlying equity returns have become somewhat more volatile, while cashflows have not (Figure 2.4.2). The Figure also highlights the Great Depression of the 1930s as the period when discount rates, rather than fundamentals were particularly important. This suggests that risk tolerance and risk perception played an important role in the financial crises of the 1930s –

Figure 2.A.3. The importance of time varying discount rates through time



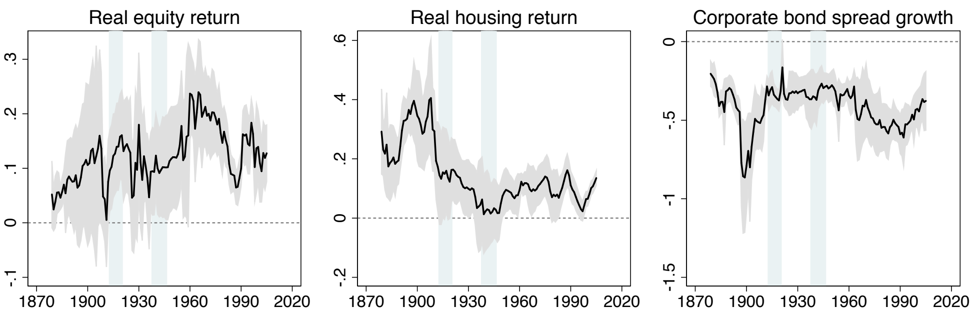
Note: The share of variation in valuation ratios *dp* that is accounted for by discount rate news, for each asset class, centered 20-year moving windows. Share equal to 100 means all of the variation in valuations is accounted for by discount rate news. Calculated using rolling 20-year window predictability regressions of 10-year ahead discounted returns. For housing and equity, the share equals the ratio of the predictive coefficient on the return to the sum of the return and dividend growth coefficients, DR share$_{i,t} = \beta_{i,2,t-9,t+10}/(\beta_{i,2,t-9,t+10} + \gamma_{i,2,t-9,t+10})$. For bonds, it is the proportion of spread variance that is explained by the co-variation with future spread growth, within the 20-year rolling window. Horizontal shaded grey areas are 90% confidence intervals. Vertical shaded blue areas are the two world wars.

a fact highlighted by a number of contemporaries, including Keynes (1936). Turning to housing, the world wars stand out as the two large shocks to fundamentals. During these periods, almost all of the housing market volatility was attributable to cashflows. Excess volatility in housing was high in the late 19th century, low in the mid-20th century, and steadily increased from the low post-war level in the six decades following 1950. Throughout the sample, the variation in corporate bond spreads is largely attributable to future returns.

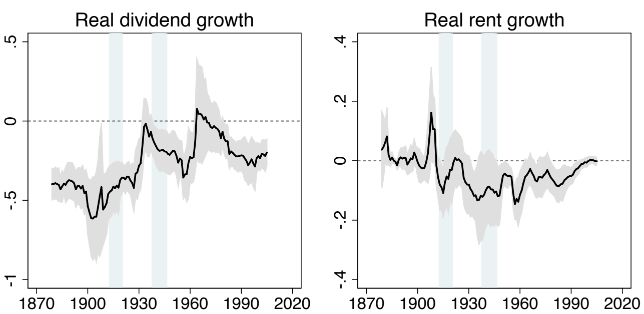
Figure 2.A.4 also separates out the importance of discount rate news relative to cashflows into standalone discount rate and cashflow news importance. Put differently, it plots the rolling-window β (Figure 2.A.4a) and γ (Figure 2.A.4b) coefficients over time. The predictability for equity returns is strongest over the recent time period, while that of housing returns – during the late 19th century. Dividend growth predictability largely stems from the period before the 1930s, which helps explain the lack of dividend growth predictability found in the post-1925 US data (Cochrane, 2008).

Figure 2.A.4. Return and cashflow predictability through time

(a) Predictability of real returns:



(b) Predictability of real dividend and rent growth



Note: Return and cashflow predictability regression coefficients, calculated using regressions over 20-year rolling windows. Panel (a) shows the $t+1$ return coefficient $\beta_{i,2,t-9,t+10}$ in the predictive regression (2.2.4), calculated over the window $t-9$ to $t+10$ years, for each year t. Estimates use OLS with fixed effects and country-clustered standard errors. Horizontal shaded grey areas are 90% confidence intervals. Vertical shaded blue areas are the two world wars.

2.A.4 Cross-asset predictability: additional results

Table 2.A.8. Predictability of cashflows across asset classes

	(1)	(2)	(3)	(4)
	Unconditional		Conditional	
	Equity dg_{t+1}	Housing dg_{t+1}	Equity dg_{t+1}	Housing dg_{t+1}
Dividend-price ratio		0.004	-0.090***	0.007
		(0.008)	(0.033)	(0.007)
Rent-price ratio	-0.054**		-0.028	-0.025*
	(0.027)		(0.038)	(0.013)
Bond spread	-0.011	0.006	-0.008	0.007
	(0.009)	(0.004)	(0.008)	(0.004)
R^2	0.006	0.005	0.032	0.013
Observations	1225	1220	1225	1220

Note: OLS regressions of cashflows on one asset class on discount rates of other asset classes. All specifications use country fixed effects. Predictor (y) variables in rows. Dividend-price and rent-price ratios are in logs. Corporate bond spread is in percentage points dg is log real dividend or rental growth. Country-clustered standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

Table 2.A.9. Predictability using macro-financial factors: alternative specifications

	(1)	(2)	(3)
	Equity r_{t+1}	Housing r_{t+1}	Corporate bond $\Delta spread_{t+1}$
		Panel 1: Conditional predictability	
Dividend-price ratio	0.042***		
	(0.014)		
Rent-price ratio		0.072***	
		(0.011)	
Bond spread			-0.301***
			(0.063)
$\Delta_3 Real\ Consumption_t$	-0.224	0.233***	-0.114
	(0.144)	(0.073)	(0.413)
$Surplus\ Consumption_t$	-0.006	-0.107	-0.382
	(0.151)	(0.081)	(0.674)
cay_t	0.009	0.014	0.383
	(0.043)	(0.015)	(0.261)
$\Delta_3 Bank\ Leverage_t$	-0.073*	-0.018	-0.091
	(0.038)	(0.015)	(0.154)
$\Delta_3 Real\ Bank\ Assets_t$	-0.018	0.006	0.422
	(0.056)	(0.024)	(0.268)
$\Delta_3 Real\ Credit_t$	-0.009	0.062***	0.170
	(0.041)	(0.019)	(0.196)
$log(MCAP_t/GDP_t)$	-0.034***	0.003	-0.003
	(0.008)	(0.003)	(0.034)
$Term\ Spread_t$	0.014**	0.006***	-0.013
	(0.006)	(0.001)	(0.031)
R^2	0.044	0.123	0.135
Observations	1430	1385	1103
		Panel 2: 5-year ahead average returns	
$\Delta_3 Real\ Consumption_t$	-0.110**	0.012	-0.055
	(0.055)	(0.059)	(0.206)
$Surplus\ Consumption_t$	-0.027	0.082	-0.176
	(0.087)	(0.061)	(0.213)
cay_t	-0.007	0.020	0.017
	(0.039)	(0.023)	(0.061)
$\Delta_3 Bank\ Leverage_t$	-0.096***	-0.020*	0.107
	(0.020)	(0.011)	(0.093)
$\Delta_3 Real\ Bank\ Assets_t$	-0.073**	-0.030	0.050
	(0.032)	(0.028)	(0.064)
$\Delta_3 Real\ Credit_t$	0.045	0.023	0.028
	(0.030)	(0.025)	(0.090)

Table 2.A.9. Predictability using macro-financial factors: alternative specifications, continued

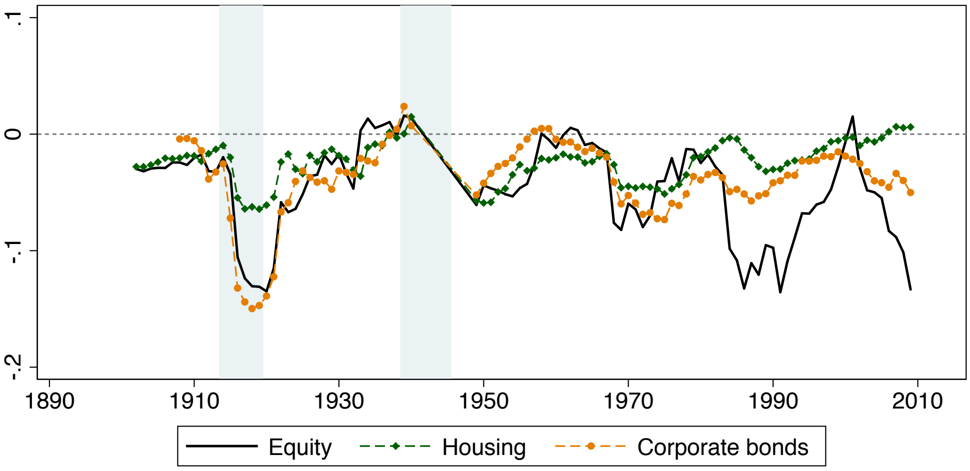
	(1)	(2)	(3)
	Equity r_{t+1}	Housing r_{t+1}	Corporate bond $\Delta spread_{t+1}$
$log(MCAP_t/GDP_t)$	-0.044***	-0.002	-0.021
	(0.007)	(0.004)	(0.037)
$Term\ Spread_t$	0.004	0.005**	0.025***
	(0.003)	(0.002)	(0.010)
R^2	0.141	0.048	0.044
Observations	1439	1357	1055
Panel 3: Post 1950			
$\Delta_3 Real\ Consumption_t$	-0.564*	0.482***	1.308
	(0.307)	(0.130)	(1.047)
$Surplus\ Consumption_t$	-0.050	-0.036	-1.414
	(0.352)	(0.137)	(1.068)
cay_t	-0.008	-0.004	0.231
	(0.062)	(0.023)	(0.337)
$\Delta_3 Bank\ Leverage_t$	0.012	-0.033*	0.051
	(0.048)	(0.020)	(0.111)
$\Delta_3 Real\ Bank\ Assets_t$	-0.063	-0.017	0.373
	(0.073)	(0.027)	(0.364)
$\Delta_3 Real\ Credit_t$	-0.051	-0.001	0.196
	(0.076)	(0.033)	(0.263)
$log(MCAP_t/GDP_t)$	-0.045***	0.007	-0.033
	(0.014)	(0.005)	(0.038)
$Term\ Spread_t$	0.009	0.006***	0.039
	(0.009)	(0.001)	(0.041)
R^2	0.060	0.111	0.015
Observations	931	911	755

Note: OLS regressions with country fixed effects. Predictor (x) variables in rows. $\Delta_i Real\ Consumption_t$ is the log change in real consumption per capita from $t-i$ to t. $Surplus\ Consumption_t$ is the real consumption per capita at t relative to a backward-looking 10-year moving average trend, from t-10 to t. cay_t is a proxy for the consumption-wealth ratio, estimated as the deviations from the cointegrating relationship between real consumption, a proxy for financial wealth (the combined capitalization of the equity, housing and government bond markets) and real wages. $\Delta_i Bank\ Leverage_t$ is the change in the log of bank leverage from $t-i$ to t. $\Delta_i Real\ Bank\ Assets_t$ is the change in the log of real bank assets from $t-i$ to t. $\Delta_i Real\ Credit_t$ is the change in the log of real credit to non-financials from $t-i$ to t. $log(MCAP_t/GDP_t)$ is the log of the market capitalization to GDP ratio. $Term\ Spread_t$ is the percentage point yield spread between the long and short term government debt. Dependent (y) variables in columns. r is the log real total return, and Δ spread is the percentage point change in the bond spread. Country clustered standard errors in parentheses. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

2.A.5 Explanations for the puzzle: additional results

2.A.5.1 Asset-specific risk

Figure 2.A.5. Time variation in asset-specific risk: alternative discount rate measure

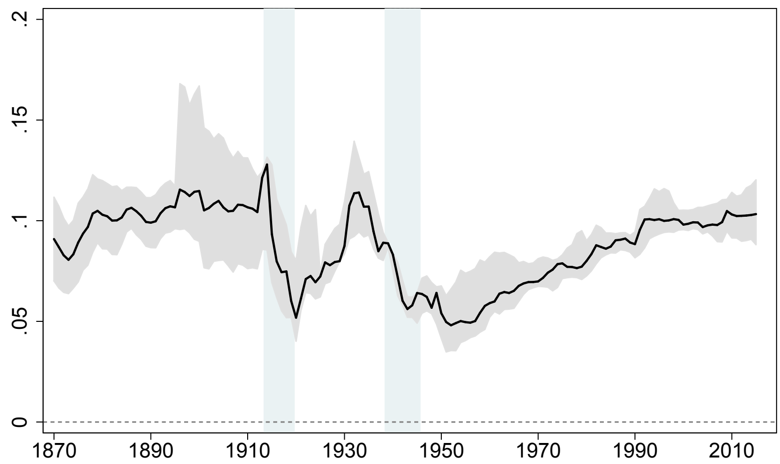


Note: Covariance between the return on equity, housing and corporate bonds, and a proxy for the aggregate discount factor, constructed as the average of asset-specific sentiment on each asset class, normalised to mean 0 and standard deviation of 1 over the full sample. Centered 10-year rolling windows.

Section 2.7 compared the time variation in riskiness of different asset classes using the covariance with a simple proxy of the discount factor m – the average of standardised asset-specific valuations. Figure 2.A.5 plots the same m, R covariance using a more sophisticated proxy for m which utilises the information on other macro-financial predictors such as the surplus consumption ratio and bank leverage. I construct this proxy as the average of asset-specific sentiment – the inverse of the expected return on each asset class, with expected returns calculated using a regression on the own valuation ratio and a broad range of the macro-financial risk factors (see Section 2.7.3 for details on the estimation). Similarly to Figure 2.7.1, this measures suggests that each of the three asset classes is risky, and that this riskiness varies little over time or across asset classes.

2.A.5.2 Non-monetary payoffs

Figure 2.A.6. The share of rental expenditure in GDP



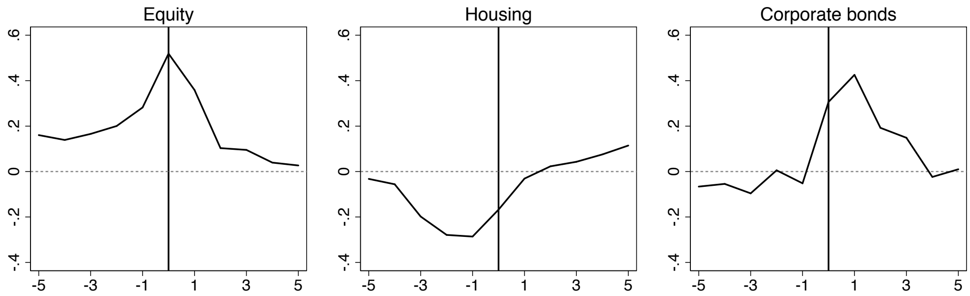
Note: Ratio of total rental expenditure to GDP, which acts a proxy for the marginal utility of housing services. Data for Australia, France, Germany, Sweden, UK and USA. Sweden and USA data start in the 1930, Australia and France around 1900, Germany and UK in 1870.

Figure 2.A.6 provides further evidence on whether the utility of housing services is likely to vary much from year to year. It shows how the rental income to GDP ratio has evolved over the long run in 6 of the countries in my sample. The solid black line shows an unweighted average, while the shaded areas mark the interquartile range. The rent to GDP ratio is a proxy for the housing share in consumption, which in turn is a measure of the marginal utility of housing services. This ratio is relatively stable over time and displays little movement at business cycle frequency. The only sizeable variation occurs around the two world wars – the periods when asset-specific expected returns do actually show some positive co-movement (Figure 2.6.1).

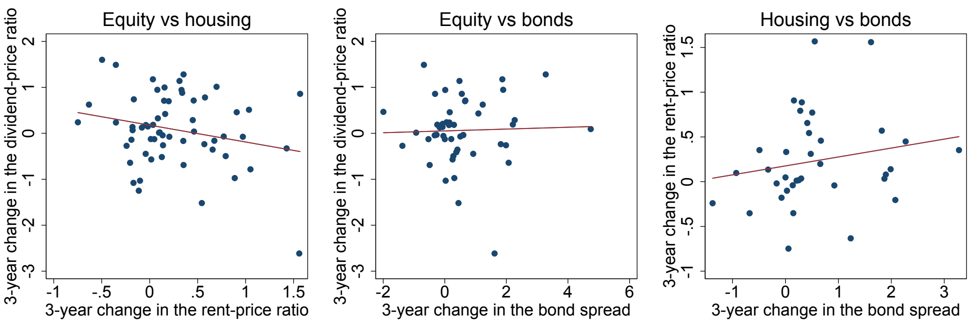
2.A.6 Real effects of time-varying sentiment: additional details

Figure 2.A.7. Risky asset valuations around financial crises

(a) Discount rate proxies during the crisis



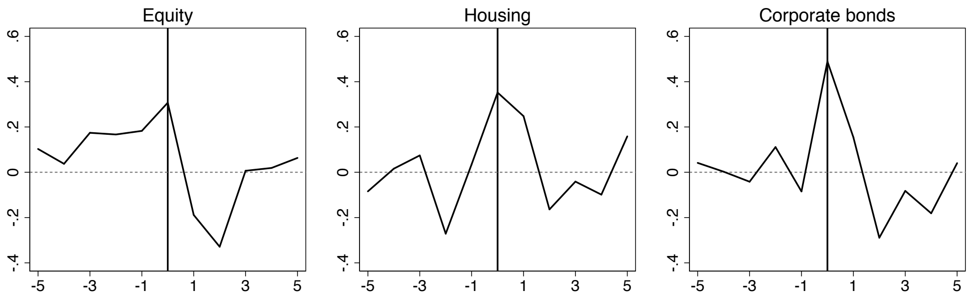
(b) Co-movement of asset-specific discount rate proxies around the crisis



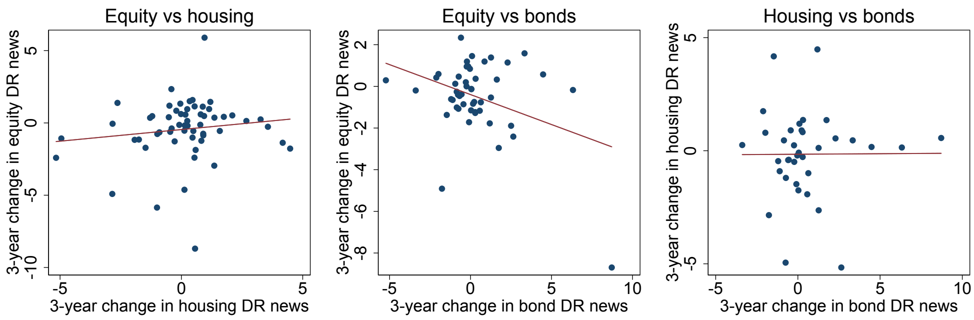
Note: Financial crisis at $t = 0$. All data are standardised to mean 0 and standard deviation of 1 in the full sample. Correlations are for the 3-year changes around the crisis year (from $t-1$ to $t+1$). Each point on the scatter plot is an individual financial crisis event.

Figure 2.A.8. Discount rate news around financial crises

(a) Discount rate news during the crisis



(b) Co-movement of asset-specific discount rate news around the crisis



Note: Financial crisis at $t = 0$. All data are standardised to mean 0 and standard deviation of 1 in the full sample. Correlations are for the 3-year averages around the crisis year (from $t-1$ to $t+1$). Each point on the scatter plot is an individual financial crisis event. Discount rate news correspond to changes in the present in future expected returns.

References

Adam, Klaus, Johannes Beutel, and Albert Marcet. 2017. "Stock Price Booms and Expected Capital Gains." *American Economic Review* 107 (8): 2352–2408.

Adam, Klaus, and Sebastian Merkel. 2018. "Stock Prices Cycles and Business Cycles." unpublished.

Adrian, Tobias, and Hyun Song Shin. 2010. "Liquidity and Leverage." *Journal of Financial Intermediation* 19 (3): 418–37.

Bai, Jushan, and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics* 18 (1): 1–22.

Baker, Malcolm, and Jeffrey Wurgler. 2000. "The Equity Share in New Issues and Aggregate Stock Returns." *Journal of Finance* 55 (5): 2219–2257.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59 (4): 1481–1509.

Bao, Jack, Jun Pan, and Jiang Wang. 2011. "The Illiquidity of Corporate Bonds." *Journal of Finance* 66 (3): 911–946.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. 2015. "X-CAPM: An Extrapolative Capital Asset Pricing Model." *Journal of Financial Economics* 115 (1): 1–24.

Barberis, Nicholas, and Andrei Shleifer. 2003. "Style Investing." *Journal of Financial Economics* 68 (2): 161–199.

Baron, Matthew, and Tyler Muir. 2018. "Intermediaries and Asset Prices: Evidence from the US, UK, and Japan, 1870-2016." Working Paper.

Barro, Robert J. 2006. "Rare Disasters and Asset Markets in the Twentieth Century." *Quarterly Journal of Economics* 121 (3): 823–866.

Barro, Robert J., and Jose F. Ursua. 2008. "Consumption Disasters in the Twentieth Century." *American Economic Review* 98 (2): 58–63.

Benhabib, Jess, and Alberto Bisin. 2016. "Skewed Wealth Distributions: Theory and Empirics." NBER Working Paper 21924.

Biais, Bruno, and Richard C. Green. 2007. "The Microstructure of the Bond Market in the 20th Century." Working paper.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2017. "Diagnostic Expectations and Stock Returns." NBER Working Paper 23863. National Bureau of Economic Research.

Brunnermeier, Markus K., and Lasse Heje Pedersen. 2009. "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22 (6): 2201–2238.

Bry, Gerhard, and Charlotte Boschan. 1971. "Programmed Selection of Cyclical Turning Points." In *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. NBER, 7–63.

Campbell, John. 1991. "A Variance Decomposition for Stock Returns." *Economic Journal* 101 (405): 157–79.

Campbell, John Y., and John Ammer. 1993. "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns." *Journal of Finance* 48 (1): 3–37.

Campbell, John Y., and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* 107 (2): 205–251.

Campbell, John Y., and Robert J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1 (3): 195–228.

Cochrane, John H. 2017. "Macro-Finance." *Review of Finance* 21 (3): 945–985.

Cochrane, John H. 2008. "The Dog That Did Not Bark: A Defense of Return Predictability." *Review of Financial Studies* 21 (4): 1533–1575.

Cochrane, John H. 2011. “Presidential Address: Discount Rates.” *Journal of Finance* 66 (4): 1047–1108.

Cochrane, John H., and Monika Piazzesi. 2005. “Bond Risk Premia.” *American Economic Review* 95 (1): 138–160.

Constantinides, George M., and Darrell Duffie. 1996. “Asset Pricing with Heterogeneous Consumers.” *Journal of Political Economy* 104 (2): 219–240.

Eggertsson, Gauti B., and Paul Krugman. 2012. “Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach.” *Quarterly Journal of Economics* 127 (3): 1469–1513.

Engsted, Tom, and Thomas Q Pedersen. 2010. “The Dividend-Price Ratio Does Predict Dividend Growth: International Evidence.” *Journal of Empirical Finance* 17 (4): 585–605.

Fama, Eugene F., and Kenneth R. French. 1989. “Business Conditions and Expected Returns on Stocks and Bonds.” *Journal of Financial Economics* 25 (1): 23–49.

Fleckenstein, Matthias, Francis A. Longstaff, and Hanno Lustig. 2010. “Why Does the Treasury Issue Tips? The Tips-Treasury Bond Puzzle.” NBER Working Paper 16358.

Gabaix, Xavier. 2012. “Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance.” *Quarterly Journal of Economics* 127 (2): 645–700.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty. 2017. “Accounting for Wealth Inequality Dynamics: Methods, Estimates and Simulations for France (1800-2014).” BANK OF FRANCE WORKING PAPER No. WP 633.

Gârleanu, Nicolae, and Stavros Panageas. 2015. “Young, Old, Conservative, and Bold: The Implications of Heterogeneity and Finite Lives for Asset Pricing.” *Journal of Political Economy* 123 (3): 670–685.

Giesecke, Kay, Francis A. Longstaff, Stephen Schaefer, and Ilya Strebulaev. 2011. “Corporate Bond Default Risk: A 150-year Perspective.” *Journal of Financial Economics* 102 (2): 233–250.

Giesecke, Kay, Francis A. Longstaff, Stephen Schaefer, and Ilya A. Strebulaev. 2014. “Macroeconomic Effects of Corporate Default Crisis: A Long-term Perspective.” *Journal of Financial Economics* 111 (2): 297–310.

Giglio, Stefano, and Bryan Kelly. 2018. “Excess Volatility: Beyond Discount Rates.” *Quarterly Journal of Economics* 133 (1): 71–127.

Gilchrist, Simon, and Egon Zakrajšek. 2012. “Credit Spreads and Business Cycle Fluctuations.” *American Economic Review* 102 (4): 1692–1720.

Golez, Benjamin, and Peter Koudijs. 2018. “Four Centuries of Return Predictability.” *Journal of Financial Economics* 127 (2): 248–263.

Gompers, Paul A., and Andrew Metrick. 2001. “Institutional Investors and Equity Prices.” *Quarterly Journal of Economics* 116 (1): 229–259.

Goyal, Amit, and Ivo Welch. 2008. “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction.” *Review of Financial Studies* 21 (4): 1455–1508.

Greenwood, Robin, and Samuel G. Hanson. 2013. “Issuer Quality and Corporate Bond Returns.” *Review of Financial Studies* 26 (6): 1483–1525.

Gyourko, Joseph, and Donald B. Keim. 1992. “What Does the Stock Market Tell Us About Real Estate Returns?” *Real Estate Economics* 20 (3): 457–485.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh. 2017. “Predicting Relative Returns.” NBER Working Paper 23886.

Haddad, Valentin, and Tyler Muir. 2018. “Do Intermediaries Matter for Aggregate Asset Prices?” Working paper. Working Paper.

He, Zhiguo, Bryan Kelly, and Asaf Manela. 2017. “Intermediary Asset Pricing: New Evidence from Many Asset Classes.” *Journal of Financial Economics* 126 (1): 1–35.

He, Zhiguo, and Arvind Krishnamurthy. 2013. "Intermediary Asset Pricing." *American Economic Review* 103 (2): 732–70.

Hu, Grace Xing, Jun Pan, and Jiang Wang. 2013. "Noise as Information for Illiquidity." *Journal of Finance* 68 (6): 2341–2382.

Ibbotson, Roger G., and Laurence B. Siegel. 1984. "Real Estate Returns: A Comparison with Other Investments." *Real Estate Economics* 12 (3): 219–242.

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1): 161–182.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2017. "The Rate of Return on Everything, 1870–2015." NBER Working Paper 24112.

Jordà, Òscar, Björn Richter, Moritz Schularick, and Alan M. Taylor. 2017. "Bank Capital Redux: Solvency, Liquidity, and Crisis." NBER Working Paper 23287.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45 (2): 3–28.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2016. "Macrofinancial History and the New Business Cycle Facts." In *NBER Macroeconomics Annual 2016, Volume 31*. Edited by Jonathan A. Parker Martin Eichenbaum. Chicago, Ill.: University of Chicago Press, 213–263.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money*. London: Macmillan.

Klovland, Jan Tore. 2004. "Bond Markets and Bond Yields in Norway 1820–2003." In *Historical Monetary Statistics for Norway 1819–2003. Norges Bank Occasional Paper No. 35*. Edited by Jan T. Klovland Øyvind Eitrheim and Jan F. Qvigstad. Chapter 4, pp. 99–181.

Knoll, Katharina. 2017. "Our Home in Days Gone By: Housing Markets in Advanced Economies in Historical Perspective." Doctoral dissertation. Free University of Berlin. Chapter 3. As Volatile As Houses: Return Predictability in International Housing Markets, 1870–2015.

Knoll, Katharina, Moritz Schularick, and Thomas M. Steger. 2017. "No Price like Home: Global House Prices, 1870–2012." *American Economic Review* 107 (2): 331–352.

Krishnamurthy, Arvind, and Tyler Muir. 2017. "How Credit Cycles Across a Financial Crisis." NBER Working Paper 23850.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2017. "Income and Wealth Inequality in America , 1949–2013." CEPR Discussion Paper 20547.

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2018. "The Big Bang: Stock Market Capitalization in the Long Run." EHES Working Paper 136.

Lettau, Martin, and Sydney Ludvigson. 2002. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance* 56 (3): 815–849.

Lettau, Martin, Matteo Maggiori, and Michael Weber. 2014. "Conditional Risk Premia in Currency Markets and Other Asset Classes." *Journal of Financial Economics* 114 (2): 197–225.

Lettau, Martin, and Stijn Van Nieuwerburgh. 2008. "Reconciling the Return Predictability Evidence." *Review of Financial Studies* 21 (4): 1607–1652.

Liu, Crocker H., and Jianping J.P. Mei. 2003. "The Predictability of Returns on Equity REITs and their Co-movement with Other Assets." In. *Asset Pricing*, 21–45.

López-Salido, David, Jeremy C. Stein, and Egon Zakrajšek. 2017. "Credit-Market Sentiment and the Business Cycle." *Quarterly Journal of Economics* 132 (3): 1373–1426.

Martıúnez-Toledano, Clara. 2018. "Housing Bubbles and Wealth Inequality: Evidence from Spain." Unpublished.

Mediobanca. Various years. "Indici e Dati Relativi ad Investimenti in Titoli Quotati."

Mian, Atif, and Amir Sufi. 2014. "What Explains the 2007–2009 Drop in Employment?" *Econometrica* 82 (6): 2197–2223.

Minsky, Hyman P. 1977. "The Financial Instability Hypothesis: An Interpretation of Keynes and an ALternative to "Standard" Theory." *Nebraska Journal of Economics and Business* 16 (1): 5–16.

Muir, Tyler. 2017. "Financial Crises and Risk Premia." *Quarterly Journal of Economics* 132 (2): 765–809.

Nozawa, Yoshio. 2017. "What Drives the Cross-Section of Credit Spreads?: A Variance Decomposition Approach." *Journal of Finance* 72 (5): 2045–2072.

Piazzesi, Monika, Martin Schneider, and Selale Tuzel. 2007. "Housing, Consumption and Asset Pricing." *Journal of Financial Economics* 83 (3): 531–569.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Harvard University Press.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics* 133 (2): 553–609.

Piketty, Thomas, and Gabriel Zucman. 2014. "Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010." *Quarterly Journal of Economics* 129 (3): 1255–1310.

Ross, Stephen A. 1977. "Return, Risk and Arbitrage." In *Risk and Return in Finance*. Edited by Irwin Friend and James L. Bicksler. Cambridge, MA.: Ballinger.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review* 71 (3): 421–436.

Shiller, Robert J. 1982. "Consumption, Asset Markets and Macroeconomic Fluctuations." *Carnegie-Rochester Conference Series on Public Policy* 17: 203–238.

Shiller, Robert J. 2000. *Irrational Exuberance*. Princeton, N.J.: Princeton University Press.

Stambaugh, Robert F. 1999. "Predictive regressions." *Journal of Financial Economics* 54 (3): 375–421.

Chapter 3

The Big Bang: Stock Market Capitalization in the Long Run*

Joint with Kaspar Zimmermann

3.1 Introduction

How has the stock market evolved since the late 19th century, and which forces drive its long-run evolution? The consensus in existing literature centers around the "great reversals" hypothesis of Rajan and Zingales (2003): that stock markets were large at the turn of the 20th century, stagnated between 1913 and 1980, and have regained their importance today. The explanations for this trend largely relate to changes in quantities of listed equity, with the ultimate drivers, in turn, relating to institutions and political economy considerations (La Porta, Lopez-de-Silanes, Shleifer, and Vishny, 1997; Rajan and Zingales, 2003). The data underpinning this consensus are, however, relatively scant. Existing long-run market capitalization estimates are limited to several selected benchmark years, with little information on what happens in-between, and comparability issues plaguing the comparisons across countries and time.

This paper introduces a comprehensive annual dataset on the long-run evolution of stock market capitalization in advanced economies. The data cover 17 countries over the period 1870–2016, and are constructed from a wide range of primary and secondary historical sources, with many of these previously unused or newly compiled using hand-collected archival data. Together with the extensive documentation in the Data Appendix, these series provide a new resource for researchers to

* This work is part of a larger project kindly supported by a research grant from the Bundesministerium für Bildung und Forschung (BMBF). We are grateful to Leslie Hannah, Kristian Rydqvist, Björn Richter, Moritz Schularick, and the participants of the Bonn Macrohistory Seminar, the 8th Eurhistock workshop and the EABH Institutional Investors conference for providing helpful comments. We are indebted to a large number of researchers who helped with individual country data.

study the development of the stock market, equity finance, and capital structure throughout the last 145 years. Our paper uses these data to make two further contributions. First, we sketch out the main trends and patterns in the market capitalization data, mapping out the long-run evolution of stock market size at a higher frequency and for a broader sample than was previously possible. Second, we combine the capitalization data with additional information on stock returns, fundamentals and discount rates from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019), to re-assess the main drivers of secular movements in stock market cap. Our analysis reveals a series of new facts which are materially at odds with the current consensus.

First, the evolution of stock market size over time resembles a hockey stick. The ratio of stock market capitalization to GDP was flat at around one-third between 1870 and 1985 before undergoing a rapid, sustained and historically unprecedented expansion in the 1980s and 1990s. This structural increase – the “big bang” – leaves today’s market capitalization to GDP ratios at around 1, three times the historical norm. Second, the time variation in stock market size is almost entirely driven by changes in prices, rather than quantities. The recent big bang is a sharp and persistent price appreciation, and historical swings of market cap typically track the developments in the stock price index. Third, much of the variation in stock market size is driven by cyclical and secular movements in the equity risk premium, with the recent big bang coinciding with a sharp decline in the dividend-price ratio. Furthermore, the market cap to GDP ratio predicts future equity returns, outperforming standard metrics such as the dividend-price ratio, and sharp run-ups in stock market cap share many characteristics with stock market bubbles. This suggests that rather than being associated with political norms and financial efficiency, stock market capitalization serves as a “Buffet indicator” of investor risk appetite.

The implications of our findings go far beyond the topics of financial structure and stock market development. The stock market constitutes an important component of national wealth, and one that is held by the relatively rich. Our analysis for listed equities can, therefore, be seen as a laboratory for understanding the broader trends in wealth-to-income ratios and wealth inequality, which have been subject to much recent debate (Piketty, 2014; Saez and Zucman, 2016; Alvaredo, Atkinson, and Morelli, 2018). Our findings suggest that these trends may be driven by changes in asset returns and valuations, rather than the pace of capital accummulation or labour income growth – a finding that also emerges from recent work by Bach, Calvet, and Sodini (2016) and Kuhn, Schularick, and Steins (2017). Indeed, the structural increase in market capitalization during the big bang marks a turning point towards an upward trend in national wealth and inequality. The importance of time varying risk premia in driving equity capitalization further suggests that insights from the asset pricing literature can help explain the broad macro-financial trends in the level and distribution of aggregate wealth. Finally, the fact that market capitalization predicts future equity returns shows that variation in wealth can have important implications for asset prices and risk premia, echoing Lettau and Ludvigson (2002)'s earlier findings on the predictive power of the consumption-wealth ratio.

We establish these and other related findings in four parts. First, we document the evolution of stock market capitalization across countries and time. The hockey-stick big bang pattern is a robust feature of the data: the recent expansion in stock market cap took place in every country in our sample, and in the vast majority of countries, the current level of stock market size far exceeds anything observed before the 1980s. At individual country level, market capitalization shows some correspondence to legal norms and initial capitalization levels. The long-run trends also underscore a rise in the global importance of the US equity market at the expense of those in the UK and France: while all three countries enjoyed a roughly equal share of the global market in the early 20th century, by the mid-20th century the US was clearly dominant, with its market accounting for close to 70% of the total 17-country capitalization.

The second part of our paper assesses whether variation in stock market cap is primarily driven by prices or quantities. We decompose changes in the market cap to GDP ratio into capital gains, GDP growth and net equity issuance using the methodology employed by Piketty and Zucman (2014) to decompose changes in wealth-to-income ratios into savings and price effects. We find that net equity issuance is sizeable – around 4% of market capitalization, or 1% of GDP – but relatively constant over time, and plays very little role in the short, medium and long-run swings in stock market cap. Instead, most of the cyclical and structural variation in capitalization is accounted for by stock price movements. We show that if stock price growth after 1985 had equalled its historical average, there would have been no big bang. On the contrary, if we set net issuance to its pre big bang average, the counterfactual evolution of stock market cap closely tracks actual observed data. The big bang, therefore, has little to do with changes in capital market entry, financial development, or physical accumulation of corporate equity. Rather, it is simply a sharp equity price increase that happened at a similar time across all advanced economies.

The third part of our paper looks into the factors which underpin the upsurge in equity prices during the big bang. Since firm valuations ultimately equal the discounted sum of future post-tax cashflows, the structural increase can be attributed to higher dividend payments, lower discount rates, or lower taxes. We explore the time trends in these three metrics and their correlation with the structural and cyclical movements in market cap in order to evaluate the likely contribution of each of these drivers.

We find that the structural increase in equity prices during the 1980s and 1990s is driven by a combination of higher dividends accruing to shareholders, and a higher valuation – or a lower discount rate – of the underlying dividend stream. Total dividend payments increased from 1% of GDP in 1985 to 2.5% of GDP in 2015. At the same time, the dividend-price ratio – a proxy for the rate at which these dividends are discounted – has fallen from a historically stable average of 4.5% to less than 3%. Together, these two developments can explain close to the entirety of the big bang.[1] Taxation, on the other hand, seems to be of only second-order importance. Corporate and income taxes did fall between the 1980s and today, but their levels have remained much higher than in the first half of the 20th century, a time when stock market capitalization was close to its historical average. These findings are confirmed by running cross-country explanatory regressions. Dividends to GDP and the dividend-price ratio are strongly correlated with market capitalization across different time horizons and historical periods, while changes in taxes are not.

What lies behind the recent trends in cashflows and discount rates? Some of the increase in cashflows is likely attributable to greater market power – and hence higher profitability – of larger firms, which also tend to be listed (De Loecker and Eeckhout, 2017). Discount rates can fall either because of a lower safe rate, or a lower equity risk premium. But despite a recent decline (Holston, Laubach, and Williams, 2017), safe rates are currently close to their long-run average levels, while market capitalization is not. This suggests that much of the recent decline in the discount rate is attributable to a structurally lower equity risk premium driven, for example, by lower macroeconomic risk (Lettau, Ludvigson, and Wachter, 2008; Bianchi, Lettau, and Ludvigson, 2016) or higher investor demand for advanced-economy risky assets (Bernanke, 2005).

In the final part of the paper, we show that the link between risk premia and market capitalization goes far beyond the long-run structural trends discussed above. The stock market cap to GDP ratio turns out to be a reliable measure of the time varying risk premium in the equity market. High market capitalization predicts low equity returns, and low – rather than high – cashflows. In fact, market cap substantially outperforms the standard price-dividend ratio variable as an equity return predictor. We show that this superior performance is driven by two factors. First, GDP is a better measure of the underlying corporate fundamentals than dividends, being, for example, relatively unaffected by changes in dividend policy. Second, capitalization is a better valuation metric, because it includes information on quantities as well as prices, which allows it to better capture the time variation in factors such as investor sentiment.[2] The data lend support to both of these claims. On the first, we show that the ratio of stock prices to GDP does a better job at predicting returns than the ratio of stock prices to dividends, suggesting that GDP is more informative about firm fundamentals than dividends. On the second, we show that net issuance predicts future returns, even when controlling for the current price-dividend ratio.

1. In the data, stock market capitalization increased by 80% of GDP during the big bang. A back-of-the-envelope calculation suggests that if dividends increase by 1.5–2% of GDP, and this increase is discounted at 2%–3% per year, valuations should increase by between 50% and 100% of GDP.

2. Baker and Wurgler (2000) show that if investors time the market to issue equity when pricing conditions are relatively favourable, periods of elevated investor sentiment should be accompanied by increases in both share prices and quantities.

Taking this analysis further reveals that rapid increases in stock market capitalization share many characteristics with stock market bubbles: they are accompanied by high returns and rising equity valuations, and followed by low equity returns and a higher risk of an equity market crash. These results connect nicely with parallel evidence on sectoral stock market run-ups and subsequent crashes in the United States (Greenwood, Shleifer, and You, 2018). Warren Buffet called stock market capitalization "the best single measure of where valuations stand at any given moment" (Buffett and Loomis, 2001). Taken together, our findings underscore the high relevance of the "Buffet indicator" for the return predictability literature.

Our work is related to three strands of existing literature. The first strand seeks to document the evolution of stock market size, with the earliest estimates dating back to occasional surveys of stock market activity commissioned by wealthy financiers and by stock exchanges (Burdett, 1882; Green, 1887). More recently, economic historians have provided more extensive estimates of market capitalization for individual countries (Hoffmann, 1965; Roe, 1971; Waldenström, 2014). When it comes to cross-country data, Goldsmith (1985) and Piketty and Zucman (2014) have constructed estimates of national wealth which include listed equities, and Rajan and Zingales (2003) – of the stock market capitalization itself, albeit with much of this analysis limited to selected benchmark years. Our paper is the first to compile long-run cross-country market capitalization data at annual frequency. The paper is complemented by an extensive Data Appendix which contains a detailed discussion of the various quality checks and comparisons with other existing capitalization estimates.

The second strand seeks to understand the drivers of stock market size and equity wealth. A number of papers, including Atje and Jovanovic (1993), Levine and Zervos (1996), La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1997) and Musacchio (2010), have equated stock capitalization with financial development and equity issuance, and sought to link market capitalization movements to legal norms and broader market-friendly regulations. Other authors have, however, instead focussed on changes in stock valuations, with McGrattan and Prescott (2005) attributing the recent increase in equity wealth of US corporations to lower corporate taxes, and De Loecker and Eeckhout (2017) emphasising the role of higher mark-ups. Our findings support this second, valuation-based view on stock market wealth. We further show that the importance of valuation changes goes far beyond the recent US data, and that a substantial part of these changes is attributable to time varying risk premia.

A third strand of literature has documented the importance of time variation in the risk premium on equities through return predictability regressions. A number of studies have shown that US equity returns can be predicted by the dividend-price ratio, and that this expected return or discount rate variation explains most of the changes in equity values (Campbell and Shiller, 1988; Cochrane, 2008). A number of alternative return predictors have also been explored, with Lettau and Ludvigson

(2002) documenting the predictive power of the consumption-wealth ratio, Rangvid (2006) – of the stock price to GDP ratio, Cooper and Priestley (2008) – of the output gap, and Baker and Wurgler (2000) – of equity issuance. We show that market cap to GDP captures some of the additional informational content of these alternative predictors, and is therefore able to outperform the price-dividend ratio. The fact that much of the superior power comes from using GDP instead of dividends, and quantities as well as prices, points towards importance of factors such as dividend smoothing (Chen, Da, and Priestley, 2012) and investor sentiment (Baker and Wurgler, 2000) for the empirical predictability relationships.

Stock markets today are larger than at any point in recent history. This, however, does not mean that financial markets are substantially more developed. Rather, this means that stock valuations are unusually high, and have been so for the best part of the last three decades. These high valuations could harbour positive news about high future corporate profitability or low levels of risk. But our analysis suggests that the rise of the stock market entails a darker side. Much of the increase in valuations is driven by a low equity risk premium – a factor that tends to fluctuate substantially over time and can quickly mean-revert, inducing large swings in stock prices, capitalization and household wealth. Indeed, the structural increase in market capitalization during the big bang has been accompanied by higher volatility, with several large surges in market cap followed by reversals to the structurally higher post-1980 mean.

3.2 A new dataset on historical stock market capitalization

This paper introduces a new dataset on the historical size of stock markets in advanced economies. The data consist of statistics on total stock market capitalization, on an annual basis, in 17 countries, from 1870 to today. The countries included are Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. Our data measure the total market value of all ordinary shares of domestic companies listed on the domestic exchanges at the end of each calendar year.

We use a wide range of primary and secondary sources to construct the data series, many of these new and previously unused. The secondary sources consist of financial history books and research articles, and publications of stock exchanges, statistical agencies, central banks and trade bodies. Where reliable secondary sources were not available, we construct the capitalization measure by aggregating the total market values of individual stocks, using data on stock prices and number of shares or listed capital value from stock exchange bulletins and gazettes, stock exchange handbooks and companies' published accounts. Most of these primary source data were newly compiled through a series of archival visits to the respective countries' stock exchanges, central banks and national libraries, while some were also helpfully shared with us by other researchers. We generally produce annual estimates of capitalization, but for instances where these were not available, we obtain capitalization data for benchmark years and construct the annual series using changes in the book capital of listed companies and share prices. An extensive Data Appendix, Tables D.1–D.17 and Figures D.1–D.17 detail the sources used for each country, and compare our estimates to others in the existing literature. For the decomposition and predictability analysis in Sections 3.4–3.6, we complement these data with statistics on equity prices, returns, and dividends from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) and Kuvshinov (2018), and data on taxes and corporate profits, much of it sourced from the work of Piketty and Zucman (2014).

The main challenge in constructing stock market capitalization indices is getting appropriate coverage of all ordinary shares listed on domestic stock exchanges, that are issued by domestic firms. This means that, first of all, the series should only include ordinary shares and exclude preferred shares and other securities listed on the stock exchange, such as preference shares and bonds (Hannah, 2018, offers a discussion of these issues in the early London Stock Exchange data). Some of the earlier statistical estimates bundle these different securities together, or sometimes only provide figures for both unlisted and listed equity liabilities. We therefore ensure that our estimates capture ordinary shares only, by where necessary constructing our own benchmark year estimates, or using supplementary stock exchange data and research publications to make this distinction.

The second challenge is that the capitalization measure should sum the securities listed on all domestic stock exchanges, net of any cross listings. Wherever possible, we therefore rely on data that cover all the major stock exchanges in the country, constructing our own estimates from microdata when necessary, as in the case of the pre World War 1 German stock market cap (see Appendix Table D.7). It is, however, not always possible to obtain information on the capitalization of smaller stock exchanges, especially one that goes beyond benchmark years. For most countries in our sample, the bias from excluding smaller exchanges is small because by the late 19th century, stock markets in many countries were already quite centralised, and many securities that were chiefly traded on smaller markets were often also quoted on the main stock exchange. The potential for bias is the greatest for early US data, where several large stock exchanges and an active curb market were in operation (Sylla, 2006). For the US and several other countries we, therefore, rely on benchmark year estimates to proxy the size of regional and curb exchanges relative to the main market.

The third challenge relates to excluding foreign stocks. For most of our estimates, the foreign stock share is either well measured (e.g. in recent data) or small (as for most of the mid-20th century data), so the measurement issues mainly concern the large international stock exchanges in the early 20th century, in particular the London stock exchange. We rely on a mixture of secondary sources and own estimates to adjust the equity market capitalization for foreign stocks, such that the remaining biases should be small, with the most likely direction leading us to slightly overstate the domestic stock market capitalization in the financial center countries during the early 20th century.

The Data Appendix contains a detailed discussion of the various quality checks and comparison with other existing capitalization estimates. In general, our data are in line with previous country-specific estimates constructed by financial historians and statisticians. When it comes to cross-country estimates of Goldsmith (1985) our estimates are typically below his national balance sheet data, because the Goldsmith (1985) estimates often include unlisted stocks, preference shares or bonds in the capitalization total, whereas ours focus on listed ordinary shares only. Our estimates are sometimes above and sometimes below those of Rajan and Zingales (2003), depending on the specific country and time period. For example, our estimates of the early 20th century US market capitalization are higher than those of Rajan and Zingales (2003), while those for the UK are lower, which goes towards addressing the criticisms of the Rajan and Zingales (2003) data raised by Sylla (2006), relating to the inclusion of curb and regional exchanges, and the exclusion of bonds and foreign shares.

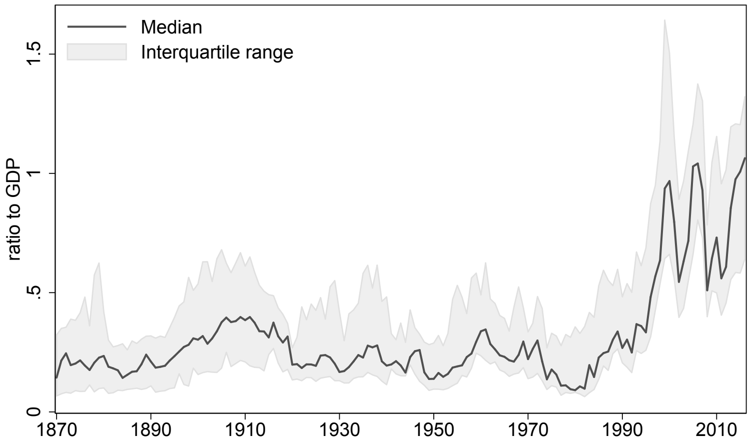
Our dataset advances the existing knowledge of market capitalization along two main dimensions. First, the estimates go beyond benchmark years which allow us to discern general trends, making sure they are not skewed by year-specific outliers, to look at both short-, medium- and long-run changes in capitalization within a unified framework, and, together with the data on stock prices and other financial variables, decompose the changes in capitalization into prices and quantities, and conduct a more detailed analysis of what the underlying drivers of stock market size are. None of the analysis in Sections 3.3–3.6 would have been possible to conduct without annual data. Second, our estimates are based on updated and generally higher-quality source data compared to previous work, covering an extended time period from 19th to 21st century, all based on a consistent definition of market cap. The next four sections present several novel facts and findings that emerge from these new data.

3.3 The Big Bang

Figure 3.3.1 shows the ratio of stock market capitalization to GDP across the 17 economies in our sample between 1870 to today. The solid black line is the sample median, and the shaded area is the interquartile range of country-level data.

From the end of the industrial revolution and up to the late 1980s, the size of the stock market had been relatively stable, at around one-third of a country's output. This was true both across time, with the median stock market cap to GDP ratio always below 0.5 during this period, and across countries, with the interquartile range oscillating between 10% and 60% of GDP. Market capitalization has experienced several pronounced swings during that time: the boom of the early 1900s

Figure 3.3.1. Stock market capitalization in advanced economies



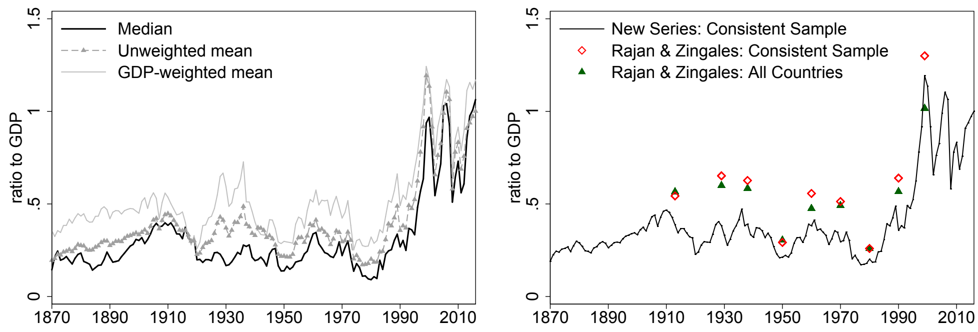
Notes: Stock market capitalization to GDP ratio, 17 countries. The solid line and the shaded area are, respectively, the median and interquartile range of the individual country capitalization ratios in each year.

during which capitalization roughly doubled, and the subsequent collapse during World War 1 when it halved again; the modest decline after World War 2, and the downturn during the stagflation of the 1970s. But each time, market capitalization eventually returned to its historical average level, around one-third of GDP.

Over the last several decades however, the stock market has undergone a historically unprecedented expansion. The median market cap to GDP ratio increased from 0.2 in 1980 to 1 in 2000, with some countries' stock markets growing to more than three times the size of their gross output. Moreover, this surge in stock market cap seems to have been persistent – despite sharp equity price corrections in the early 2000s and the Global Financial Crisis of 2008–09, market cap to GDP ratios today remain around three times larger than the historical norm. We loosely term this sudden and rapid growth of the stock market in the 1980s and 1990s as "the big bang".

Figure 3.3.2a shows that this hockey stick pattern holds regardless of how we aggregate the individual country data: the time trend of the unweighted and GDP-weighted stock market cap series is very similar to that of the median shown in Figure 3.3.1, and shows the sharp and persistent increase starting in the 1980s. Figure 3.3.3 further plots the trends for each individual country in our sample. The big bang is very much a cross-country phenomenon. The sharp equity market expansion in the 1980s and 1990s is evident in every single country in our dataset. In the vast majority of countries, the longer term time series pattern follows a hockey stick similar to that in Figure 3.3.1, with the peaks reached during the big bang period unsurpassed and unprecedented over the remainder of the sample. For a few countries in our dataset, the big bang can be seen as a return to some previously high

Figure 3.3.2. Alternative market capitalization estimates



(a) Alternative aggregators **(b)** Comparison to Rajan and Zingales (2003)

Notes: Stock market capitalization to GDP ratio. Left-hand panel: Median, unweighted and GDP-weighted averages of 17 countries. Right-hand panel: Estimates in our data compared to those of Rajan and Zingales (2003), unweighted averages. The consistent sample includes all countries in our dataset apart from Finland, Portugal and Spain.

level of market capitalization that was in place before a structural decline, financial or economic shocks – such as the two world wars or, in the case of Portugal, the Carnation revolution – reduced the size of the respective stock markets, only for them to experience a renaissance over the recent decades.

In comparison to existing literature, the big bang hockey stick differs from the U-shape "great reversals" pattern documented by Rajan and Zingales (2003) (henceforth RZ). RZ compiled data on market capitalization and other financial development indicators at benchmark years between 1913 and 1999, and argued that markets were well developed during the early 20th century, subdued during the mid-20th century, and bounced back over the more recent period. Figure 3.3.2b compares our market capitalization estimates to those of RZ. To improve comparability, we have excluded Finland, Portugal and Spain, which are present in our sample but not that of RZ, from our series (solid black line). The figure also presents the original RZ estimates for 22 countries (green triangles), and their estimates for the 14 countries in our reduced consistent sample (red diamonds).

We can see that the differences between our estimates and those of RZ is not driven by sample composition. Even though their sample includes some countries that are absent in ours, this makes little difference: countries with high market capitalization to GDP ratios in the 1913 RZ data – such as Cuba and Egypt – are counterbalanced by others with relatively low ratios, such as Russia and India. Some of the differences can be attributed to the improved quality of our data. Earlier estimates of stock market capitalization sometimes lacked accuracy because they included securities other than the ordinary shares of domestic companies – for example, bonds – or did not include data from smaller stock exchanges. But as with the sample composition, these differences balance out to a certain extent: excluding bonds or

Figure 3.3.3. Stock market capitalization to GDP ratio in individual countries

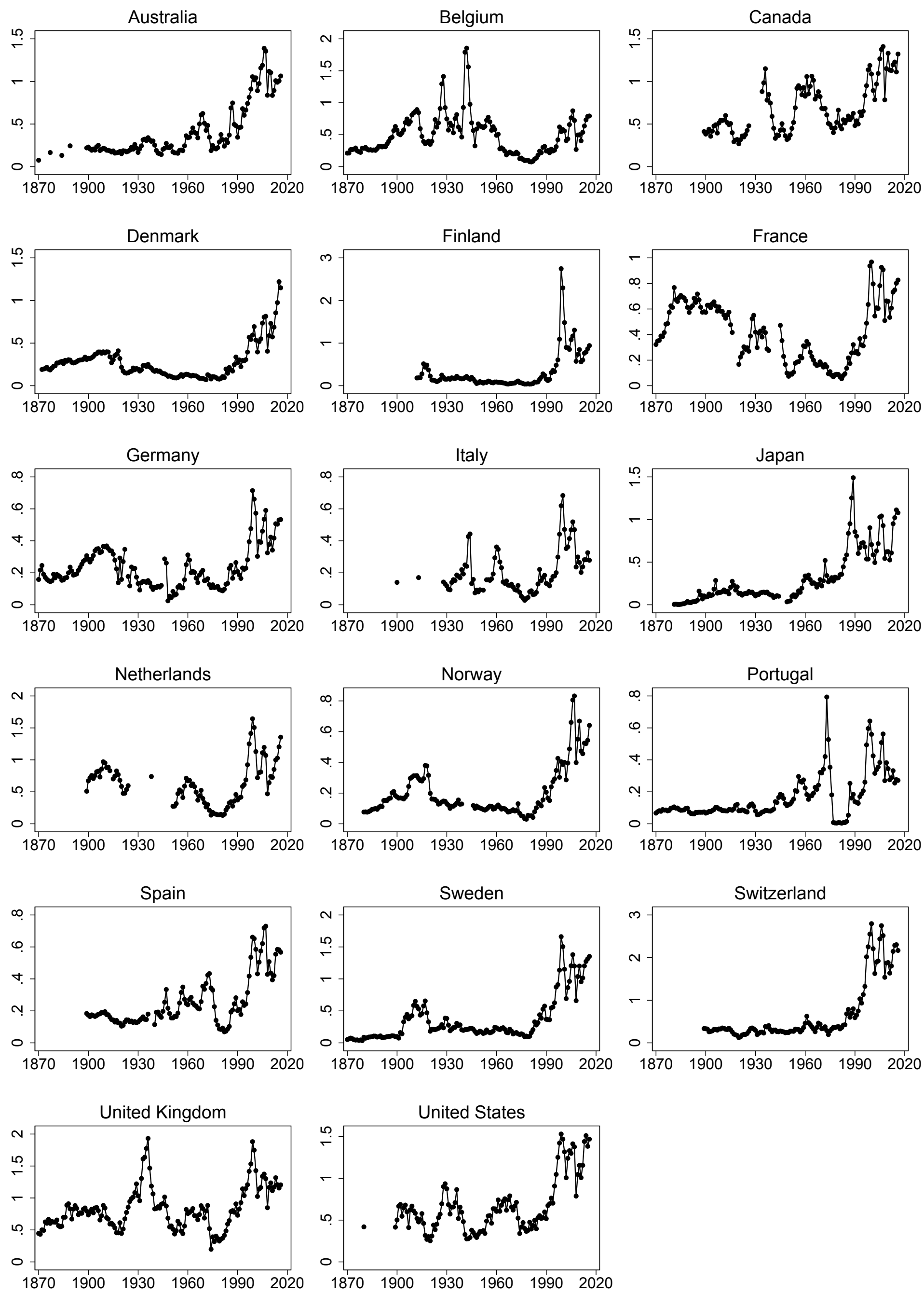
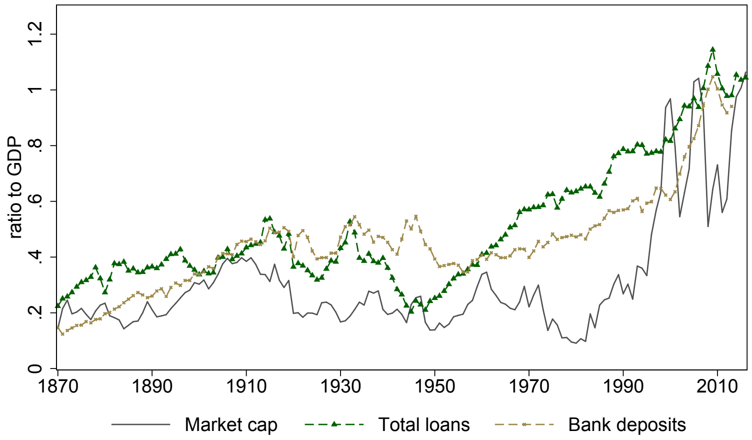


Figure 3.3.4. Stock market capitalization and other measures of financial development



Notes: Median ratio of stock market capitalization, total loans and bank deposits to GDP, 17 countries.

foreign shares reduces some of the market cap estimates, while including other stock exchanges increases them. Altogether, our aggregate stock market capitalization estimates are somewhat below those of RZ, especially for the mid-20th century period, but the figures are broadly comparable. The extensive Data Appendix presents a detailed comparison of our data with alternative estimates for each country, including those of RZ. The averages in Figure 3.3.2b do obscure some sizeable differences for individual countries, as can be seen in the comparisons for Australia, Canada, Japan and Norway in Appendix Figures D.1, D.3, D.9 and D.11.

The main reason that, up to this point, the big bang has been somewhat hidden from view, is the lack of annual data on stock market capitalization. Because equity prices are volatile, stock market capitalization varies substantially from year to year. The annual standard deviation in the market cap to GDP ratio is close to 0.4, around the same size as the mean of the series. The choice of the benchmark year thus has a significant influence on long-run market cap comparisons, and can obscure the underlying trends in the data. For their comparison, RZ mostly relied on years 1913, 1980 and 1999. But Figure 3.3.2b shows that 1980 was a trough of the equity price cycle, while 1913 and 1999 were peaks. Focussing only on these individual years makes the long-run market cap pattern more similar to a U shape. Adding the 18 years of data beyond 1999 further helps establish that the increase in market capitalization in the 1980s and 1990s was a persistent structural shift, rather than a short-lived equity boom.

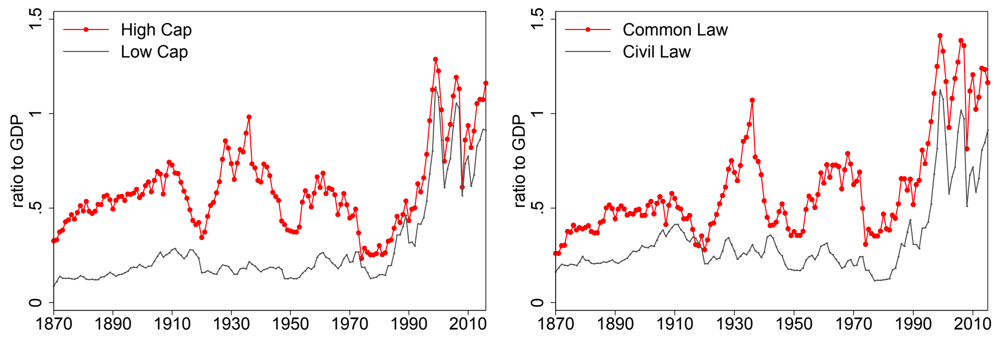
Does this mean that financial markets in the early 20th century, and as recently as 1980, were far less developed than they are today? We postpone the detailed discussion of this question until the next section, but some of the broad patterns in the data indicate that this may not be the case. First, the evolution of other measures of financial development points to a far more gradual and slow-moving improvement over the last 150 years. Figure 3.3.4 shows the evolution of total credit to the non-financial sector (green triangles), and total bank deposits (brown crosses) alongside market cap, all expressed as a ratio to GDP. The credit data come from Jordà, Schularick, and Taylor (2017), and deposit data – from Jordà, Richter, Schularick, and Taylor (2017). Both of these measures show a steady growth in the late 19th century, followed by a plateau and a fall around World War 2, before a steady rise starting in the 1950s and continuing until today. The time pattern of the changes is quite different to stock market cap: the 20th century trough occurs around the time of World War 2 rather than World War 1, and the recovery starts much earlier, and continues for a longer time and at a slower pace than the big bang.

Much of the literature on financial development has also emphasised the importance of persistence, or initial conditions in shaping future financial growth (King and Levine, 1993). This pattern of historical persistence is, to an extent, also echoed in our market capitalization measure. Figure 3.3.5a splits our countries into two groups: those which had large stock markets in 1910 (red diamonds), and those that did not (solid black line). Countries with large stock exchanges during that time consist of the financial centres in the UK, US and France, and smaller but highly developed and internationally integrated markets of the Netherlands and Belgium, as well as Canada, whose high capitalization was largely driven by the large caps of Canadian railway and financial stocks (Michie, 1988). This group of countries already had much larger stock exchanges as early as in 1870, and their advantage persisted throughout the 20th century. The big bang, however, marks a point of convergence between these two groups of countries: from 1990 onwards, average stock market capitalization in countries with initially small stock markets was similar to those with initially large markets. To some extent, this process of convergence already started before the big bang, as the high-cap group of countries was more heavily hit by the shocks of World War 2 and the 1970s stagflation.

A similar convergence pattern emerges when we group the countries according to their legal norms, shown in Figure 3.3.5b. La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1997) hypothesised that stock markets in common law countries tend to be more developed because of the more market-friendly legal norms. This pattern is largely borne out by the evidence in Figure 3.3.5b: common law countries (red circled line) – which, in our dataset, consist of Britain, Canada, US and Australia – have generally had larger market capitalization than civil law countries (solid black line), in particular during the mid-20th century.[3] But the differences had not always been large, and the two groups of countries have converged somewhat during the big

3. Also consistent with the legal origin thesis, civil law countries tend to have more bank-based, rather than market-based financial systems. But interestingly, this is not only because their market-based financial intermediation is relatively less developed (as shown in Figure 3.3.5b). Banking systems in civil law countries also tend to be more developed – relative to GDP – than those in common law countries. Appendix Figure 3.A.2 shows that civil law countries tend to have higher deposit-to-GDP and, especially, loan-to-GDP ratios, both throughout history and in present day.

Figure 3.3.5. Market capitalization across different groups of countries



(a) High vs low initial market cap

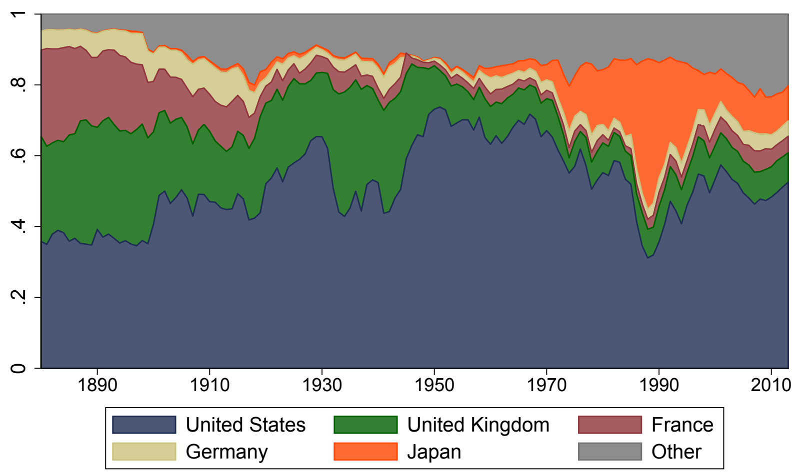
(b) Legal origin

Notes: Stock market capitalization to GDP ratio by country group, unweighted averages. Left-hand panel: High cap countries are Belgium, Canada, France, Netherlands, the UK and the US. Low cap countries are all other countries in our dataset. Right hand panel: Common law countries are Australia, Canada, the UK and the US. Civil law countries are all other countries in our dataset.

bang. Furthermore, market capitalization in both the "high-cap" group of countries in Figure 3.3.5a, and the common law countries in Figure 3.3.5b tends to be more volatile, or cyclical, with large peaks in the 1930s and 60s, and troughs around the two world wars and in the 1970s.

The new dataset also allows us to investigate the relative importance of individual domestic equity markets. Figure 3.3.6 shows the share of each country's stock market in the world market capitalization (i.e. the total of our 17 countries). It reports separate shares for the US, UK, France, Germany and Japan and lumps all other countries together. In 1880 world capital markets were roughly divided between three major players: the United States, France and Great Britain. This distribution, however, changed markedly during the subsequent 50 years. While the US was able to quickly increase its market share between 1880 and 1930, the French stock market's global importance more or less vanished. UK's market share also dwindled, albeit at a slower pace than France's. After the Second World War global equity markets became almost entirely dominated by the United States, with US equities accounting for roughly 70% of the global market cap in 1950. Even though the US has lost importance over recent decades, the size of its stock market today is still comparable to that of the other 16 economies grouped together. New equity markets have gained importance, with other countries slowly catching up, and Japan's market share expanding during the high growth era after World War 2. Even though Japan still has an important equity market today, it is the Japanese stock market bubble of the 1990s that stands out in the data. Capitalization of Japanese listed companies grew from 5% of the global market in 1970 to 40% in 1989 – comparable in size to the US – and collapsed thereafter.

Figure 3.3.6. World market capitalization shares



Notes: Shares of individual countries' capitalization in world total. Capitalization shares are computed by transforming domestic stock market capitalization into US dollars using historical exchange rates and dividing it by the sum of capitalizations of all 17 countries. Shares of the United States, the United Kingdom, France, Germany and Japan are shown separately. All other countries are combined together into one joint item.

Our long-run data show that stock market size had been relatively stable before a relatively recent upsurge in the 1980s and 1990s. This upsurge occurred across countries and has no historical precedent. It constitutes a structural break in the evolution of market cap, rather than a reversal to some previously high stock market cap level. At country level, market capitalization tends to be persistent, and shows some relation to legal norms – but the big bang also resulted in a convergence of stock market size across different economies. At the global level, total stock capitalization has been dominated by US equities until recent decades.

Can we interpret these patterns as changes in financial development? Was there no financial development for 100 years between 1870 and 1980, and are countries far more financially developed today than they were 30–40 years ago, and at any point from 1870 to today? To answer these questions, we need to understand what drives changes in stock market cap over these long periods of time, and across countries. Section 3.4 decomposes the market capitalization changes into quantities and prices, and Section 3.5 looks into the deeper underlying drivers of these structural trends.

3.4 Decomposing the Big Bang

We first seek to understand whether stock market cap growth is driven by quantities or prices – i.e. stock market issuance, by both new and existing firms, or the valuation of issued stocks. To do this, we decompose the market cap to GDP growth into issuances, valuations and GDP growth using a similar technique to the Piketty and

Zucman (2014) decomposition of growth in wealth-to-income ratios.[4] To derive the decomposition, we first note that total market capitalization *MCAP* is simply the the sum of the capitalizations – or quantity Q times prices P – of each individual share listed on the exchange:

$$MCAP_t = \sum_{i=1}^{N} P_{i,t} Q_{i,t}, \tag{3.4.1}$$

where N is the total number of listed shares. Rewriting equation (3.4.1) in difference terms, the change in market cap either comes about from higher quantities Q – i.e. issuance, or higher prices P:

$$MCAP_t = MCAP_{t-1} + Issuances_t + Capital\ Gains_t \tag{3.4.2}$$

$$Issuances_t = (Gross\ issues_t - Redemptions_t)/MCAP_{t-1} \tag{3.4.3}$$

$$Capital\ Gains_t = MCAP_{t-1} * P_t/P_{t-1} \tag{3.4.4}$$

Here $Issuances_t$ is total net equity issuance in proportion to previous year's market cap, $Capital\ Gains_t$ is the capital appreciation of the previous year's capitalization, and P_t is the value-weighted stock price index. Dividing through by GDP, rearranging and taking logs, we can write down the following linear approximation of the growth in the market cap to GDP ratio:

$$g_t^{MCAP/GDP} \approx iss_t + r_t^{eq} - g_t \tag{3.4.5}$$

Equation (3.4.5) breaks market cap to GDP growth down into three components: issuances (i.e. quantities), capital gains (i.e. changes in P), and real GDP growth. Here, $g^{MCAP/GDP}$ is the geometric growth in the market cap to GDP ratio, $g^{MCAP/GDP} = log(MCAP_t/GDP_t) - log(MCAP_{t-1}/GDP_{t-1})$. iss_t is the yearly net stock issuance relative to previous year's market cap, again expressed in terms of geometric growth: $iss_t = log(1 + Issuances_t/MCAP_{t-1})$. r_t^{eq} is the real equity capital gain, $r_t^{eq} = log(P_t/P_{t-1}) - log(1 + \pi_t)$, where π_t is the CPI inflation rate. g_t is real GDP growth, $g_t = log(GDP_t/GDP_{t-1}) - log(1 + \pi_t)$.

Table 3.4.1 shows this decomposition in our data, for the full sample and three different subperiods, which roughly correspond to the trend in market capitalization shown in Figure 3.3.1.[5] The subperiods cover the initial pre-WW1 market cap growth (column 2), the mid-20th century stagnation (column 3), and the big bang (column

4. Piketty and Zucman (2014) decompose the growth in the ratio of wealth to income into capital gains on wealth, income growth and saving rates – equivalent to, respectively, capital gains on equity, GDP growth and net issuance in our decomposition.

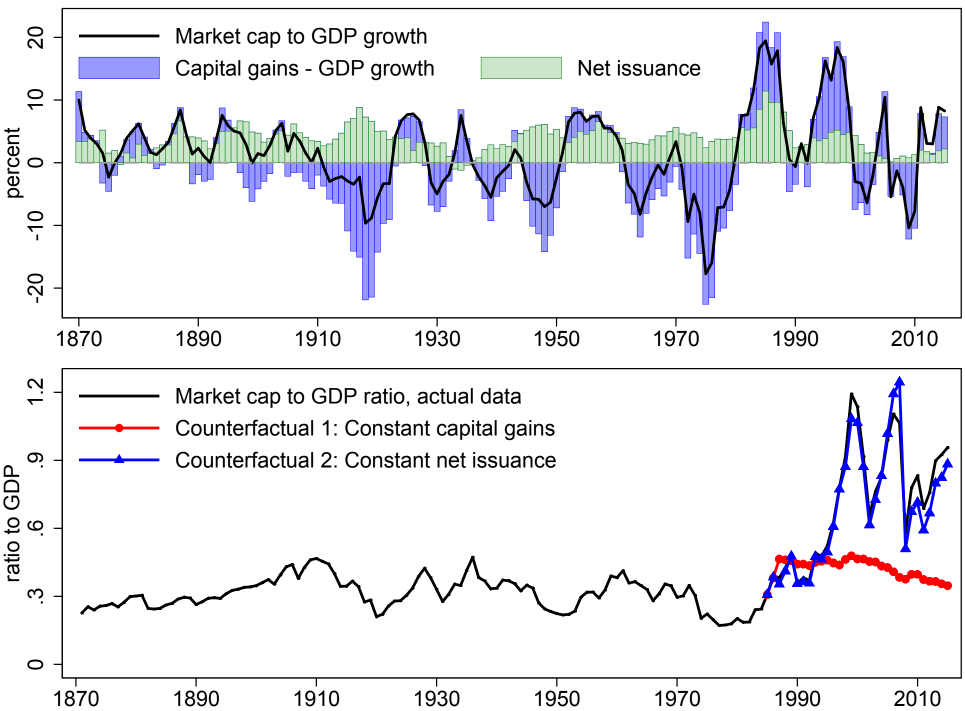
5. Table 3.4.1 also includes a small approximation error, which arises because of the log approximation, and because real GDP and real equity price growth use different deflators.

Table 3.4.1. Market capitalization growth decomposition

	(1) Full sample	(2) Pre 1914	(3) 1914–1985	(4) Post 1985
Market capitalization growth	1.54	2.47	-0.13	4.42
Decomposition of market capitalization growth into:				
Implied issuance to market cap	3.86	3.76	4.09	3.43
+ Real capital gain on equity	0.41	0.96	-1.17	3.41
− Real GDP growth	2.84	2.41	3.26	2.28
+ Approximation residual	0.11	0.16	0.21	-0.14
Observations	2068	448	1116	504

Notes: Decomposition of market cap to GDP ratio growth into issuances, capital gains and GDP growth based on equation (3.4.5). Market cap growth is the change in the log of market cap to GDP ratio. Implied issuance is the change in market cap not explained by equity prices or GDP growth. The sum of implied issuance and real capital gains, minus real GDP growth is equal to total market cap growth, subject to a small approximation residual from using log growth rates. Average of pooled cross-country observations.

Figure 3.4.1. Decomposition trends and counterfactual



Notes: Top panel: Decomposition of annual stock market cap to GDP growth into issuances, and capital gains less GDP growth, using equation (3.4.5). Centered five-year moving averages. For variable descriptions, see notes to Table 3.4.1 and main text. Bottom panel: Counterfactual market cap to GDP ratio evolution during the big bang. Constant capital gains counterfactual forces the real capital gains during 1985–2015 to equal the pre-1985 average. Constant net issuance counterfactual forces net issuance relative to market cap during 1985–2015 to equal the pre-1985 average. All data are unweighted averages of 17 countries.

4).[6] Starting with the full sample results in column 1, the average geometric growth in the stock market cap to GDP ratio of 1.6% is modest, but over 150 years it adds up to the market cap increase of close to 80% GDP, from 20% of GDP in 1870 to 100% of GDP in 2015.[7] Hardly any of this long-run growth is attributable to real capital gains, which average just 0.4% per year, far below the real GDP growth of 2.8%. Had there been no capital market issuance throughout the period, the market cap to GDP ratio would have been falling. The shortfall is made up by positive net issuance, which, on average, amounts to around 4% of market cap, or a little over 1% of GDP.

High net issuances were the driving factor propping up market capitalization over the long run. Long-run averages aside, however, most of the time variation in the market cap to GDP ratio can be attributed to changes in real capital gains. This can be most easily seen in Figure 3.4.1, which decomposes five-year moving average annual market capitalization growth (black line) into issuances (blue bars) and capital gains minus GDP growth (green bars).[8] Net issuances are very stable from year to year, and show little secular or cyclical patterns. Most of the variation in the stock market cap to GDP ratio is driven by short and medium run swings in real capital gains. Furthermore, these capital gain movements tend to drive market capitalization changes at horizons far longer than the typical business or financial cycle frequency. This can be most easily seen by going through the decomposition trends across the different historical subperiods, presented in Table 3.4.1 columns 2–4. The differences in market capitalization growth across these time periods, which generally last between 40 and 70 years, are largely driven by capital gains.

Table 3.4.1 column 2 presents the growth decomposition for the initial increase in market cap, from 0.2 of GDP in 1870 to 0.4 of GDP in 1913. Looking at this period in isolation, one could conclude that the main driver behind this increase was net stock issuance, and ultimately financial development, since issuance growth makes up the largest contribution to the growth in market cap. But in the bigger picture, this issuance growth of 3.9% is roughly equal to the full-sample average. The underlying drivers of this initial market cap increase are slightly above-trend real equity price growth (1% p.a vs long-run average of 0.4% p.a.) combined with slightly below-trend real GDP growth (2.4% p.a vs long-run average of 2.8% p.a.). The initial market cap increase can, therefore, be attributed to the near-absence of

6. We choose 1985 as a benchmark for two reasons. First, 1985 marks the point of recovery in stock market returns from the trough in 1980 to close to their historical average, and allows us to look through the equity market cycle. Second, it roughly coincides with the famous “big bang” financial sector liberalisation in the UK under Thatcher, which soon after took hold in many other countries in our sample.

7. The 1.6% growth rate applied to the average market cap to GDP ratio of 0.4 means an average annual increase in market cap of around 0.6% of GDP (1.6% × 0.4), adding up to 80% of GDP over 145 years.

8. Capital gains and GDP growth are combined to reduce complexity, but Appendix Figure B.1 presents all three series separately.

large shocks to equity valuations at the same time as the general macroeconomic performance was relatively weak (see also Figure 3.4.1).

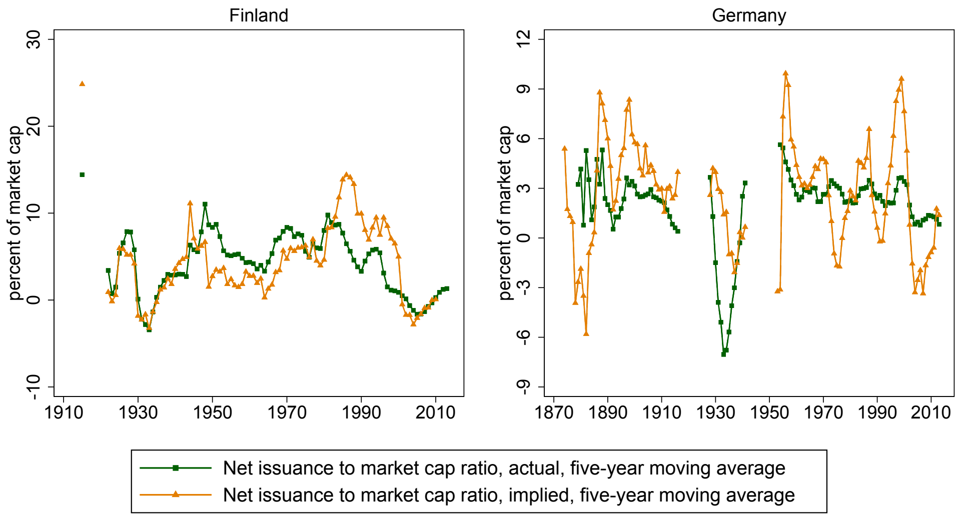
Moving on to the market cap stagnation during 1914–1985, Table 3.4.1 column 2 shows that, indeed, the average market cap to GDP growth during this time period was approximately zero. The relatively robust net issuance (on average, 4.1% of market cap) was held back by negative real capital gains (-1.2% p.a.), and higher than average real GDP growth (3.3% p.a.). Figure 3.4.1 shows that these negative capital gains were a result of several large shocks that hit the equity market during this period. The largest aggregate shocks occurred during World War 1 and the 1970s stagflation. World War 2 and the Great Depression also had a negative, but smaller, effect on the stock market. These aggregate trends mask further shocks that hit individual countries, with the largest of these occurring during the Portuguese Carnation Revolution of 1974. In its aftermath, the Portuguese stock market lost roughly 98% of its value (see Appendix Figure D.12). The impact of other political shocks, such as the Spanish Civil war and the Nazi occupation was sometimes negative, but generally small (see Le Bris, 2012, for the case study of occupied France).

Column 4 of Table 3.4.1 captures the period of the big bang, or explosive and persistent growth in market capitalization in the 1980s and 1990s. On average, market cap to GDP ratios grew by around 4.4% per year, or 3.2% of GDP (4.4% times the average market cap to GDP ratio of 0.7). This growth was not driven by net issuances: these were on average slightly lower than over the full sample.[9] Lower real GDP growth made a positive, but relatively small contribution (2.3% p.a. vs 2.8% p.a. full-sample average). Instead, the big bang is largely driven by higher real capital gains. Stock prices grew at a rate of 3.4% per year in real terms, eight times the full-sample average. Figure 3.4.1 shows that these increases largely occurred in the 1980s and 1990s, and were only partially tempered by the burst of the dot-com bubble and the Global Financial Crisis of the 2000s.

The bottom panel of Figure 3.4.1 further illustrates this result. It displays two counterfactual market cap evolutions together with the actual data (solid black line). The first counterfactual, marked by red diamonds, shows what the market cap evolution after 1985 would have been if we fixed the capital gains to their pre-1985 average. Under this scenario, all changes in the stock market cap from 1985 onwards are attributable to net issuance and real GDP growth. Without abnormally high capital gains in the 1980s and 1990s, market capitalization stays relatively constant, and even shows a mild decline over the last 20 years. The second counterfactual (blue triangles) instead fixes issuances to their pre-1985 mean, and attributes all the growth in stock market cap after 1985 to real capital gains. In line with the dis-

9. Because stock market cap to GDP grew over this period, net issuance relative to GDP (rather than market cap, as shown in Table 3.4.1) was actually slightly higher than in the previous subperiods, but the differences are small (average net issuance is 1.2% of GDP over the full sample, and 1.5% of GDP after 1985.

Figure 3.4.2. Implied and actual net equity issuance



Notes: Actual and implied net equity issuance in year *t* as a proportion of stock market capitalization at the end of year *t* − 1. Centered five-year moving averages. Actual issuance is either the change in book value of listed firms, or the market value of gross equity issues less gross redemptions. Implied issuance is calculated as the change in stock market capitalization that is not explained by capital gains divided by last years market capitalization.

cussion above, this counterfactual closely follows the actual data. In essence, the big bang is simply a marked and persistent increase in stock market valuations.

How robust are the findings in Table 3.4.1? The first thing to note is that our net issuance data are simply a proxy, or residual: the change in market capitalization that is not accounted for by capital gains. It is, therefore, subject to measurement error. This measurement error would arise if there is an inconsistency between the equity price index and the stock market capitalization measure. Such inconsistencies are typically driven by either timing or coverage differences. In terms of timing, we seek to measure both market capitalization and equity prices at the end of each calendar year. But this is not always possible, especially when it comes to historical data. If, for example, market capitalization is measured at end of the year, equity prices at mid-year, and the stock price increases during the second half of the year, this increase would be interpreted as higher net issuance. In terms of coverage, market capitalization, by definition, covers all listed firms. The stock price index, however, may be based on a subsample of firms, or an unweighted average of all firms, which means that the price gain in the index may not be reflected of the all-firm market cap weighted average. For the vast majority of the sample, we use best-practice all-firm value weighted equity price indices, as detailed in Jordà, Knoll, et al. (2019). But for some countries and years, we rely on a subsample of firms, or weights other than market capitalization, which may create a discrepancy.

To check the extent of this measurement error, we need to compare our implied issuance data with actual net equity issuance. To be consistent with the decomposition in (3.4.5), net equity issuance should capture all changes in listed capital by listed firms, and any new listings, measured at market value. Such a measure is difficult to obtain for the historical sample, which is precisely why we rely on the decomposition proxied by equation (3.4.5) in the first place. But for a few countries, we were able to obtain high quality issuance data that allow for such a historical comparison.

Figure 3.4.2 compares the actual and implied equity issuance series for two countries with the best historical data coverage – Germany and Finland. The light-orange line is the implied net equity issuance computed using equation (3.4.5). The dark-green line is the actual net issuance data. For most years, this is measured as the change in total book value of capital of listed firms. For the more recent period (post-1950 for Germany and post-1990 for Finland), it measures the market value of net issuance by listed firms. Both series use five-year moving averages to get a better overview of the trends. For both countries, the implied and actual net equity issuance have similar magnitudes and move closely together. The implied issuance series tends to be more volatile because of the measurement error discussed above. This analysis suggests that, if anything, the actual net issuance is more stable across time than the implied issuance data. This supports the finding in Table 3.4.1 that net issuance, despite a large contribution to the overall growth of market cap over the long run, makes little difference to the time variation in that growth, including the rapid increase in market capitalization during the big bang.

We also check whether aggregate trends in Table 3.4.1 mask cross-country heterogeneity. In particular, as discussed in Section 3.3, the US stock market is by far the largest globally, so the big bang on a global scale would largely be influenced by developments in the US, rather than the cross-country averages in Table 3.4.1. Table 3.4.2 presents the decomposition results for each country, for the periods before (columns 1 - 3) and after the big bang (columns 4–6), with aggregate market cap growth (columns 1 and 4) made up by net equity issuance (columns 2 and 5) and the $r_t^{eq} - g_t$ gap (columns 3 and 6).

Before 1985, market capitalization growth in most countries was low or slightly negative. Only Japan, which started with very low market capitalization and underwent a rapid stock market boom in the 1970s and early 1980s, experienced a robustly positive capitalization growth during this time period, driven by high net equity issuances. For every other country in the sample, positive net equity issuance is roughly offset by the negative gap between equity returns and GDP growth, both around 2–4 per cent. The decline in the global importance of the Paris and London stock exchanges (Figure 3.3.6), and the devastating impact of the Portuguese Carnation Revolution are evidenced by the below-average market cap to GDP growth in the corresponding countries. The low growth rates in France and Portugal are

Table 3.4.2. Decomposition of market cap to GDP growth by country and period

Country	(1)	(2)	(3)	(4)	(5)	(6)
	Pre 1985			Post 1985		
	$g_t^{MCAP/GDP}$	iss_t	$r_t^{eq} - g_t$	$g_t^{MCAP/GDP}$	iss_t	$r_t^{eq} - g_t$
Australia	.61	2.28	-1.72	3.33	4.47	-.99
Belgium	-.15	2.85	-3.36	4.93	2.91	2.74
Canada	-.42	2.54	-3.42	2.38	1.39	.71
Germany	1.09	2.64	-1.66	2.82	1.86	1.13
Denmark	.1	2.64	-2.98	5.81	1.08	4.87
Finland	-.76	5.17	-5.16	7.02	4.86	3.43
France	-.46	5.55	-6.34	5.89	4.1	2.05
Italy	1.11	6.91	-5.22	3.06	4.88	-1.65
Japan	5.88	10.21	-4.46	2.15	2.02	-.29
Netherlands	1.65	3.7	-1.97	4.35	3.37	1.54
Norway	.67	5.64	-4.69	4.59	3.61	.73
Portugal	-1.32	4.81	-5.05	9.79	15.27	-2.87
Spain	-.15	4.55	-5.18	5.78	4.79	1.05
Sweden	1.57	3.87	-2.31	4.32	.24	3.96
Switzerland	.82	2.58	-1.71	4.06	.9	2.86
UK	.32	1.35	-1.23	1.94	2.08	.07
USA	.28	1.97	-2.16	3.2	.3	2.89

Notes: Decomposition of log market cap to GDP growth into issuances, capital gains less GDP growth using equation (3.4.5). $g_t^{MCAP/GDP}$ is the growth in stock market capitalization, iss_t is net issuance relative to last period's market cap, and $r_t^{eq} - g_t$ is the difference between capital gains on equity and the GDP growth rate. Using log growth rates creates a small approximation residual. Period coverage differs across countries.

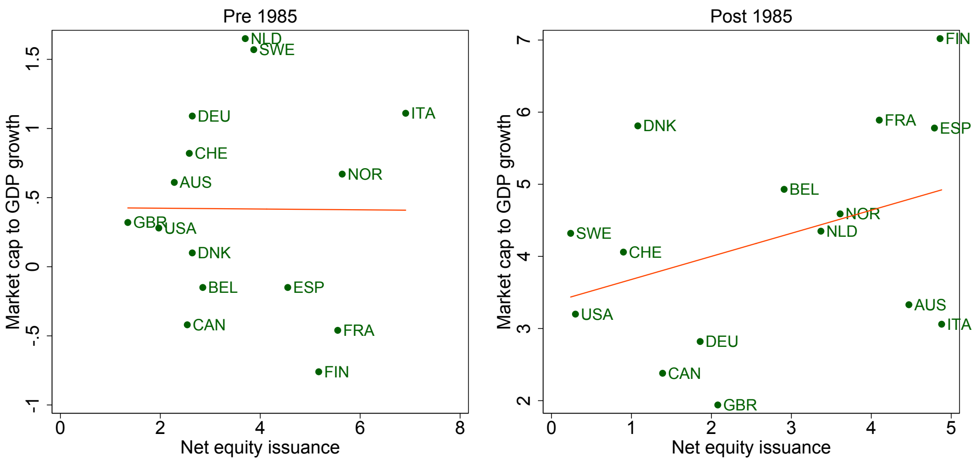
largely attributed to low equity returns, and the stagnation of the UK market – to low issuance.

Turning to the period of the big bang, market capitalization in all countries apart from Japan – which stagnated after the burst of its stock market bubble – grew at high rates, typically close to 5% p.a. This growth was driven by sharp increases in the $r_t^{eq} - g_t$ gap, which, in contrast to the pre-1985 period, is positive or close to zero in the majority of countries. Net issuance is positive in every country, but close to full sample average everywhere apart from Portugal – a special case reflecting the re-emergence of the stock market following the revolution.

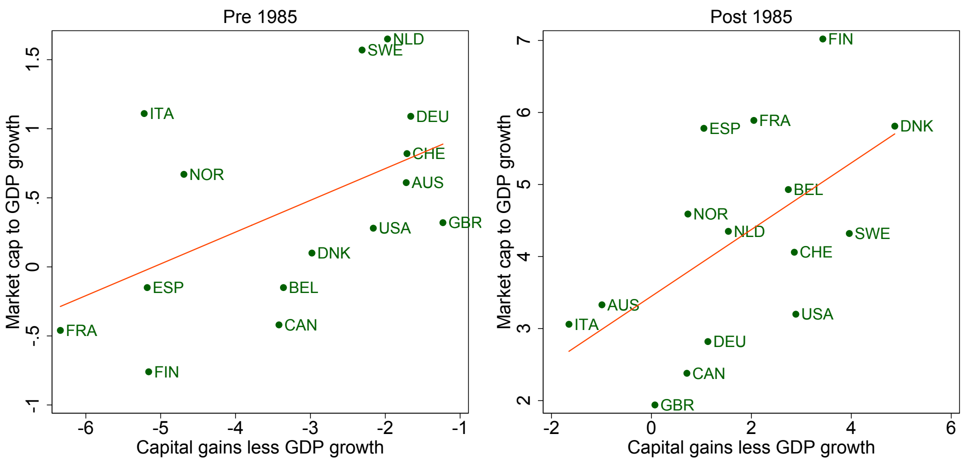
Figure 3.4.3 graphically illustrates the cross-country correlation of market capitalization growth (y axis) with changes in net issuance (x axis, Figure 3.4.3a), and capital gains less GDP growth (x axis, Figure 3.4.3b), again, for the pre and post big bang periods. We exclude Japan and Portugal from the sample, so that the analysis is not overshadowed by the effects of the Japanese stock market bubble and the Portuguese revolution. Both before and after the big bang, the cross-country differences in the market cap to GDP ratio are largely explained by capital gains, not issuances. Net equity issuance shows no correlation with cross-country market

Figure 3.4.3. Cross-country correlations between market cap, issuances and capital gains

(a) Market cap growth and equity issuance



(b) Market cap growth and capital gains



Note: Full sample and post-1980 averages of (log) growth in stock market cap to GDP ratio, issuances relative to market cap, and capital gains less GDP growth rate. Japan and Portugal outliers excluded (Japan's rise from very low market cap in the 19th century to the stock bubble in the 1980s, and Portugal's Carnation revolution otherwise skew the overall results).

cap growth before 1985, and only a small positive correlation after 1985 (Figure 3.4.3a). By contrast, capital gains show a large positive correlation (Figure 3.4.3b). After 1985, market capitalization grows almost one for one with the $r_t^{eq} - g_t$ gap, as shown by the near-45 degree slope of the line in the right-hand panel of Figure 3.4.3b.

The decomposition of stock market cap growth into issuance, capital gains and GDP growth suggests that even at long horizons, the time trends and cross-country differences in market capitalization are largely a result of changing prices, not quantities. This is particularly true for the period that saw the rapid expansion of the stock market over the recent decades – the big bang, but is also the case for earlier historical periods. The next section studies the long-run evolution of the possible underlying drivers of these changing stock valuations.

3.5 Structural drivers of stock valuations

The 1980s and 1990s saw a structural increase in stock market valuations. To shed light on the possible drivers of this increase, it helps to go back to equation (3.4.1), and express the stock price as the sum of expected future cashflows $CF_{i,t+j}$, net of tax τ_{t+j}, discounted at rate r_t:

$$MCAP_t = \sum_{i=1}^{N} P_{i,t} Q_{i,t} = \sum_{i=1}^{N} Q_{i,t} \sum_{j=1}^{\infty} \frac{CF_{i,t+j}(1-\tau_{t+j})}{(1+r_t)^j} \tag{3.5.1}$$

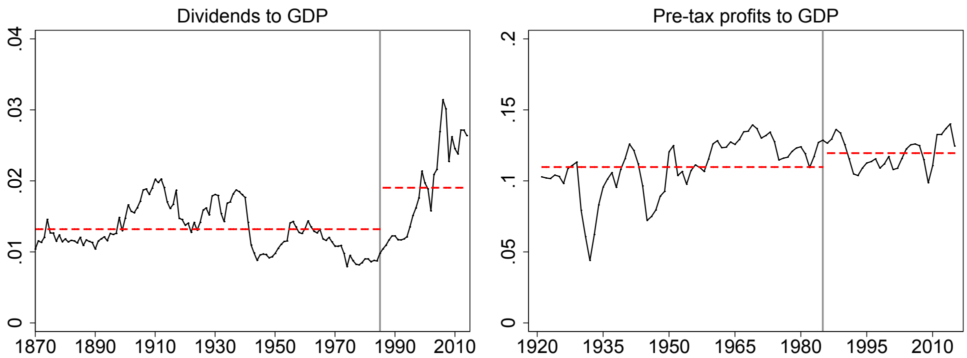
After ruling out quantity based explanations in section 3.4, an increase in stock valuations can occur for the three following reasons: higher expected cashflows $CF_{i,t+j}$, lower taxes τ_{t+j}, or lower discount rates r_t. We start by examining the long-run evolution of the fundamentals underlying stock valuations – pre-tax cashflows CF and taxes τ – before discussing the historical trend in the rate at which these future fundamentals are discounted, r.

3.5.1 Pre-tax cashflows

Over the long run, equity cashflows CF_t should correspond to total dividends paid. The left-hand panel of Figure 3.5.1 shows the evolution of dividend payments of listed companies relative to GDP.[10] The big bang coincided with a structural and persistent increase in dividends, which rose by a factor of 2.5 between 1985 and 2015, from 1% to 2.5% of GDP. This substantial increase in dividends to GDP occurred virtually universally across countries, with the notable exception of the US.

10. The dividend-to-GDP ratio for each country is calculated as the dividend yield D_t/P_t, or dividends paid throughout the year D_t divided by the end-of-year share price P_t, multiplied by the market cap to GDP ratio at the end of the year, $MCAP_t/GDP_t$.

Figure 3.5.1. Gross equity cashflows and the big bang



Note: Unweighted averages, 17 countries (left-hand panel) and four countries (right-hand panel). Solid black line indicates the start of the big bang in 1985. Dashed horizontal lines show the average of the series before and after the big bang.

The dividend-to-GDP ratio also shows positive co-movement with market cap over the earlier historical period.

Realised dividend payments are, however, an imperfect proxy for the present value of expected future cashflows in equation (3.5.1). Year-on-year dividend changes can be driven by variation in payout policies of firms, or intertemporal substitution between current and future payouts. Tracking the overall profitability of the corporate sector may be a better way to gauge the underlying trend in expected future cashflows. The right-hand panel of Figure 3.5.1 displays the total pre-tax profits of private corporations, again relative to gross output.[11] Unlike dividends, the trend in the level of total profits is far less clear, with only a slight increase occurring since the 1980s.[12] Appendix Figure B.2 shows that after-tax profits have also changed little since 1980. This evidence is consistent with Gutierrez (2017), who shows that in cross-country data, profit shares over the recent decades have been surprisingly stable.

Several potential explanations can help reconcile the differential trends in dividend payments and corporate profitability. First, the profit measure in Figure 3.5.1 may understate the growth in economic profits because it does not account for changes in "factorless income" – a part of GDP factor income that is not assigned to any specific production factor. Eggertsson, Robbins, and Wold (2018) show that in the US, the total of corporate profits and factorless income has increased substantially over recent decades. But Karabarbounis and Neiman (2018) argue that a large part of this increase in factorless income is driven by lower risk premiums,

11. The dividend-to-GDP data cover the full sample, whereas profits to GDP cover four countries – Canada, France, Japan and the US – and start in 1920.

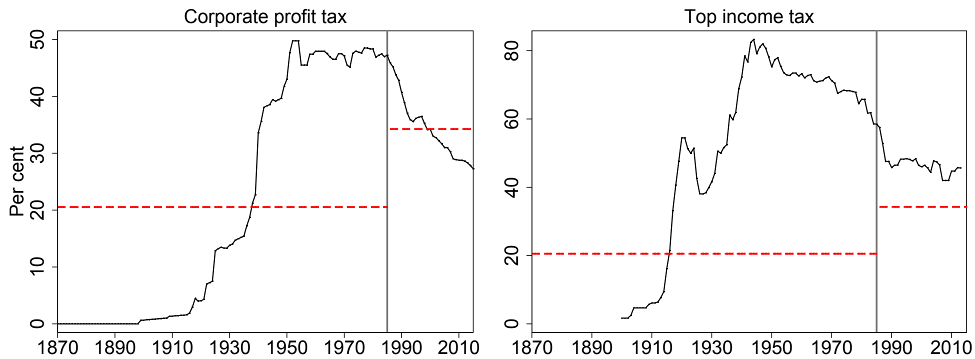
12. The increase becomes slightly larger if we limit the sample to the US, and goes away if we exclude the large, mostly unanticipated, profitability shocks of the Great Depression, World War 2 and the Global Financial Crisis.

which is precisely what we want to exclude from our cashflow measures; hence we do not include this type of income in our baseline calculation in Figure 3.5.1. Second, the dividends in Figure 3.5.1 are for listed firms only, whereas the profit data cover both listed and unlisted firms. Some of the increase in dividends can, therefore, be attributed to a compositional shift, whereby profits of listed firms grow at the expense of unlisted firms, for example if large listed companies accumulate market power (De Loecker and Eeckhout, 2017, provide evidence of increasing market power in the US). Appendix Figure B.3 shows that indeed, US listed firms' profits have grown faster than those of unlisted firms over recent decades. Third, firms in advanced economies have been declaring an increasingly large share of their profits off-shore to minimise corporate tax payments (Zucman, 2014). These off-shore profits would be absent from the advanced economy profit data. Finally, higher dividend payments may be a result of changes in payout policy, and have little to do with expected future profits or cashflows.

3.5.2 Taxes

A reduction in taxes τ_t increases the cashflows received by investors, and should drive up valuations P_t even if the pre-tax cashflows CF_t remain unchanged. For example, McGrattan and Prescott (2005) argue that a large part of the recent increase in equity valuations in the US can be attributed to changes in the corporate tax code. Corporate cashflows are generally taxed on two levels: first a corporate tax is applied to total profits, and then any distributions or realised capital gains are taxed as income. Sometimes allowances for double-taxation are made, so that, for example, dividends are only taxed once. Regardless, we consider both types of taxes in isolation, and it turns out that they follow a similar historical trend.

Figure 3.5.2. Taxation and the big bang



Note: Unweighted averages, 4 countries (right-hand panel) and 7 countries (right-hand panel). Solid black line indicates the start of the big bang in 1985. Dashed horizontal lines show the average of the series before and after the big bang.

Figure 3.5.2 plots the long-run evolution of taxes on corporate profits (left-hand panel) and top personal incomes (right-hand panel) across countries, thus capturing the two levels of taxation discussed above. The corporate income tax measures the deductions from cashflows before distribution. The top income tax only serves as a rough proxy for dividends and capital gains taxation, but given that stocks are typically owned by households in the top percentiles of the income and wealth distribution, and dividends are typically taxed as income, it remains informative of the likely marginal tax rate on distributed profits. The income tax data are an average of seven countries: Canada, France, Germany, Italy, Japan, UK and the US and come from from Roine and Waldenström (2012) and Piketty (2014). The corporate tax data are based on a somewhat smaller subset of four countries with long-run data – Australia, Germany, Japan and the US – but cover the remaining countries later on. All countries within the sample do, however, follow a similar time pattern.

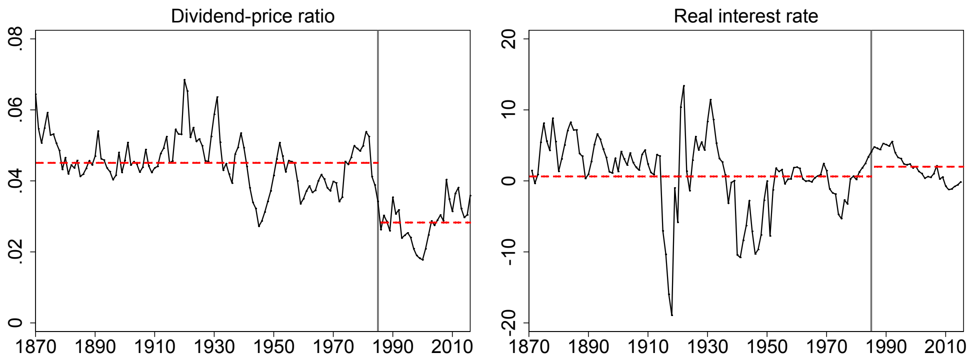
Both corporate and top income taxes were close to zero in the late 19th and early 20th century, before rapidly shooting up to reach levels of close to 50% and 80% respectively shortly after World War 2. During the 1980s, governments began cutting taxes, with rates eventually falling to around 30% for corporates and 40% for top incomes. On the surface, the timing of these tax cuts roughly coincides with the big bang (Figure 3.5.2 vertical black line in 1985). But looking at the longer run historical picture, the relationship becomes much weaker. Both corporate and income taxes were near zero up to 1910s or 1920s, and below current levels up until World War 2. The sample averages before 1985 are well below the post 1985 levels. And yet, stock market capitalization and stock valuations are much higher today than in the early 20th century (Figure 3.3.1).

3.5.3 Discount rates

Ex ante discount rates on equities are not directly observable. But Campbell and Shiller (1988) show that the discount rate can be proxied by the dividend-price ratio, as long as this ratio helps forecast future stock returns. Kuvshinov (2018) shows that high stock valuations – or low dividend-price ratios – do forecast low future stock returns at short and medium term horizons in our historical sample. This means that the dividend-price can be used to proxy the equity discount rate ($1 + r_t$ in equation (3.5.1)).

The left-hand panel of Figure 3.5.3 plots the long-run evolution of the average dividend-price ratio across the 17 countries in our sample. The vertical line marks the start of the big bang in 1985, and the two dashed lines plot the pre- and post-big bang sample averages of the series. Discount rates were stable until the 1980s at around 4–5% p.a. Around the start of the big bang, discount rates fell sharply and reached the all-time historical trough of 2% at the height of the dot-com bubble in 2000. Since then, discount rates recovered somewhat and increased in the Global

Figure 3.5.3. Discount rates and the big bang



Note: Unweighted averages, 17 countries. Solid black line indicates the start of the big bang in 1985. Dashed horizontal lines show the average of the series before and after the big bang. Real interest rate is the short-term government bill rate minus inflation.

Financial Crisis, but remain substantially below their pre-1985 mean.[13] The trend is apparent not only in price-dividend, but also price-earnings ratios, which are relatively unaffected by changes in payout policy, and the fact that distributions to shareholders in recent decades have often taken the form of stock buybacks, rather than dividends. Appendix Figure B.4 shows that the US buyback-adjusted earnings-price ratio shows a similar downward trend after 1985. These facts suggest that at least part of the increase in stock valuations, and hence market cap, since the 1980s can be attributed to lower discount rates.

The fall in the discount rate since the 1980s could correspond to either lower risk premiums, or a lower risk-free rate. It is difficult to discern these two factors because data on ex ante risk free rates are not available for our long-run sample. The right-hand panel of Figure 3.5.3 looks at ex post realised risk-free rates instead, and plots the average real short-term government bill rate across the 17 countries in our sample, as well as the pre and post big bang averages. Even though real rates have fallen since the 1980s, this fall does not appear particularly large from a historical perspective, with the current safe rate level close to the long-run average. It could, however, be the case that the inflation shocks in our data hide a somewhat more subtle decline in the ex ante real rate (Del Negro, Giannone, Giannoni, and Tambalotti, 2019, provide evidence supportive of this claim). But as Appendix Figure B.5 shows, using inflation forecasts instead of realised inflation to construct a better proxy of the ex ante real rate, results in roughly the same long-run trends. And even the 1950s and 1960s – two time periods without substantial positive inflation surprises during which we would expect to see a close correspondence between ex

13. Consistent with this, structural break analysis suggests that most countries in the sample experienced a structural break in the dividend-price ratios in late 1980s or 1990s. Results available from authors upon request.

ante and ex post realised rates – saw similar levels of the real rate to today, but much lower market capitalization.[14]

Finally, the upturn in the discount rate since its trough in 2000 (Figure 3.5.3) is consistent with recent evidence presented in Farhi and Gourio (2018), who argue that the risk premium in the US has risen over the period 2001–2016. This recent uptick in the risk premium – albeit, taken relative to the all-time high of the dot-com boom – has moderated the post-1980 upsurge in market cap, with the long-run time trend stabilising at levels close to 100% of GDP.

3.5.4 Quantifying the contribution of different drivers

The long-run evidence in Figures 3.5.1, 3.5.2 and 3.5.3 suggests that the big bang is driven by a combination of more favourable stock market fundamentals, and lower ex ante compensation for risk demanded by equity investors. To quantify the relative contribution of potential drivers, we turn to cross-country explanatory regressions. Table 3.5.1 regresses the stock market cap to GDP ratio on dividends to GDP, dividend-price ratio, real interest rates and corporate taxes. In order to capture structural trends, we analyse the relationship in levels and changes happening over a medium (5-year) and long (10-year) horizon.[15]

Consistent with the long-run trends in Figures 3.5.1 and 3.5.3, stock market cap is strongly correlated with changes in discount rates and cashflows. On the contrary, market cap shows no correlation with current or future corporate taxes, and real interest rates. Contemporaneously, a one percentage point increase in the dividend to GDP ratio predicts around 20–25 percentage points higher market cap to GDP, consistent with the average price-dividend ratio of around 30 in our sample. But market cap shows little co-movement with future dividend payments. One percentage point lower dividend-price ratios predict 7–10 percentage point higher market cap.

The lack of correlation with taxes is relatively robust: for example, it also holds for income taxes, effective corporate tax rates – i.e. total taxes paid as a share of corporate profits – and under a variety of alternative regression specifications. Appendix Table B.1 shows that if we limit the sample to the post-1985 period, thereby excluding the early-20th century period of low taxes and low market cap, we observe a small negative correlation between taxes and capitalization in some specifications, but the effects remain relatively weak. Appendix Table B.2 shows that using inflation forecasts instead of realised inflation to obtain a better proxy for the ex ante safe rate results in a somewhat stronger correlation between the safe rate and market

14. These time periods also show a similar trend level of the world interest rate to today in the estimates of Del Negro et al. (2019).

15. The levels specification takes the level of each variable at time t, but the results are unchanged if we smooth the data by applying a 10-year moving average filter to look through short-term trends. Results are available from authors upon request.

Table 3.5.1. Quantifying the relative contribution of stock market cap determinants

	Levels		5-Year Changes		10-Year Changes	
	(1)	(2)	(3)	(4)	(5)	(6)
D_t/GDP_t	25.27***	26.21***	23.75***	30.39***	25.08***	26.40***
	(2.4)	(2.8)	(2.8)	(6.1)	(2.3)	(4.3)
D_{t+1}/GDP_{t+1}	1.36	0.89	0.45	2.35	3.31*	-1.45
	(1.9)	(1.9)	(1.1)	(2.5)	(1.8)	(2.6)
D_t/P_t	-8.89***	-9.89***	-6.77***	-8.44***	-5.88***	-6.73***
	(1.2)	(1.8)	(1.2)	(1.8)	(1.1)	(1.9)
r_t	0.03	-0.19	-0.05	-0.32	-0.16	-0.42
	(0.1)	(0.2)	(0.1)	(0.3)	(0.1)	(0.3)
τ_t^{corp}		-0.09		0.16		0.11
		(0.2)		(0.1)		(0.2)
τ_{t+1}^{corp}		0.00		0.05		-0.33*
		(0.2)		(0.1)		(0.2)
R^2	0.743	0.718	0.476	0.486	0.608	0.509
Observations	1879	849	1652	681	1475	520

Note: Regressions with country fixed effects and robust standard errors. Standard errors in parentheses. Columns (1) and (2) show regression coefficients of dividends to GDP, the dividend to price ratio and corporate tax rates on the market capitalization level. Column (3)–(5) report the regression coefficients of the analysis in five year and 10 year changes. For the five-year and ten-year change regressions, the one-year ahead variables such as D_{t+1}/GDP_{t+1} become five-year ahead variables, i.e. $D_{t+5}/GDP_{t+5} - D_t/GDP_t$. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

cap, but the regression coefficients of the dividend-price ratio remain in the same ballpark as in Table 3.5.1.[16]

Turning to pre-tax cashflows and risk premiums, since the early 1980s, dividend payments have increased by 2 percent of GDP (Figure 3.5.1), and the dividend-price ratio has fallen by around 1.5–2 percentage points (Figure 3.5.3). Using the above regression coefficient estimates, higher dividend cashflows have contributed to a roughly 50 percentage points increase in stock market cap to GDP (2×25), and lower discount rates added a further 16 percentage points (2×8). These two effects have also reinforced each other, with higher cashflows discounted at lower rates. Together they explain almost all of the increase in stock market cap during the big bang (80% of GDP), with little room left for additional factors such as taxes and equity issuance.

What are the potential mechanisms that could explain these observed patterns? The explanations that assign a dominant role to cashflows center around rising mark-

16. Inflation forecasts are based on past GDP growth and past inflation and estimated for each monetary regime individually. See footnote of Figure B.5 for more information.

ups of large firms. De Loecker and Eeckhout (2017) show that over recent decades, market power in the United States has increased substantially, and Diez, Leigh, and Tambunlertchai (2018) find a similar pattern in other advanced economies. De Loecker and Eeckhout (2017) argue that market power can explain increasing stock valuations in the US. Relatedly, Greenwald, Lettau, and Ludvigson (2014) show that the US labour share has substantial explanatory power for stock returns.

Turning to discount rate based explanations, the fact that the big bang has taken place in multiple countries, and has coincided with increases in valuations of other risky assets such as housing (Knoll, Schularick, and Steger, 2017) points towards the importance of global macroeconomic factors. These could include lower levels of macroeconomic risk – for example through falling consumption volatility, one of the explanations for the recent decline in the equity premium in the US (Lettau, Ludvigson, and Wachter, 2008). But as Appendix Figure B.6 shows, consumption volatility has been steadily declining since the late 19th century, while the global equity premium decline and the big bang are much more recent phenomena. Bianchi, Lettau, and Ludvigson (2016), instead, link the structural fall in the equity risk premium to the emergence of inflation targeting as the dominant monetary regime. Aside from lower risk, the fall in the equity premium could be driven by structurally higher demand for risky assets, both domestic – through increases in market access and participation rates – and foreign, through higher precautionary savings by the emerging economies (Bernanke, 2005). Consistent with the domestic demand channel, data from the Survey of Consumer Finances in Appendix Figure B.7 show that stock market participation in the US has increased substantially since the 1980s.

The evidence presented in the last two sections suggests that structurally lower equity risk premia have played an important role in driving the recent increases in stock market capitalization, and the big bang. This means that rather than measuring changes in stock issuance and financial development, market capitalization can serve as an informative indicator of time variation in the risk premium on stocks. We examine this claim more systematically in the next section.

3.6 Stock market capitalization and the equity risk premium

If stock market capitalization is, to an extent, driven by movements in the equity risk premium, it may, in of itself, contain information on investor valuations of stock market fundamentals, and in turn future equity return and risk. In this section, we draw on the vast literature that examines the cyclical patterns of returns and risk premia in financial markets to evaluate the impact of shifts in stock market capitalization on future stock market performance. We start by running return predictability regressions to test whether cyclical movements in market cap, on average, help forecast future returns for the equity investors. We then proceed by focussing on the tails of the market cap and return distributions, and test whether sharp run-ups in stock market cap display a range of characteristics typically associated with stock market bubbles.

3.6.1 Market capitalization as a predictor of stock returns

Stock market capitalization is correlated with equity risk premium measures such as the dividend-price ratio (Figure 3.5.3 and Table 3.5.1). This correlation, however, does not fully capture the link between market cap and risk premia. On the one hand, dividend-price ratios can change in response to expected cashflows as well as expected returns or risk premia (Campbell and Shiller, 1988). Stock market capitalization may, then, also reflect expected cashflow rather than expected return movements. On the other hand, it may be that stock market cap itself is a better measure of the equity risk premiums than the dividend-price ratio. In this case, the correlations in Figure 3.5.3 and Table 3.5.1 would understate the importance of time varying risk premia in determining the size of the equity market. These propositions can be easily tested within the framework of return predictability regressions (see Cochrane, 2011, for a summary). If market cap is driven by future cashflows, high market cap should forecast high future dividend growth. If it is driven by discount rates or risk premia, high market cap should forecast low future equity returns. We test these two hypotheses by running the following predictive regressions:

$$r_{t+1} = \beta_0 + \beta_1 log(MCAP_t/GDP_t) + \beta_2 log(P_t/D_t) + u_t \qquad (3.6.1)$$

$$dg_{t+1} = \gamma_0 + \gamma_1 log(MCAP_t/GDP_t) + \gamma_2 log(P_t/D_t) + e_t \qquad (3.6.2)$$

Here r_{t+1} is the log of real, or excess equity return (real measured net of inflation, and excess – relative to the short-term risk-free rate), dg_t is log of real dividend growth, $MCAP_t/GDP_t$ is the market capitalization to GDP ratio, and P_t/D_t is the price-dividend ratio – the variable that is commonly used in such predictive regressions. We use price-dividend ratios instead of dividend-price ratios to allow for an easier joint interpretation of the coefficients. If $\beta_1 < 0$, high market capitalization predicts low future returns, and signals low discount rates. If $\gamma_1 > 0$, high market cap signals high future cashflows. We run the regressions in logs to be consistent with the formulation in equation (3.5.1), but the results are unchanged if we run them in levels.

Table 3.6.1 presents the results of the return predictability regressions. The numbers show the predictive coefficients β and γ, when used to predict real returns (columns 1 and 2), excess returns (columns 3 and 4), and dividend growth (columns 5 and 6). The top panel shows predictability at an one-year ahead horizon and the bottom panel at a five-year ahead horizon. Several results stand out.

First, high stock market capitalization forecasts low equity returns. The estimated coefficient β_1 is negative for both real and excess returns, and at a one-year as well as five-year horizon. Using the richer specifications in column 2, a 10 percentage

Table 3.6.1. Stock market capitalization as a predictor of equity returns and dividends

	Real returns		Excess returns		Real dividend growth	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: *One-year ahead returns and dividend growth*						
$log(MCAP_t/GDP_t)$	-0.037***	-0.029**	-0.033***	-0.028***	-0.008	-0.053***
	(0.007)	(0.010)	(0.006)	(0.008)	(0.009)	(0.015)
$log(P_t/D_t)$		-0.030		-0.018		0.161***
		(0.017)		(0.017)		(0.035)
R^2	0.015	0.019	0.011	0.012	0.000	0.066
Observations	1986	1986	1986	1986	1986	1986
Panel 2: *Five-year ahead average returns and dividend growth*						
$log(MCAP_t/GDP_t)$	-0.039***	-0.035***	-0.033***	-0.030***	-0.004	-0.033***
	(0.006)	(0.007)	(0.005)	(0.004)	(0.006)	(0.009)
$log(P_t/D_t)$		-0.014		-0.012		0.102***
		(0.012)		(0.010)		(0.016)
R^2	0.084	0.088	0.064	0.068	0.001	0.182
Observations	1883	1883	1883	1883	1883	1883

Note: Returns and dividend growth are measured in logs. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Regressions with country fixed effects. Country-clustered standard errors in parentheses.

point increase in the stock market cap to GDP ratio (25% in relative terms) forecasts a 0.8 percentage points lower return 1 year ahead ($0.25 \times (-0.03) \times 1.048$), and a 5.2 percentage points lower cumulative return 5 years ahead ($0.25 \times (-0.04) \times 1.048 \times 5$). Columns 3 and 4 show that stock market cap predicts excess equity returns with a similar statistical significance, sign and magnitude of the coefficient. Consistent with the long-run evidence in Section 3.6 and Figure 3.5.3, this suggests that time variation in stock market cap is, to a larger reflect, capturing the changes in the discount rate and – given the strong predictive performance for excess returns – in particular the risk premium for equities.

Second, market cap is a better predictor of equity returns than the more commonly used price-dividend ratio. Once included in the same specification (columns 2 and 4), the coefficient on the price-dividend ratio becomes insignificant, and the R^2 stays roughly the same as with market cap alone. This is especially true for longer horizon regressions, and those for excess equity returns.

Third, high stock market capitalization is not a sign of high future cashflows. The coefficient γ_1 on real dividend growth in column 5 is statistically insignificant. Once the price-dividend ratio is added to the regression (column 6), the coefficient even turns significantly negative. 10 percentage point higher stock market cap forecasts 1.3 ppts lower real dividend growth one year ahead ($0.25 \times (-0.053) \times 1.003$), and 4.2 ppts lower dividend growth five years ahead ($0.25 \times (-0.033) \times 1.003 \times 5$). High price-dividend ratios are, on the contrary, a sign of high future cashflows.

Appendix Table C.1 presents the predictability regression estimates for the longer 10-year horizon, and for the post-1985 period. These are particularly important for our understanding of the long-run trends and structural breaks that underly the big bang. If anything, the patterns in the data discussed above become stronger. A 10 percentage point increase in stock market cap to GDP forecasts 17 ppts lower cumulative real equity returns, and 13 ppts lower real dividend growth 10 years ahead. After 1985, a 10 ppt increase in stock market cap to GDP forecasts 3.2 ppts lower returns one year ahead.

Warren Buffett famously called stock market capitalization the "the best single measure of where valuations stand at any given moment" (Buffett and Loomis, 2001). Our findings largely confirm his priors. But why does market capitalization do so well as an equity return predictor? The natural explanation is that it contains information on fundamentals, and issuance or quantities, that is not captured by other commonly used valuation measures. The superior performance of the market cap to GDP ratio when compared to the dividend-price ratio could be because GDP is a better measure of fundamentals than dividends. Dividends are often criticised for not incorporating all future cashflows to the shareholder and for being excessively smooth relative to firm profitability. Indeed, the price to GDP ratio and the output gap are reliable predictors of equity returns in the United States (Rangvid, 2006; Cooper and Priestley, 2008). Market cap to GDP might also outperform the dividend to price ratio due to the additional informational content from equity quantities. Even though quantity changes play a relatively small role in long-run structural trends (Section 3.4), cyclical swings in net equity issuance may still tell us something about future returns. Existing evidence for the US suggests that high equity issuance tends to precede periods of substandard market returns (Nelson, 1999; Baker and Wurgler, 2000). With our new data, we can test whether such patterns hold in our richer cross-country setting.

We first study whether GDP includes information about firm fundamentals that is not included in dividends. Table 3.6.2 replaces market cap to GDP with a price-GDP ratio. The format follows that of Table 3.6.1: we regress log real and excess returns, and real dividend growth, one and five years ahead, on the price-GDP ratio, alone and alongside the price-dividend ratio. Following Rangvid (2006) we construct the price-GDP ratio by dividing the nominal stock market index by nominal GDP. Unlike Rangvid (2006), we deduct a time trend from $log(P_t/Y_t)$ to obtain a stationary series.[17]

17. Equity capital gains were relatively minor during the first 100 years of our sample. GDP therefore grew at a faster pace than the stock market index creating a downward sloping trajectory of the price to GDP series. This differs from the 1929–2003 US series used in Rangvid (2006), which was largely stationary due to the relatively high capital gains in the corresponding sample.

Table 3.6.2. The price-GDP ratio as a predictor of equity returns and dividends

	Real returns		Excess returns		Real dividend growth	
Panel 1: *One-year ahead returns and dividend growth*						
	(1)	(2)	(3)	(4)	(5)	(6)
$log(P_t/Y_t)$	-0.059***	-0.052***	-0.038***	-0.034***	-0.031**	-0.062***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.012)	(0.009)
$log(P_t/D_t)$		-0.032**		-0.021*		0.145***
		(0.013)		(0.012)		(0.030)
R^2	0.033	0.038	0.013	0.015	0.006	0.063
Observations	2245	2194	2144	2093	2190	2190
Panel 2: *Five-year ahead average returns and dividend growth*						
$log(P_t/Y_t)$	-0.062***	-0.060***	-0.036***	-0.034***	-0.029***	-0.049***
	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)
$log(P_t/D_t)$		-0.010		-0.013		0.099***
		(0.013)		(0.010)		(0.016)
R^2	0.029	0.029	0.067	0.071	0.032	0.206
Observations	2162	2111	2048	1997	2095	2095

Note: Returns and dividend growth are measured in logs. P/Y is the detrended log of a cumulative stock price index divided by nominal GDP. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Regressions with country fixed effects. Country-clustered standard errors in parentheses.

Table 3.6.2 shows that the price-GDP ratio is a reliable predictor of equity returns and excess returns at short and medium term horizons. Furthermore, the price-GDP ratio incorporates information that is not included in the price-dividend ratio. Once included in the same specification, both variables predict returns one year ahead, and the price-GDP ratio dominates the horse race between the two predictors over longer horizons. These results suggests that GDP contains information about fundamentals that is not included in dividends. We investigate this question more explicitly by studying the predictive power of the dividend-GDP ratio in Appendix Table C.2. High dividends relative to GDP predict low future dividend growth, which suggests that dividends show a tendency to converge towards a stable share of output, with the short-run deviations from this stable ratio potentially reflecting deviations of dividends from fundamentals due to, for example, changes to corporate payout policy.

Table 3.6.3 tests whether the additional information on quantities helps improve the predictive performance of the stock market cap to GDP ratio. Here we run the predictive regression of one and five year ahead stock returns and dividend growth on the current level of net equity issuance. To reduce the potential measurement error in the net issuance series, which is estimated as a residual (see Section 3.4), we use the three-year backward-looking moving average of net issuance instead of

Table 3.6.3. Net equity issuance as a predictor of equity returns and dividends

	Real returns		Excess returns		Real dividend growth	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: *One-year ahead returns and dividend growth*						
Issuance/GDP	-0.844*	-0.780*	-0.619**	-0.559*	-0.251	-0.421
	(0.400)	(0.383)	(0.290)	(0.283)	(0.335)	(0.380)
$log(P_t/D_t)$		-0.046***		-0.043***		0.121***
		(0.010)		(0.012)		(0.029)
R^2	0.011	0.021	0.006	0.015	0.001	0.049
Observations	1903	1903	1903	1903	1903	1903
Panel 2: *Five-year ahead average returns and dividend growth*						
Issuance/GDP	-0.350**	-0.307*	-0.390***	-0.348**	-0.081	-0.172
	(0.159)	(0.149)	(0.126)	(0.126)	(0.104)	(0.104)
$log(P_t/D_t)$		-0.035***		-0.034***		0.074***
		(0.008)		(0.009)		(0.013)
R^2	0.010	0.040	0.013	0.042	0.000	0.120
Observations	1806	1806	1806	1806	1806	1806

Note: Returns and dividend growth are measured in logs. *Issuance/GDP* is implied net issuance relative to GDP smoothed by averaging over three years, from $t-3$ to t. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Regressions with country fixed effects. Country-clustered standard errors in parentheses.

the annual data. Returns and dividend growth are expressed in logs, but as before, the results in levels are very similar.

Table 3.6.3 shows that high net equity issuance robustly predicts low future real and excess returns, one and five years ahead, but it does not predict high future cash-flows. A 1% of GDP increase in net equity issuance signals 1 percentage point lower returns one year ahead, and 2 percentage points lower returns five years ahead.[18] The return coefficients on the issuance to GDP ratio remain significant once the price-dividend ratio is added to the regression, but unlike the market cap regression in Table 3.6.1, the dividend-price ratio retains its predictive power. This confirms our prior that these variables measure two different things, quantities and prices, both of which help predict future returns. The strength of the market cap to GDP ratio is that it combines these two metrics.

Why could GDP be a better measure of fundamentals than dividends, and capitalization – inclusive of issuance – a better measure of valuations than prices? Unlike dividends, GDP is relatively unaffected by time variation in payout policy such as dividend smoothing and the use of stock buybacks as implicit dividend payouts. Chen,

18. For one year, Table 3.6.3 suggests a 1*0.78 = 0.8% relative increase in gross real return, from 1.04 to 1.05 p.a.

Da, and Priestley (2012) and Chen (2009) show that, respectively, dividend smoothing and unstable corporate policy can adversely affect the empirical performance of return predictability regressions. Turning to quantities, time variation in net equity issuance should better capture changes in investor sentiment than information on prices alone. Baker and Wurgler (2000) show that if firms time the market to issue equity when investor sentiment is elevated, periods of high equity valuations and low risk premia should be accompanied by high issuance activity. Baker and Wurgler (2000) also show that the equity share in total issues is a powerful predictor of returns in US data. Greenwood and Hanson (2013) argue that times of elevated sentiment also open the market up to poorer quality issuers. Deteriorating issuer quality can help explain why, conditional on price valuations, high market capitalization forecasts low, rather than high dividend growth.

We have shown that on average, high market capitalization is followed by low subsequent returns. Next we turn our attention to the more extreme tails of the return and market cap distribution, and examine whether sharp run-ups in market cap are followed by equally sharp corrections, and whether they share other characteristics with stock market bubbles.

3.6.2 Equity bubbles and crashes

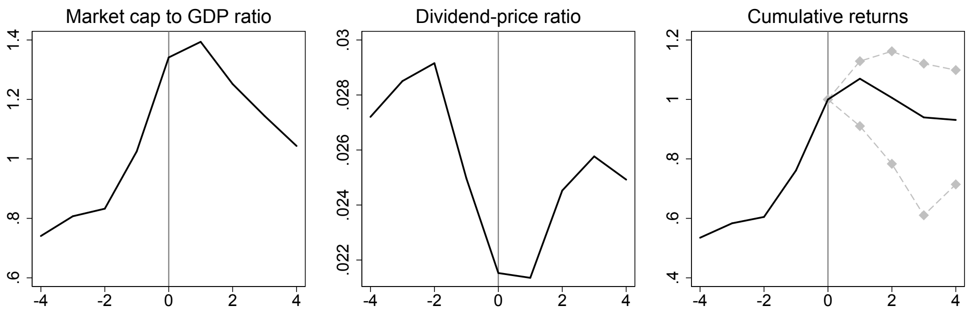
To test whether changes in market capitalization can inform us about stock market bubbles and crashes, we investigate how the equity market behaves during and after sharp increases in stock market cap. To define a "sharp" increase, we follow existing literature, and in particular the Greenwood, Shleifer, and You (2018) definition of sector-specific equity market bubbles. Greenwood, Shleifer, and You (2018) define equity market run-ups as sharp increases in real equity returns over two years, that follow persistently high returns over a longer time period. In a similar vein, we define a market cap run-up as a 35% GDP or more increase in market cap over 2 years, that follows a cumulative increase of at least 17.5% GDP over 5 years. Adding the 5-year growth assumption allows us to focus on run-ups and exclude recoveries from temporarily low market cap levels. This definition gives us roughly the same number of run-ups as the Greenwood, Shleifer, and You (2018) definition that is based on growth in the total return index.[19]

Figure 3.6.1a shows the trends in stock capitalization, dividend-price ratios and stock returns during and after such run-ups in market cap. The typical market cap increase during the run-up is around 60% of GDP (Figure 3.6.1a left-hand panel).

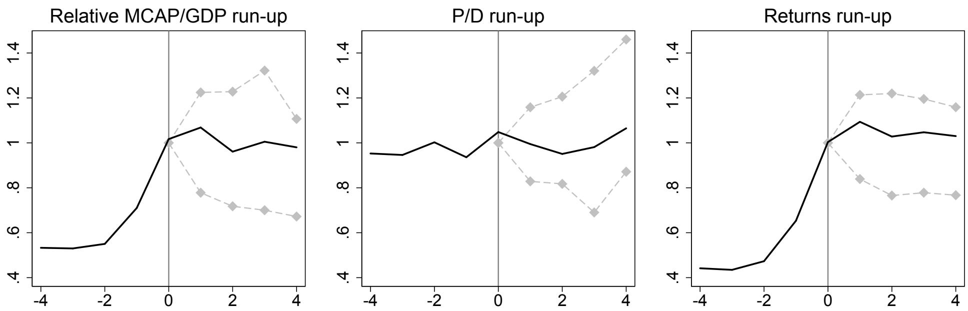
19. Greenwood, Shleifer, and You (2018) define a stock market run-up as 100% growth in the sector specific stock return index over two years, and at least 50% growth over 5 years. Applied to our aggregate data, this gives us around 30 run-up observations, a number that we target with the thresholds for increases in market cap. Because we apply the Greenwood, Shleifer, and You (2018) definition to the aggregate index, rather than sector-specific returns, this, given the cross-sector diversification gains, makes our definition somewhat more conservative.

Figure 3.6.1. Market capitalization and stock market bubbles

(a) Stock valuations and returns around sharp increases in market cap



(b) Stock returns around run-ups in alternative valuation measures



Note: Average market cap to GDP, dividend-price ratio and cumulative real return during and after stock market run-ups. Panel (a): run-up defined as a 35% GDP or higher increase in market cap over 2 years ($t = -2$ to $t = 0$), and 17.5% GDP or higher increase over 5 years ($t = -5$ to $t = 0$). Panel (b), left-hand graph: run-up defined as a doubling of market cap to GDP over 2 years, and 50% relative increase over 5 years. Panel (b), middle graph: run-up defined as a log price-dividend increase of 0.7 or more over 2 years, and 0.35 or more over 5 years. Panel (b), right-hand graph: run-up defined as a cumulative real total return of 100% or more over 2 years, and 50% or more over 5 years. Returns and relative market cap indexed to 1 at $t = 0$.

The middle panel of Figure 3.6.1a shows that a run-up in market cap is accompanied by increases in more conventional measures of stock valuations: the dividend-price ratio, on average, falls by almost a percentage point (one-quarter of its long-run mean) during these episodes. Stock returns (right-hand panel) are also high. The boom, however, quickly runs out of steam and is typically followed by a sharp correction in market cap. After the sharp increase, stock market cap falls on average by 40% of GDP, dividend-price ratios increase, and real returns are on average negative over the subsequent four years. The average geometric equity return in the sample is around 4.5% p.a., so a 10% cumulative fall in the four-year return represents an almost 30 percentage point drop relative to the counterfactual of mean return growth ($4.5 \times 4 + 10$). The low average real returns also come with a higher risk of equity market crashes. The grey lines in Figure 3.6.1a show the 75th and 25th percentile of cumulative equity returns after the run-up in stock market cap. Within

Table 3.6.4. Predicting equity market crashes

	(1)	(2)	(3)	(4)
$log(MCAP_{t-1}/GDP_{t-1})$	0.45***		0.71***	0.50***
	(0.14)		(0.13)	(0.16)
$\Delta_3\ log(MCAP_{t-1}/GDP_{t-1})$		1.06***	0.73**	0.57
		(0.32)	(0.29)	(0.35)
$log(D_{t-1}/P_{t-1})$				-0.78***
				(0.17)
Country fixed effects			✓	✓
ROC	0.62	0.64	0.71	0.71
Number of Crashes	125	123	123	123
Observations	1939	1862	1862	1807

Note: Dependent variable is the equity market crash dummy at time *t*. All episodes with real equity returns falling by more than 25% in one year or within a two year window, and with no crashes in the two previous years are dated as crashes. Logit coefficient estimates with country clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

the one-quarter of the worst outcomes, returns fall by as much as 40% in cumulative terms over 4 years.

Figure 3.6.1b shows the evolution of equity returns under alternative definitions of stock market run-ups. The left-hand panel looks at *relative* increases in stock market cap – i.e. a doubling of the market cap to GDP ratio over two years, and at least a 50% increase over five years. This places less emphasis on the events where market cap was already high before the run-up – such as the dot-com boom – and a greater emphasis on increases from low values of stock market cap. The returns follow a similar pattern to Figure 3.6.1a, even though the average fall is not quite as pronounced. The middle panel shows returns after steep increases in the log price-dividend ratio (a doubling of the price-dividend ratio over two years and at least a 50% increase over five years). The real returns are flat, but do not generally come with overly high tail risk. The comparison of market cap and price-dividend ratio run-ups echoes the findings of Section 3.6.1: in general, market cap seems to be a better measure of discount rates, and stock market over- or under-valuation than the dividend-price ratio. Finally, Figure 3.6.1b right-hand panel focuses on run-ups in real returns, with the definition from Greenwood, Shleifer, and You (2018): a 100% increase in real returns over 2 years, and at least a 50% increase over 5 years. The results are similar to market cap run-ups, although again, the bubble aftermath is slightly less severe, with on average flat rather than declining returns, and slightly lower losses in the "market crash" tail.

Table 3.6.4 formally tests whether high levels, or growth in market cap signals a higher probability of a market crash. We define an equity market crash as a return realisation in the bottom 5th percentile of the return distribution, for one and 2 year ahead returns, which gives us a threshold of -25%.[20] To assess the link between market capitalization and equity crash risk, we estimate the following logit model:

$$Prob(C_{i,t} = 1) = \Lambda\left(MCAP_{i,t-1}/GDP_{i,t-1}, X_{i,t-1}, \beta\right), \tag{3.6.3}$$

where C is the equity crash dummy, X are other predictors, β is the estimated coefficient vector, Λ is the logistic distribution function, and i and t are country and time indices. Table 3.6.4 reports the estimated β coefficients and standard errors. Consistent with the stylised facts in Figure 3.6.1, high market cap to GDP ratios (column 1), or high growth in market cap (column 2) predict a heightened probability of a crash. These results hold when controlling for country fixed effects (column 3), and the dividend-price ratio (column 4). The ROC of 0.65–0.7 compares the predictive performance of our regression with a random sorting into crash and non-crash observations, and shows that our model does substantially better than the naive prediction (ROC of 0.5).[21] Appendix Table C.3 shows that high, or growing, stock market cap predicts crash risk across different time periods, when controlling for credit growth, and for a range of alternative crash definitions.

Sharp increases in market capitalization are signs of brewing trouble in the equity market. Run-ups in the market cap to GDP ratio are, on average, followed by sharp reversals in risk premiums and valuations, and a higher risk of an equity market crash. These risks are also borne out in our aggregate market capitalization trends in Figure 3.3.1. The long-run structural increase in capitalization during the big bang has been accompanied by several boom-bust cycles, including the dot-com boom of the 1990s and the Global Financial Crisis, with rapid run-ups in market cap followed up by market crashes and reversals to the structurally higher long-run mean. Just like the big bang, these booms and busts have typically occurred in most of the countries in our sample, suggesting an increasingly important role for time variation in the global equity premium during the recent decades (Jordà, Schularick, Taylor, and Ward, 2018).

3.7 Conclusion

This paper has presented a new dataset of annual stock market capitalization in 17 advanced economies from 1870 to today. Exploring the trends in the data, and their co-movement with various financial and economic variables has revealed several surprising facts. First, the historical evolution of stock market cap resembles a hockey stick: the market cap to GDP ratio was roughly flat up until the 1980s, at

20. We set the crash dummy $C_{i,t}$ equal to 1 if there is a -25% equity return in year t, or a -25% cumulative equity return in years t and $t+1$. Appendix Table 3.6.4 evaluates the results under a number of alternative crash definitions, and finds them more or less unchanged.

21. For further details on the application of ROC curves to the financial extreme event analysis, see Schularick and Taylor (2012).

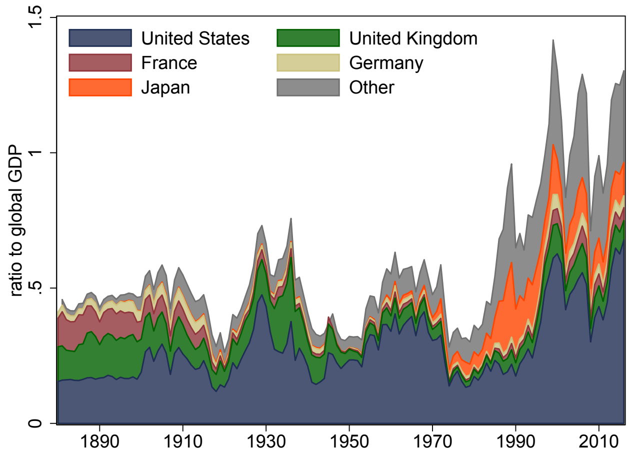
which point it expanded sharply, and remains high today. We term this sudden and unprecedented expansion in stock market size the "big bang".

At the same time, the forces that underly this structural shift in stock market size pose challenges for equating stock market growth with financial development. For the most part, changes in market capitalization are driven by shifting stock valuations, with much of these valuation shifts in turn driven by changing equity risk premia. In one sense, this explanation is somewhat unsatisfactory: it attributes even long-run changes in stock market size to a "dark matter" in stock valuations that is unrelated not only to market access and financial development, but also to future stock market fundamentals. In another sense, it is revealing because it tells us that changes in listed equity wealth, including the big bang, have little to do with changes in capital market entry, stock market efficiency, or physical accumulation of corporate stocks. Consistent with this "risk premium view" of stock market wealth, we find evidence that high levels of market capitalization predict low subsequent equity returns, and a heightened risk of stock market crashes.

Appendix 3.A The Big Bang: Additional Results

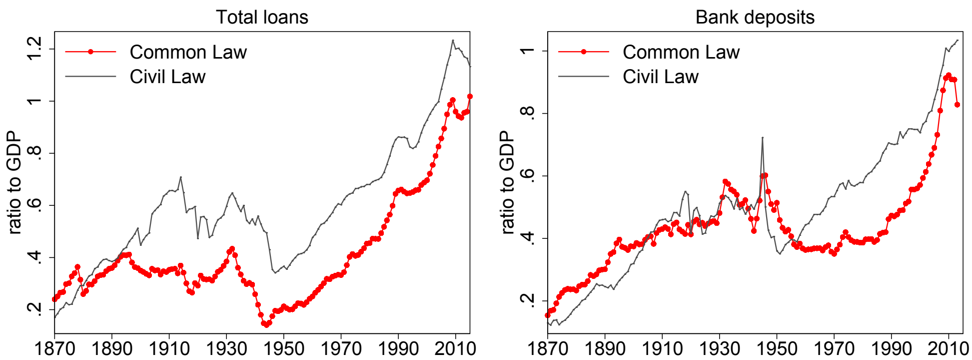
3.A.1 Trends in market capitalization

Figure 3.A.1. World market capitalization



Notes: The ratio of global market capitalization to global GDP. Global variables are the sum of the 17 countries in our sample, converted to US dollars. Missing values are interpolated to maintain sample consistency. Country shares correspond to the US dollar value of the specific country's stock market relative to global GDP.

Figure 3.A.2. Loans, deposits and legal norms

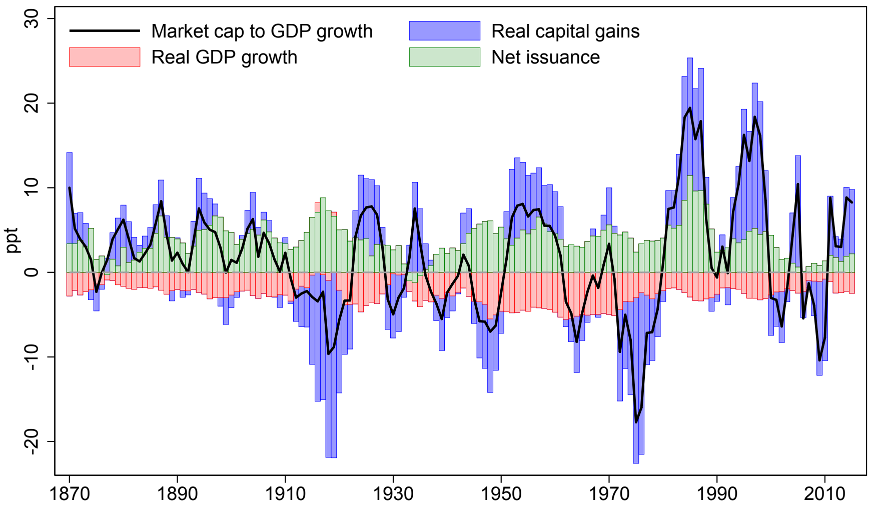


Notes: The ratio of total loans and bank deposits to GDP by country group, unweighted averages. Common law countries are Australia, Canada, the UK and the US. Civil law countries are all other countries in our dataset.

3.A.2 Drivers of stock valuations

3.A.2.1 Decomposition into capital gains, net issuance and GDP growth

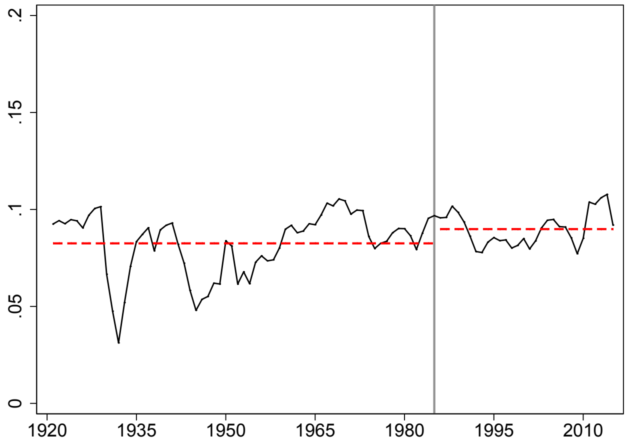
Figure B.1. Decomposition Trends with 3 Components



Notes: Decomposition of annual stock market cap to GDP growth into issuances, real capital gains and real GDP growth, using equation (3.4.5). Five-year moving averages. Market cap growth is the change in the log of market cap to GDP ratio. Implied issuance is the change in market cap not explained by equity prices or GDP growth. Using log growth rates creates a small approximation residual.

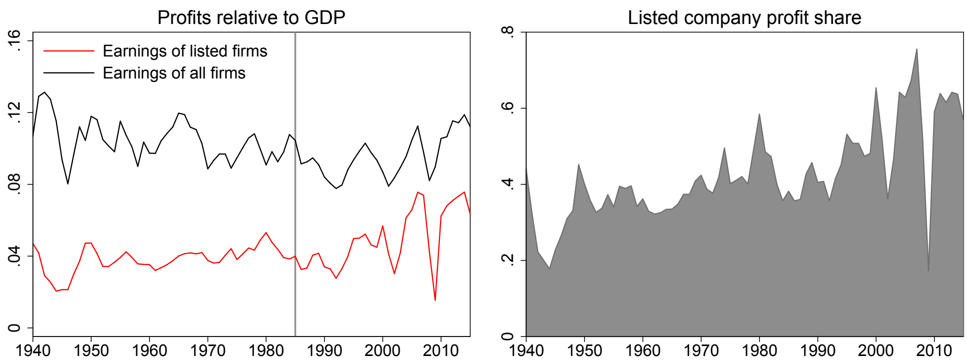
3.A.2.2 Long-run trends in equity cashflows

Figure B.2. Post-tax profits relative to GDP



Note: Unweighted average of four countries. Black vertical line indicates the start of the big bang in 1985. Dashed horizontal lines show the average of the series before and after the big bang.

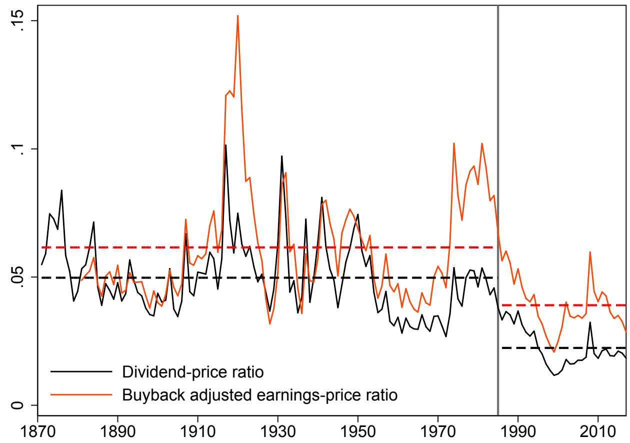
Figure B.3. Earnings share of listed companies in the United States



Note: Earnings of listed companies compared to earnings of all companies from national accounts. Earnings of listed companies relative to GDP is calculated by combining the average of monthly earnings figures and end of year price data from Shiller (2015) with our end of year market cap to GDP estimate.

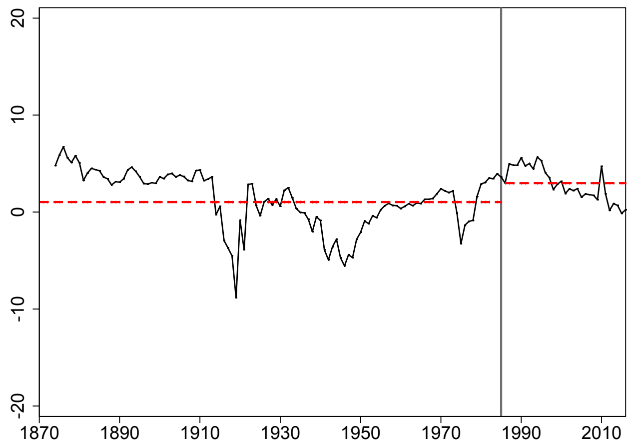
3.A.2.3 Long-run trends in equity discount rates

Figure B.4. Dividend-price and buyback adjusted earnings-price ratio in the United States



Note: Dividend-price ratio and cyclically adjusted total return earnings-price ratio (inverse of P/E10 CAPE) from Shiller (2015). December values. Black vertical line in 1985 signifies the big bang. Dashed horizontal lines show the averages of the series before and after the big bang.

Figure B.5. A proxy for the ex ante real interest rate



Note: Real interest rate calculated by substracting predicted inflation from a long-term government bond yield. Inflation is predicted using last years inflation, two lags of GDP growth and changes in inflation in a cross-country fixed effects regression. To account for monetary regimes, we split the sample into four periods 1870-1913, 1913-1945, 1946-1973, 1974-2015 and estimate the inflation expectation for each period separately. Black vertical line in 1985 signifies the big bang. Dashed horizontal lines show the averages of the series before and after the big bang.

3.A.2.4 Quantifying the contribution of different drivers

Table B.1. Quantifying the contribution of market cap determinants, post-1985 sample

	Levels		5-Year Changes		10-Year Changes	
	(1)	(2)	(3)	(4)	(5)	(6)
D_t/GDP_t	32.94***	32.45***	37.63***	37.38***	34.86***	36.00***
	(3.0)	(3.0)	(6.3)	(6.5)	(9.5)	(7.7)
D_{t+1}/GDP_{t+1}	-0.03	-0.52	3.73	3.58	-0.52	-0.96
	(1.9)	(1.8)	(3.1)	(3.3)	(3.4)	(2.4)
D_t/P_t	-18.55***	-18.73***	-11.31***	-11.71***	-6.78**	-9.38***
	(3.1)	(3.1)	(2.8)	(3.2)	(2.6)	(2.9)
r_t	-0.14	0.24	-0.89*	-0.90*	-1.11*	-1.59**
	(0.4)	(0.3)	(0.4)	(0.5)	(0.6)	(0.7)
τ_t^{corp}		-0.47**		0.14		0.69
		(0.2)		(0.2)		(0.4)
τ_{t+1}^{corp}		0.13		0.17		-1.10**
		(0.3)		(0.4)		(0.5)
R^2	0.743	0.746	0.513	0.518	0.463	0.493
Observations	474	472	410	396	328	263

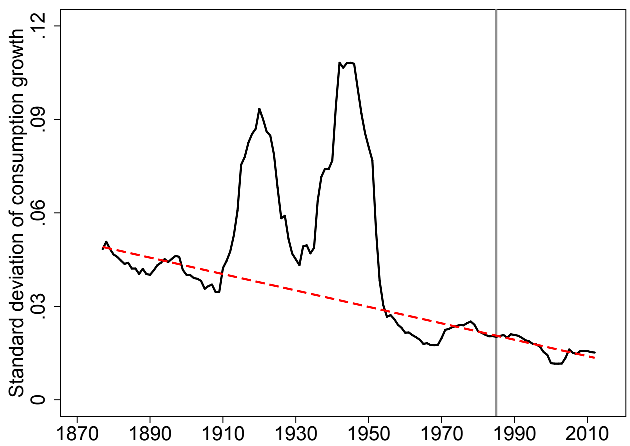
Note: Regressions with country fixed effects and robust standard errors. Standard errors in parentheses. Columns (1) and (2) show regression coefficients of dividends to GDP, the dividend to price ratio and corporate tax rates on the market capitalization level. Column (3)–(5) report the regression coefficients of the analysis in five year and 10 year changes. For the five-year and ten-year change regressions, the one-year ahead variables such as D_{t+1}/GDP_{t+1} become five-year ahead variables, i.e. $D_{t+5}/GDP_{t+5} - D_t/GDP_t$. Sample limited to post-1985 only. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Table B.2. Quantifying the contribution of market cap determinants, ex ante real rate proxy

	Levels		5-Year Changes		10-Year Changes	
	(1)	(2)	(3)	(4)	(5)	(6)
D_t/GDP_t	24.84***	25.27***	22.69***	28.03***	23.47***	23.63***
	(2.3)	(2.8)	(2.4)	(5.6)	(1.9)	(3.9)
D_{t+1}/GDP_{t+1}	1.57	0.83	0.01	1.53	2.73	-1.18
	(1.8)	(1.7)	(1.1)	(2.4)	(1.6)	(2.3)
D_t/P_t	-9.09***	-9.81***	-6.46***	-7.90***	-5.35***	-5.84***
	(1.3)	(1.9)	(1.2)	(1.7)	(1.0)	(1.4)
r_t^{pred}	-0.00	-0.68*	-1.07***	-1.75**	-1.37***	-3.04***
	(0.3)	(0.3)	(0.3)	(0.7)	(0.3)	(1.0)
τ_t^{corp}		-0.03		0.18*		0.17
		(0.2)		(0.1)		(0.2)
τ_{t+1}^{corp}		-0.03		0.11		-0.38**
		(0.2)		(0.1)		(0.2)
R^2	0.740	0.718	0.495	0.494	0.636	0.563
Observations	1986	906	1748	732	1557	561

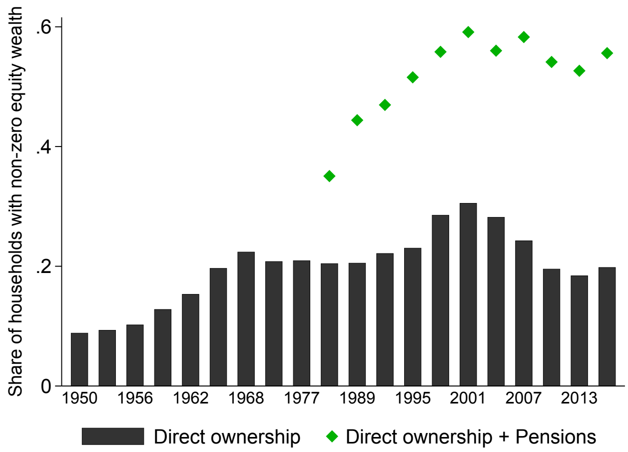
Note: Regressions with country fixed effects and robust standard errors. Standard errors in parentheses. Columns (1) and (2) show regression coefficients of dividends to GDP, the dividend to price ratio and corporate tax rates on the market capitalization level. Column (3)–(5) report the regression coefficients of the analysis in five year and 10 year changes. For the five-year and ten-year change regressions, the one-year ahead variables such as D_{t+1}/GDP_{t+1} become five-year ahead variables, i.e. $D_{t+5}/GDP_{t+5} - D_t/GDP_t$. r_t^{pred} is the real interest rate calculated by substracting predicted inflation from a long-term government bond yield. See footnote of Figure B.5 for more detail. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Figure B.6. Consumption volatility over the long run



Note: Annual standard deviation of real consumption growth over rolling decadal windows (1875 figure is the decade 1870–1880). Unweighted average of 17 countries. Black vertical line in 1985 signifies the big bang. Dashed horizontal line is the peacetime trend in the time series.

Figure B.7. Positive wealth in direct and indirect stock ownership



Note: Share of households owning stocks in the United States. Data are sourced from the Survey of Consumer Finances, kindly shared with us by Kuhn, Schularick, and Steins (2017). Direct ownership includes all households with positive stock or mutual fund wealth. The estimate of direct ownership + pensions includes all households with positive pension assets, stock or mutual fund wealth.

3.A.3 Predicting equity returns and market crashes

3.A.3.1 Predicting returns and dividend growth

Table C.1. Return predictability at long horizons and during the big bang

	Real returns		Excess returns		Real dividend growth	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: *Ten-year ahead average returns and dividend growth*						
$log(MCAP_t/GDP_t)$	-0.063***	-0.065***	-0.044***	-0.045***	-0.003	-0.050***
	(0.010)	(0.012)	(0.007)	(0.007)	(0.011)	(0.012)
$log(P_t/D_t)$		0.008		0.003		0.158***
		(0.015)		(0.013)		(0.019)
R^2	0.126	0.127	0.073	0.073	0.000	0.275
Observations	1753	1753	1753	1753	1753	1753
Panel 2: *One-year ahead returns and dividend growth after 1985*						
$log(MCAP_t/GDP_t)$	-0.129***	-0.121***	-0.103***	-0.094***	-0.029	-0.056**
	(0.014)	(0.014)	(0.015)	(0.016)	(0.022)	(0.025)
$log(P_t/D_t)$		-0.073***		-0.090***		0.254***
		(0.017)		(0.019)		(0.054)
R^2	0.063	0.075	0.036	0.052	0.002	0.109
Observations	487	487	487	487	487	487

Note: Returns and dividend growth are measured in logs. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Regressions with country fixed effects. Country-clustered standard errors in parentheses.

Table C.2. Dividend-to-GDP as a predictor of equity returns and dividends

	Real returns		Excess returns		Real dividend growth	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: *One-year ahead returns and dividend growth*						
$log(D_t/Y_t)$	-0.016	-0.028**	-0.019**	-0.028***	-0.075***	-0.054***
	(0.010)	(0.010)	(0.008)	(0.008)	(0.018)	(0.015)
$log(P_t/D_t)$		-0.063***		-0.048***		0.106***
		(0.012)		(0.011)		(0.029)
R^2	0.002	0.020	0.004	0.015	0.036	0.069
Observations	2032	2032	1986	1986	2029	2029
Panel 2: *Five-year ahead average returns and dividend growth*						
$log(P_t/Y_t)$	-0.062***	-0.060***	-0.036***	-0.034***	-0.029***	-0.049***
	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)
$log(P_t/D_t)$		-0.010		-0.013		0.099***
		(0.013)		(0.010)		(0.016)
R^2	0.029	0.029	0.067	0.071	0.032	0.206
Observations	2162	2111	2048	1997	2095	2095

Note: Returns and dividend growth are measured in logs. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Regressions with country fixed effects. Country-clustered standard errors in parentheses.

3.A.3.2 Predicting equity market crashes

Table C.3. Predicting equity market crashes: alternative specifications

	(1) Pre 1945	(2) Post 1945	(3) Post 1985	(4) War Obs.	(5) Credit Growth
$log(MCAP_{t-1}/GDP_{t-1})$	3.12***	0.64***	1.47***	0.70***	0.75***
	(0.92)	(0.14)	(0.34)	(0.12)	(0.10)
Δ_3 $log(MCAP_{t-1}/GDP_{t-1})$	1.46**	0.52**	1.26***	0.67***	0.64***
	(0.60)	(0.25)	(0.33)	(0.25)	(0.24)
Country fixed effects	✓	✓	✓	✓	✓
ROC	0.79	0.69	0.78	0.69	0.74
Number of Crashes	25	98	53	143	118
Observations	571	1178	544	2048	1895
	(1) Decade	(2) Large Crashes	(3) 1-year Crashes	(4) 3-year Crashes	(5) MCAP Crashes
$log(MCAP_{t-1}/GDP_{t-1})$	0.61***	1.00***	0.70***	0.86***	0.50***
	(0.22)	(0.22)	(0.13)	(0.11)	(0.10)
Δ_3 $log(MCAP_{t-1}/GDP_{t-1})$	0.87***	1.26**	-0.08	1.23***	0.95***
	(0.29)	(0.56)	(0.18)	(0.40)	(0.26)
Country fixed effects	✓	✓	✓	✓	✓
ROC	0.78	0.79	0.68	0.75	0.69
Number of Crashes	123	28	92	104	149
Observations	2008	1643	1862	1862	1862

Note: Dependent variable is the equity market crash dummy at time *t*. In Panel 1 and Panel 2 column 1, a crash is defined as real equity returns falling by more than 25% in one year or within a two year window, and with no crashes in the two previous years. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Standard errors in parentheses. All estimates are based on logit estimations with country fixed effects and country clustered standard errors. Panel 1: Column (1) restricts the panel to observations before 1945. Column (2) and (3) only include observations after 1945 and 1985 respectively. Column (4) adds observations from the world wars and Column (5) includes five lags of real private per capita credit growth as additional controls. Panel 2: Column (1) reports estimates with decade fixed effects and Column (2) to (5) are based on alternative crash definitions. Large crashes are all crashes with a 50% fall in real equity returns either in the first year or within a two year window. 1-year crashes are all episodes with a 25% fall of equity prices in one year and 3-year crashes are based on a three year window. MCAP Crashes uses market capitalization to GDP instead of real equity returns to date crashes.

Appendix 3.B The Big Bang: Data Appendix

This section details the sources of our market capitalization estimates for each country, and compares them to alternative estimates. The alternative estimates are generally country specific, but we always compare our data to those of Goldsmith (1985) (sourced from La Porta, Lopez-de-Silanes, and Shleifer, 2008) and Rajan and Zingales (2003) when available. All the annual estimates reflect end-of-year values, unless otherwise stated.

Australia

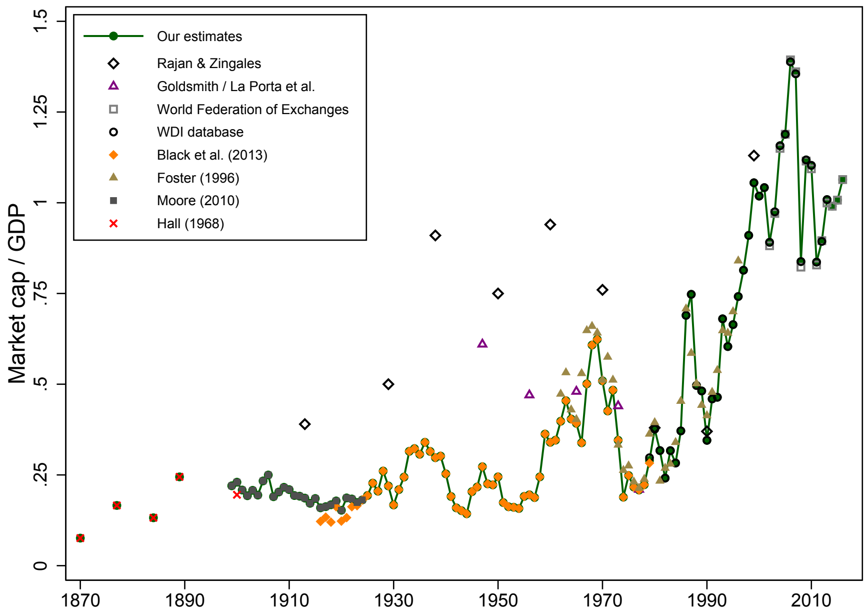
Table D.1 documents the sources of our stock market capitalization data for Australia, and Figure D.1 plots the resulting series alongside alternative existing estimates. The Australian securities market has generally been dominated by two major stock exchanges, located in Sydney and Melbourne. Hall (1968) argued that the Melbourne stock exchange was dominant in the late 19th century, largely because of large capitalizations of stocks of mining companies, and the data in Black et al. (2013) and Lamberton (1958) suggest that the Sydney stock exchange became dominant in the early 20th century. Based on this, we use the Hall (1968) estimates of the Melbourne stock market capitalization for the 19th century data, and switch to the Sydney exchange in the 20th century, using estimates of Moore (2010b) and Black et al. (2013), which are also consistent with the RBA Historical Statistics data in Foster (1996). From the 1970s onwards we switch to the total Australian firm capitalization estimates provided by the *World Federation of Exchanges* reports and the World Bank *WDI database*.

The main potential bias in the data for Australia comes from two sources: the fact that until the 1970s, we only have data for either the Sydney or the Melbourne exchange, not both; and the fact that these data include both foreign and domestic companies (again, up to the 1970s). These two biases do, however, largely seem to

Table D.1. Data sources: Australia

Year	Data source
1870–1889	Total capitalization of the Melbourne Stock Exchange, from Hall (1968)
1899–1924	Total capitalization of the Sydney Stock Exchange, from Moore (2010b). Converted to AUD using the exchange rates in Jordà, Schularick, and Taylor (2017).
1925–1978	Total capitalization of the Sydney Stock Exchange, from Black, Kirkwood, Williams, and Rai (2013).
1979–2013	Total capitalization of all Australian listed firms, shares listed on Australian exchanges. Source: World Bank *WDI database*. Almost identical to the Sydney cap in the 1970s; spliced with Black et al. (2013) data in 1979.
2014–2016	Total capitalization of all Australian listed firms, shares on Australian exchanges. Source: World Federation of Exchanges (WFE) *Statistical Reports*, various years.

Figure D.1. Australia: alternative stock market cap estimates



balance each other out: the total Australian exchange capitalization in the 1970s is very similar to that of the Sydney stock exchange, and Lamberton (1958) indicates that the Sydney stock exchange became the most important center for financial activity much earlier. Therefore we do not make any further adjustments to the early Australian data, which focus mostly on the Sydeny exchange, including both domestic and foreign companies.[22]

Our approach of focussing on the Melbourne cap in the late 19th century, and the Sydney cap in the 20th century is in line with that of Rajan and Zingales (2003). As Figure D.1 shows, however, our estimates of market capitalization are somewhat below those of both Rajan and Zingales (2003) and Goldsmith (1985), largely due to better available up-to-date statistics, for example from Black et al. (2013) and Moore (2010b).

We are grateful to the Reserve Bank of Australia and Anna Nietschke for sharing the data from Black, Kirkwood, Williams, and Rai (2013) with us, and providing other helpful references.

22. As a side note, adding up the Hall (1968) and Moore (2010b) estimates for 1899 would grossly overestimate the total cap of Australian firms because it does not adjust for cross-listings.

Belgium

Table D.2. Data sources: Belgium

Year	Data source
1870–2015	Total capitalization of all Belgian companies on the Brussels Stock exchange, SCOB Database. Data shared by Frans Buelens. See Annaert, Buelens, and De Ceuster (2012) for details.
2016–2017	Extrapolated forward using the cap of all Belgian companies listed in Belgium, from the *ECB Statistical Data Warehouse*, Security issues statistics. The ECB and SCOB data are in general very similar, but we use the SCOB data as the benchmark for greater overall consistency.

Figure D.2. Belgium: alternative stock market cap estimates

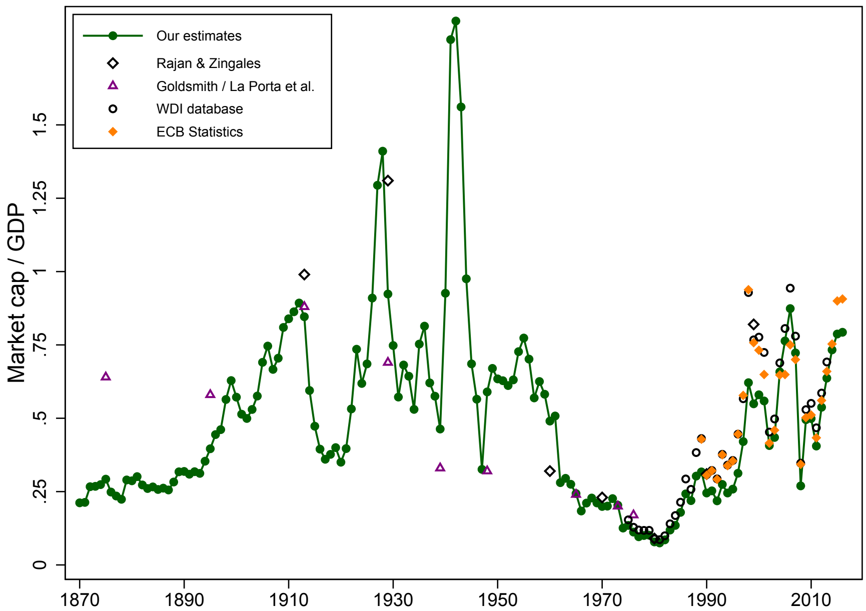


Table D.2 documents the sources of our stock market capitalization data for Belgium, and Figure D.2 plots the resulting series alongside alternative existing estimates. The data cover the Brussels stock exchange, which was the dominant stock exchange throughout the data coverage period in our paper, and are sourced from the security-level SCOB database (see Annaert, Buelens, and De Ceuster, 2012, for the description). The data cover all companies with main economic activities in Belgium, that are listed on the Brussels stock exchange. Unlike other existing estimates, the capitalization is aggregated up from security-level data for each year, and does not rely on estimation or extrapolation. For the modern period, the SCOB estimates are similar to other commonly used sources such as the *WDI* database and the *ECB Statistical Data Warehouse* data. We are grateful to Frans Buelens for sharing the SCOB market capitalization data with us.

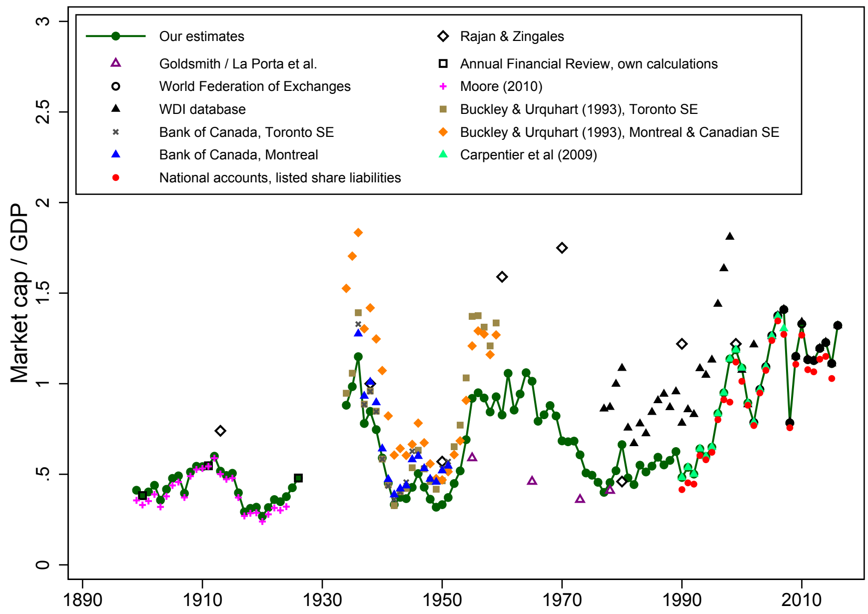
Canada

Table D.3. Data sources: Canada

Year	Data source
1899–1926	Capitalization of all Canadian firms listed on foreign exchanges. Baseline data from Moore (2010b), scaled up using own calculations from microdata in the *Annual Financial Review* in years 1900, 1911 and 1926. The scaling accounts for firms missing from the listings in Moore (2010b) data, and exclusion of foreign firms. Market cap growth for 1924–1926 estimated using the change in the share price index and assumed net issuance of 1.2% of market cap (the average of observed issuance for 1937–2016, using data from the Bank of Canada *Statistical Summaries* (*Financial Supplement*, years 1964–1969) and *Banking and Financial Statistics* database.
1934–1958	Combined capitalization of the Toronto, Montreal and Canadian Stock Exchanges from Buckley and Urquhart (1993), scaled down to exclude cross-listings, foreign shares and preference shares, in order to match the market capitalization estimate for 1959. The scaling ratio between unadjusted and adjusted capitalization (3:1) is similar to the one obtained by Carpentier, L'Her, and Suret (2009) for year 1990.
1959–1969	1970 capitalization extrapolated back using growth in share prices and net issuance of ordinary shares. Share price data from Kuvshinov (2018), issuance data from Bank of Canada *Statistical Summaries*.
1970–1974	1975 capitalization extrapolated back using the growth in market value of equity liabilities of Canadian firms, from the Bank of Canada *CANSIM* database, national balance sheet data.
1975–1989	1990 capitalization extrapolated back using growth in share prices and net issuance of ordinary shares. Share price data from Kuvshinov (2018), issuance data from the Bank of Canada *Banking and Financial Statistics*.
1990–2001	Total capitalization of all Canadian firms listed in Canada, adjusted for cross-listings, from Carpentier, L'Her, and Suret (2009).
2002–2016	Total capitalization of all Canadian listed firms, shares listed on all Canadian exchanges, adjusted for cross-listings, from the World Federation of Exchanges (WFE) *Statistical Reports*, various years

Table D.3 documents the sources of our stock market capitalization data for Canada, and Figure D.3 plots the resulting series alongside alternative existing estimates. Constructing historical market capitalization estimates is especially challenging in the case of Canada, for several reasons. First, throughout the whole of our sample period, Canada has operated at least two large and active stock exchanges, in Toronto and Montreal. The capitalizations of these two exchanges have tended to quite similar, with Montreal slightly larger in the early historical period, and Toronto – in the latter. Many available statistics provide the gross total value of securities listed on each exchange. But most large companies were listed on both of these stock exchanges, which makes adjusting gross estimates for cross-listings especially important. Even in the modern data, including the estimates of Rajan and Zingales (2003) and the World Federation of Exchanges, the total Canadian capitalization was not

Figure D.3. Canada: alternative stock market cap estimates



adjusted for cross listings until year 2002, such that the totals often double-counted the shares of large cross-listed firms (Carpentier, L'Her, and Suret, 2009). Second, the Canadian industry and financial markets were internationally integrated with the US and UK due to geographical proximity and colonial-era ties. This makes the exclusion of foreign listings from calculations important. Further, a few Canadian firms were only listed on US exchanges or in London, meaning that they should be excluded from our data. Third, many statistics group together all "stocks" issued by Canadian firms, which include both ordinary and preference shares, whereas we want to capture ordinary shares only.

The severity of these various measurement issues can be seen in the existing estimates of Rajan and Zingales (2003) (RZ) and Goldsmith (1985) (GS) in Figure D.3. RZ generally use a mix of unadjusted Toronto cap, unadjusted Montreal cap, the sum of the two, or an adjusted total, depending on the particular year. This results in changes in the series which seem to be mostly attributable to this variation in measurement: for example, between 1970 and 1980, RZ estimate that the market cap to GDP ratio fell by roughly four times, or 150% of GDP. At the same time, stock prices more than doubled. GS documents a small increase in the market cap to GDP ratio between 1973 and 1978. Finally, the RZ market capitalization estimates in the 1960s and 1970s are roughly three times those of GS, despite the fact that in principle, the RZ data should cover listed firms only, while GS covers both listed and unlisted firms. These biases are not easily remedied by other official statistics. Buckley and Urquhart (1993) and the Bank of Canada *Statistical Summaries* provide estimates of the capitalization of the Toronto and Montreal stock exchanges for the period 1934–1959, shown in Figure D.3. These estimates, however, are gross of cross-listed securities, foreign firms and preference shares. If we add up the Buckley and Urquhart (1993) estimates of the Toronto and Montreal capitalization in the 1930s, we get a market cap to GDP ratio of almost 400% right in the aftermath of the Great Depression, which seems implausibly high.

We seek to deal with the difficulties discussed above when constructing our own market capitalization estimates. For both the early 20th century, and the recent decades, we are able to calculate the total capitalization of Canadian listed firms, with all the necessary adjustments, with a high degree of accuracy. The baseline data for the early series come from Moore (2010b), who uses stock listings data to compute the total cross-listings-adjusted capitalization of the Toronto and Montreal stock exchanges. Nonetheless, these data include foreign firms, and might not include securities of smaller companies or those listed on unofficial or curb exchanges. Given that the Moore (2010b) estimates for the 1920s are so far below those of Buckley and Urquhart (1993) and Bank of Canada *Statistical Summaries* in the 1930s, and the fact that stock price aprpeciation between late 1920s and early 1930s in Canada was very small due to the Great Depression, we construct our own estimates for the early period which enable us to benchmark the Moore (2010b) data.

Our benchmark-year estimates for the early period are constructed from the microdata on individual companies in the *Annual Financial Review* publication for years 1901, 1912 and 1927.[23] Because the *Annual Financial Review* only has each company enter once, this effectively adjusts for any cross listings. In addition, these data contain information on company headquarters and operations, as well as which exchanges the firm is listed on, allowing us to control for factors such as foreign ownership. For the purpose of this calculation, we include firms incorporated and governed from Canada, but with operations overseas, such as the various Mexican tramway companies which appear in the 1911 listing, but this has little bearing on our results. It turns out that the benchmark estimates are close to the data from Moore (2010b) (see Figure D.3): around 15–20% higher for 1900 and 1926, and similar in size for 1911, due to a high number of foreign companies on the market during that year, which we adjust out but Moore (2010b) does not. Based on this, we scale up the Moore (2010b) data slightly to match the adjusted total, and bridge the 1924–1926 gap by using share price appreciation for those years, and an assumed net issuance that equals the long-run average in Canadian data.

For the recent period, the World Federation of Exchanges provide statistics which measure the adjusted total capitalization of all Canadian firms listed in Canada for years 2002–2016. Previous years' estimates from this source include some double-

23. Capitalization data refer to the end of each respective previous calendar year, i.e. end-1900, end-1911 and end-1926.

counting, hence for the period 1990–2001 we rely on data from Carpentier, L'Her, and Suret (2009), who calculate an adjusted total market cap accounting for cross-listings and excluding foreign firms and non-equity securities. For 1990, Carpentier, L'Her, and Suret (2009) estimate total capitalization which is roughly one-third of the unadjusted sum. These data match up nicely with the national balance sheet estimates for the market value of listed equity liabilities of Canadian firms (Figure D.3, red diamonds), available from the national accounts data in the *CANSIM* database of the Bank of Canada.

We have several sources available to us for the period from 1934 to 1989: the estimates from the World Bank's *WDI Database* for the period 1975–2016, historical statistics data for 1934–1959 from Buckley and Urquhart (1993), the estimates by the Bank of Canada in their *Statistical Summaries*, which are the underlying source of the Buckley and Urquhart (1993) data, as well as the computations of Rajan and Zingales (2003) and Goldsmith (1985). We also have data on net equity issuance which cover the period 1937–2016, with a gap in 1970–1974, with historical data sourced from the Bank of Canada *Statistical Summaries Financial Supplement*, and modern data from the Bank of Canada's *Banking and Financial Statistics*, as well as share price appreciation data from Kuvshinov (2018), and national balance sheet estimates of the total equity value of listed and unlisted Canadian firms. Some of these sources are, however, likely to contain a lot of measurement error. The *WDI Database* estimates before 2002 are highly noisy and, according to Carpentier, L'Her, and Suret (2009), their underlying source – the WFE database – double- or triple-counts cross-listed securities for this period. The estimates of Rajan and Zingales (2003) switch definitions in terms of exchange coverage and are also often gross of cross-listings, as discussed earlier, while the underlying definitions of the Goldsmith (1985) data are uncertain. As seen from Figure D.3, the data for all three of these sources are also rather noisy. Based on this, we decide not to use any of these sources, and restrict ourselves to the estimates of Buckley and Urquhart (1993), Bank of Canada, and the share price and net issuance data.

For the period 1960–1989, we largely rely on the data on share prices and net issuance. We extrapolate back the 1990 estimate using share price growth, and subtracting each year's net issuance, inflated at half the year's share price appreciation. The trend is similar to that obtainable from the WDI data during the 1970s and 1980s, when the growth trend in the WDI data seems reasonably accurate, and definition of the series – consistent from year to year. This gives us confidence that our data track the underlying evolution of adjusted Canadian stock market cap during this time period. For years 1970–1974, we do not have net issuance data, and use the year-on-year changes in the market value of Canadian firms' equity liabilities instead, which implicitly assumes a constant proportion of listed firms, and similar price changes between listed and unlisted equities. Given the short time period under consideration, and the fact that unlisted firm equity price changes tend to be estimated from those of listed firms, the error resulting from this extrapolation should be small.

For the period 1934–1959, we have two choices of how to use the Buckley and Urquhart (1993) (BU) and *Statistical Summaries* data. First, we can adjust the raw series to account for the extent of cross-listing, exclude foreign firms and preference shares. We can estimate each of these adjustments using the *Annual Financial Review* microdata, and data on issuance of different types of securities in the Bank of Canada *Statistical Summaries*. In total, this would adjust the BU series down by roughly a factor of 2. However, the resulting capitalization in both 1934 and 1959 would then appear too high: in 1934, too high relative to 1926 data, and in 1959, too high relative to our estimate described above.[24] In light of this, we take a different approach: we scale down the BU series to match our total market capitalization estimate for 1959, constructed by extrapolating 1990 cap back using share price and net issuance data. This results in a downward adjustment by a factor close to 3 – the ratio similar to that in the Carpentier, L'Her, and Suret (2009) adjustment for 1990, which gives us some confidence about the measurement. The resulting adjusted series are shown as the green solid line in Figure D.3.

Taken together, our estimates for Canada should go some way towards resolving the considerable uncertainty resulting from the wide range of existing estimates in Figure D.3. That being said, the severity of the potential measurement issues for Canada mean that, especially for the period 1934–1970, the series are likely to contain some measurement error.

24. Note that the stock price growth between 1934 and 1926 was close to zero due to the Great Depression, and though nominal GDP declined, the implied net issuance for 1926–1934 using this estimate would be rather large.

Denmark

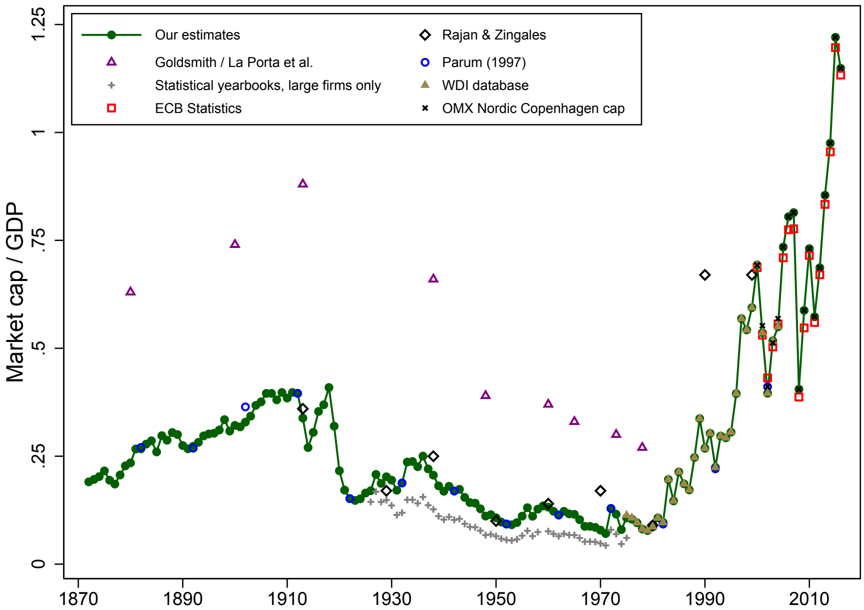
Table D.4. Data sources: Denmark

Year	Data source
1872–1899	Total market cap of all Danish firms listed in Denmark, aggregated up from individual firms' capitalization in the Green's *Dankse Fonds of Aktier* yearbooks, various years. Ordinary shares only.
1900–1925	Total market cap of all Danish firms listed in Denmark, computed as previous years' market cap * the total book cap of listed firms * market-to-book ratio of listed firms, benchmarked to Parum (1997)'s decennial market cap estimates. Book cap of listed firms estimated as book cap of all firms from Hansen and Svendsen (1968) and *Statistical yearbooks* (various years), times share of listed firms estimates from own data in 1899 and Erichsen (1902). Parum's decennial estimates sourced from Abildgren (2006).
1926–1975	Market capitalization of large listed Danish firms, scaled up to match capitalization of all firms at decennial benchmarks. Data for all firms from Parum (1997). Data for large firms are from the listings in the *Statistical yearbooks*, various years, and contain 50–60 firms for each year.
1975–2004	Total capitalization of all Danish listed firms, shares listed on Danish exchanges, from World Bank's *WDI database*. Spliced with the scaled-up capitalization of largest firms over the years 1975–1977 (the two series are very similar).
2005–2016	Total capitalization of ordinary shares on the Copenhagen stock exchange, sourced from the OMX Nordic *Yearly Nordic Statistics*.

Table D.4 documents the sources of our stock market capitalization data for Denmark, and Figure D.4 plots the resulting series alongside alternative existing estimates. Long-run estimates of the total capitalization of Danish firms for the period 1882–2002 are available from Parum (1997) and Abildgren (2006). However, these data are computed at decennial frequency only. To fill the gaps, we construct our own estimates of the total stock market capitalization of ordinary shared of listed Danish firms for each year between 1872 and 1899 using statistics on individual firms' share prices and book capital in Green's *Dankse Fonds of Aktier*. Green's yearbooks contain data on all Danish listed firms at annual frequency.

For years 1900–1925, we combine benchmark year estimates from our own microdata and Parum (1997) with statistics on share prices and book capital of listed firms. We estimate listed firms' book capital using data on total capital of all firms, available in Hansen and Svendsen (1968) up to 1914 and yearly editions of the *Statistical Yearbooks* thereafter, and estimates of the proportion of firms listed in Erichsen (1902), as well as those computed by comparing the total book capital estimates with data on share prices and market cap at benchmark years. We compute the annual change in market capitalization as the change in total book capital of all firms, times the change in the share of firms listed, times the capital appreciation in the share price index (for 1900–1914, we also compute the actual market-to-book

Figure D.4. Denmark: alternative stock market cap estimates



of listed firms, and use that instead, but the estimation gives us similar numbers to using the share index). We then adjust the growth rates of capitalization in each year to match the data at benchmark dates. The main adjustment concerns the period 1915–1922, during which the book capital of all firms nearly doubled while the book capital of listed firms remained flat, presumably following sharp delistings during the banking crisis of the early 1920s. The trend in the book capital of all firms gives us the boom-bust dynamics of high capital issuance during the book of the late 1910s, and delisting during the early 1920s, which we then rescale to match the implied larger delistings by listed firms. For years 1923–1925, very little adjustment to growth rates is necessary.

From 1926 onwards, each yearly edition of the *Statistical yearbook* publishes a summary stock listings, which includes data on capital and market-to-book of all major listed firms in Denmark. We use these data to estimate total market capitalization by scaling it up to match the total cap in Parum (1997) at decennial benchmark periods, and scaling the growth rates in-between if necessary. It turns out that the large firms in the *Statistical yaerbook* listings, which number around 50–60 in total, consistently represent around half of the total Danish market cap, and track the aggregate data very well, so very little adjustment to growth rates is necessary to match the capitalization estimates for all firms at the benchmark years.

For the recent period, market capitalization estimates for all of Denmark, or the Copenhagen stock exchange are available from the World Bank's *World Development Indicators*, ECB *Statistical Data Warehouse* and the OMX Noric *Yearly Nordic Statistics*. We use a compbination of the *WDI* and *OMX Nordic* data for our estimates, but the data are similar to the estimates of the ECB. Even though the *OMX Nordic* data in principle only cover Copenhagen, and cover foreign as well as domestic firms, in practice these numbers follow total Danish capitalization estimates almost one-for-one, and we use these data rather than the ECB statistics to avoid potential measurement error when converting the ECB data from euros to kronas.

Our estimates are substantially below those of Goldsmith (1985), with the most likely reason for the upward bias in Goldsmith (1985)'s estimates being the inclusion of unlisted equities and debt securities. Our estimates are close to those of Rajan and Zingales (2003) for the respective benchmark years.

We would like to thank Kim Abildgren for helping us locate and interpret the historical data sources for Denmark.

Finland

Table D.5. Data sources: Finland

Year	Data source
1870–1991	Total capitalization of all Finnish companies on the Helsinki Stock exchange, form Nyberg and Vaihekoski (2014a), kindly shared by Mika Vaihekoski.
1992–2017	Total capitalization of all Finnish firms, shares listed in Finland. Source: *ECB Statistical Data Warehouse*, Security issues statistics.

Figure D.5. Finland: alternative stock market cap estimates

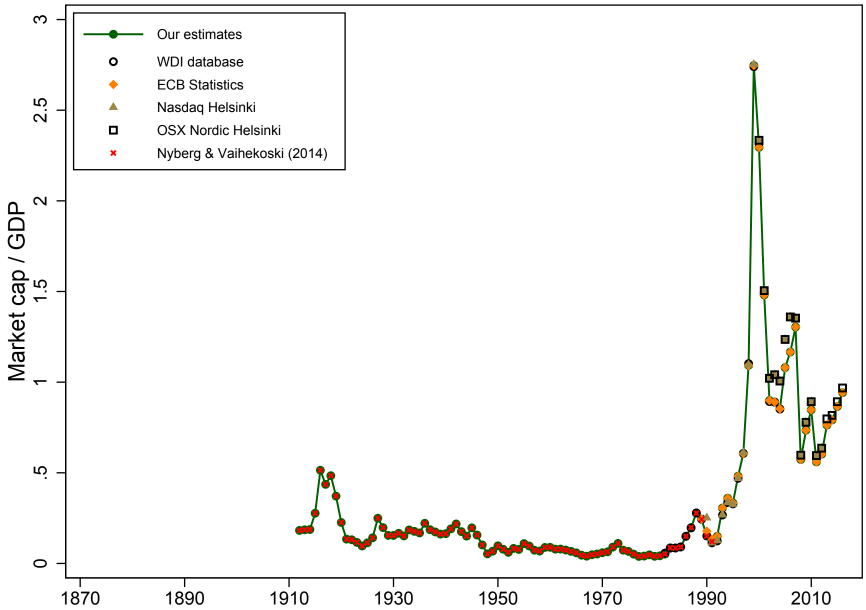


Table D.5 documents the sources of our stock market capitalization data for Finland, and Figure D.5 plots the resulting series alongside alternative existing estimates. The long-run data come from Nyberg and Vaihekoski (2014a), who have compiled a database of returns and capitalization on all stocks listed on the Helsinki exchange between its foundation in 1912 and 1991, when modern capitalization indices are available (see Nyberg and Vaihekoski, 2011; Nyberg and Vaihekoski, 2014b, for further details on the data). The Nyberg and Vaihekoski (2014a) series are aggregated up from individual share-level data, obtained from a range of historical sources, and fit the modern day series well for the overlapping period, as shown in Figure D.5. The modern data from the ECB series are very close to Helsinki stock exchange capitalization estimates from Nasdaq and OMX Nordic (Figure D.5).

We are grateful to Mika Vaihekoski for sharing data and providing help and support in locating the sources for Finland.

France

Table D.6. Data sources: France

Year	Data source
1870–1899	Stock market capitalization of the Paris stock exchange from Arbulu (1998) and Le Bris and Hautcoeur (2010), at roughly 5-year benchmarks, scaled up to proxy France total using data from Bozio (2002) (using the 1904 ratio between the Le Bris and Hautcoeur (2010) Paris series and Bozio (2002) France series as the benchmark), and year-to-year movements between the benchmark years estimated using changes in the capitalization of all French securities from Saint-Marc (1983).
1900–1988	Market capitalization of all shares of French companies listed on French stock exchanges, from Bozio (2002).
1989–2017	Total capitalization of all French firms, shares listed in France, from the *ECB Statistical Data Warehouse*, Security issues statistics.

Figure D.6. France: alternative stock market cap estimates

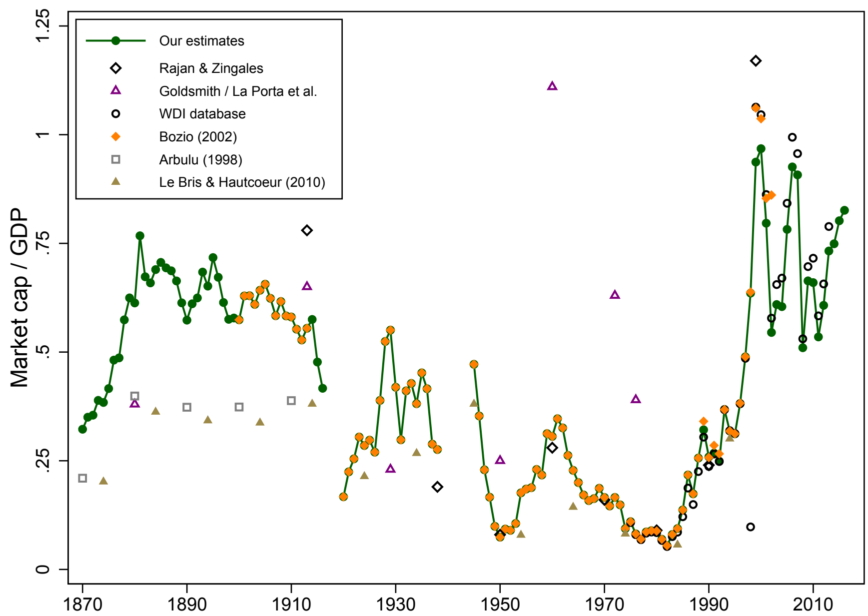


Table D.6 documents the sources of our stock market capitalization data for France, and Figure D.6 plots the resulting series alongside alternative existing estimates. Most of the data are drawn from the comprehensive study of Bozio (2002), which estimated the total capitalization of French shares listed on all French exchanges between 1900 and 2002. Between 1900 and 1963, Bozio (2002) relied on yearly capitalization of the *Cote Officielle* for the Parisian bourse, scaled up to match the total for France using data for all French stock exchanges at benchmark periods. After 1964, the Bozio (2002) data are a direct estimate of the total capitalization of all French stock exchanges. For the recent period, these data match up well with the series from the World Bank's *WDI Database*, and ECB *Statistical Data Warehouse*. We rely on the ECB data for the recent period.

For the 19th century, we make use of benchmark year estimates for the total capitalization of the Parisian bourse, from the studies of Le Bris and Hautcoeur (2010) and Arbulu (1998). We scale up these data to proxy the capitalization of all French exchanges, using the ratio of Parisian to total French market cap in year 1904. This extrapolation, therefore, implicitly assumes that the market share of regional exchanges did not change too much during the late 19th century. It is possible that the regional exchanges were somewhat more important during this early period, in which case our data would somewhat understate the total French market cap.[25] In-between the benchmark years, we use the changes in Saint-Marc (1983)'s estimates of the total capitalization of French securities, computed by scaling up capital income data, to proxy the year-to-year movements in market cap during the late 19th century.

Figure D.6 also highlights the uncertainty around earlier market capitalization estimates, especially those of Goldsmith (1985), and also to some extent the Rajan and Zingales (2003) data: on average they tend to overstate the French stock market capitalization, perhaps by including securities which are not common stocks, which can often be the case with national balance sheet estimates such as those of Goldsmith (1985), or foreign securities. In the early 1960s, the Goldsmith (1985) market capitalization estimate is almost 5 times the size of the Rajan and Zingales (2003) estimate, with our estimate, derived from Bozio (2002) data in-between these two, but closer to those of Rajan and Zingales (2003).

We are grateful to Antoine Bozio for providing help in understanding the various sources for the French market capitalization data.

25. Bozio (2002)'s estimates suggest that the relative importance of the Parisian stock exchange increased slightly between 1900 and 1913, remained roughly unchanged between 1913 and 1938, and spiked again after World War 2.

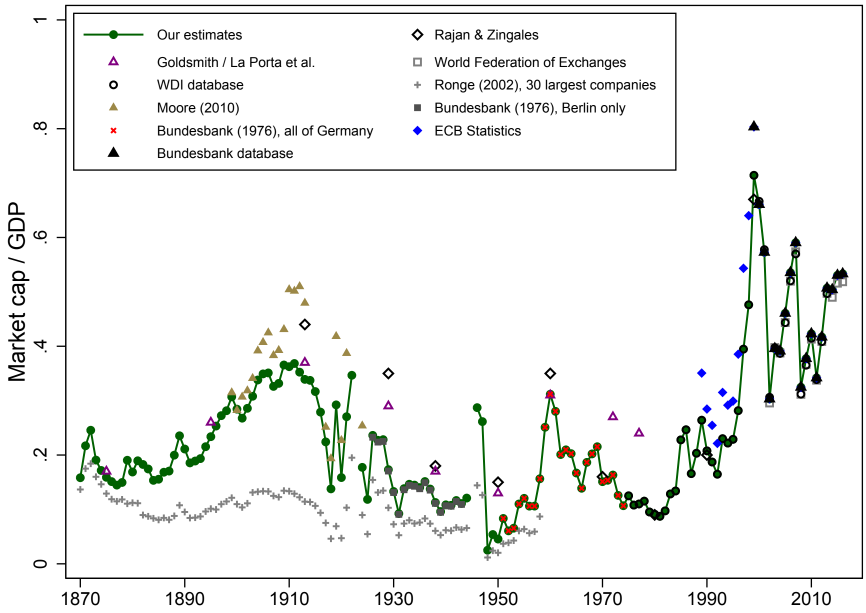
Germany

Table D.7. Data sources: Germany

Year	Data source
1870–1871	1872 total German market cap extrapolated back using the growth in the capitalization of 30 largest German listed companies from Ronge (2002).
1872–1913	Market capitalization of all German firms listed on all major German exchanges (Berlin, Frankfurt, Hamburg, Cologne, Leipzig and Munich), adjusted for cross-listings, computed by authors from microdata helpfully shared by Christian Hirsch at the Frankfurt Center for Financial Studies. The underlying data are sourced from the regional financial newspapers and stock listings, namely: the *Berliner Börsen-Zeitung*, *Berliner Börsencourier* and *Neumann's Cours-Tabellen*; Frankfurt, Munich and Leipzig *Börsen-Kursblatt*; *Frankfurter Zeitung*, *Hamburgischer Correspondent*, *Kölnische Zeitung* and *Kölner Tageblatt*.
1914–1918	Capitalization of 30 largest German listed companies from Ronge (2002) scaled up to match all German listed companies (1913 used as benchmark year for scaling).
1919–1924	Total market capitalization of shares listed on the Berlin stock exchange from Moore (2010b), scaled up to match all of Germany and down to exclude foreign firms, using data for overlapping years between the Moore (2010b) and our all-Germany series in the early 20th century.
1925	Capitalization of 30 largest German listed companies from Ronge (2002) scaled up to match all German listed companies (1913 used as benchmark year for scaling).
1926–1943	Total capitalization of shares listed on the Berlin stock exchange from Deutsche Bundesbank (1976), scaled up to match all of Germany and down to exclude foreign firms, using data for overlapping years between Moore (2010b)'s Berlin series and our all-Germany series in the early 20th century.
1944–1950	Capitalization of 30 largest German listed companies from Ronge (2002) scaled up to match all German listed companies (1943 and 1950 used as benchmark years).
1951–1974	Total market cap of all German listed firms, shares listed on German exchanges, from Deutsche Bundesbank (1976). Spliced with the scaled-up Berlin series over years 1944–1950.
1975–1998	Total market cap of all German listed firms, shares listed on German exchanges, from World Bank's *WDI Database*.
1999–2017	Total market cap of all German listed firms, shares listed on German exchanges, from the Bundesbank database (series BBK01.WU0178).

Table D.7 documents the sources of our stock market capitalization data for Germany, and Figure D.7 plots the resulting series alongside alternative existing estimates. For years 1873–1914, we construct our own best-practice estimate of the German stock market capitalization, using data on individual securities listed on all major German exchanges (Berlin, Frankfurt, Hamburg, Cologne, Leipzig and Munich), adjusted for cross-listings, and computed from microdata helpfully shared

Figure D.7. Germany: alternative stock market cap estimates



by Christian Hirsch at the Frankfurt Center for Financial Studies. Outside of these data, we rely on a number of proxies to construct the capitalization of all German companies listed in Germany from a variety of other sources. These proxies consist of the Ronge (2002) estimates of the capitalization of the largest 30 listed German companies, helpfully shared with us by Ulrich Ronge, and covering the period 1870–1958; and the total capitalization of the Berlin stock exchange computed by Moore (2010b) for years 1899–1924, and by Deutsche Bundesbank (1976) for years 1926–1943. We scale down the Berlin capitalization data to mimic the exclusion of foreign companies, and scale it up to mimic the inclusion of regional exchanges, by comparing the Berlin capitalization estimates to those for the whole of Germany for various benchmark years. Finally, we use the Ronge (2002) series to fill in the remaining gaps.

The different early-period series match up with each other rather well: for example, in the 1870s most of the total market cap can be accounted for by the 30 largest companies (the Ronge, 2002, estimates), and the top-30 share gradually decreases as new listed firms enter the market in the late 19th and early 20th centuries, before the market becoming more concentrated again during the interwar period and the 1930s. In the early 20th century, the total Berlin capitalization is actually somewhat larger than that of the German companies listed on all German exchanges, due to a large presence of foreign stocks, and the two measures (Berlin total vs all-Germany German companies) become very similar in the 1920s and 1930s as the share of foreign stocks drops after World War 1.

The post-1950 data cover all German company ordinary shares listed on German exchanges, and are sourced from the various Bundesbank publications, namely Deutsche Bundesbank (1976) and the online statistical database of the Bundesbank. These match up rather well with alternative estimates from the ECB database, the World Bank's *WDI database*, and data from the *World Federation of Exchanges*. Concerning the earlier estimates, both Rajan and Zingales (2003) and Goldsmith (1985) have tended to overestimate the size of the German stock market relative to GDP somewhat.

We are grateful to Christian Hirsch for sharing data, to Ulrich Ronge for sharing data and offering advice on the historical German seris, and to Carsten Burhop for helping us locate the historical data sources.

Italy

Table D.8. Data sources: Italy

Year	Data source
1900, 1913	Total stock market capitalization of Italian firms, estimates from Musacchio (2010).
1928–1949	Total stock market capitalization of Italian firms, shares listed in Italy, aggregated from individual stock capitalizations published in Mediobanca (Various years).
1950–1988	Total stock market capitalization of Italian firms, shares listed in Italy, using aggregate estimates published in Mediobanca (Various years). No data for 1951.
1989–2017	Total capitalization of Italian firms, shares listed in Italy, from the *ECB Statistical Data Warehouse*, Security issues statistics.

Figure D.8. Italy: alternative stock market cap estimates

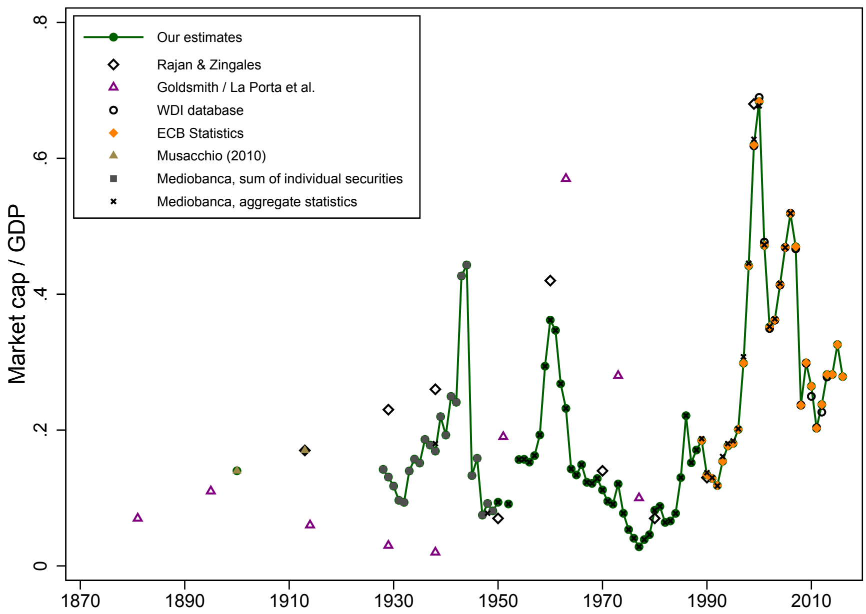


Table D.8 documents the sources of our stock market capitalization data for Italy, and Figure D.8 plots the resulting series alongside alternative existing estimates. Most of the data are sourced from the *Indici e Dati* publication, Mediobanca (Various years), which presents various aggregate and security-level statistics on Italian stocks and bonds, as well as further accounting data for the major Italian companies. For years 1928–1949, this publication publishes the market capitalization of individual Italian listed companies, and we compute our market cap measure as an aggregate of these security-level data. From 1950 onwards, *Indici e Dati* publishes aggregate market capitalization statistics relating to shares of all Italian firms listed on Italian exchanges, which becomes the main source of our data. Even though the individual security listings from the earlier years could miss out on some smaller firms, comparison of the two Mediobanca series (dark squares and x crosses in Figure D.8) suggests that these differences are, in practice, negligible. The later-years Mediobanca aggregate series match up well with alternative estimates from the World Bank's *WDI Database* and the ECB *Statistical Data Warehouse*. We use the ECB series for our estimates from 1989 onwards.

For the early years, we use Musacchio (2010) estimates of the Italian market capitalization in 1900 and 1913, with the 1913 estimate being the same as those of Rajan and Zingales (2003). We do not use the earlier Goldsmith (1985) estimates, because in years 1910, 1930 and 1940 these seem to vastly underestimate the size of the Italian stock market. The Rajan and Zingales (2003) estimates are, on average, somewhat higher than those in our paper.

We are grateful to Stefano Battilossi for providing helpful advice in locating the historical data sources for Italy.

Japan

Table D.9. Data sources: Japan

Year	Data source
1881–1899	The 1900 market capitalization extrapolated back using changes in the book capital of business corporations from Bank of Japan (1966), and stock price growth from Jordà, Knoll, et al. (2019).
1900–1924	Total capitalization of the Tokyo stock exchange from Moore (2010b).
1925–1945	The 1924 market capitalization extrapolated back using changes in the book capital of business corporations from Bank of Japan (1966), and stock price growth from Jordà, Knoll, et al. (2019).
1948–2004	Total capitalization of the Tokyo stock exchange first and second sections, from the *Statistics Bureau of Japan* historical statistics, Tables 14-25a and 14-25b.
2005–2013	Total capitalization of Japanese firms' shares listed on Japanese exchanges, from World Bank's *WDI Database*.
2014–2016	Total capitalization of Japanese firms listed on the Tokyo stock exchange, from the *World Federation of Exchanges* statistical reports.

Figure D.9. Japan: alternative stock market cap estimates

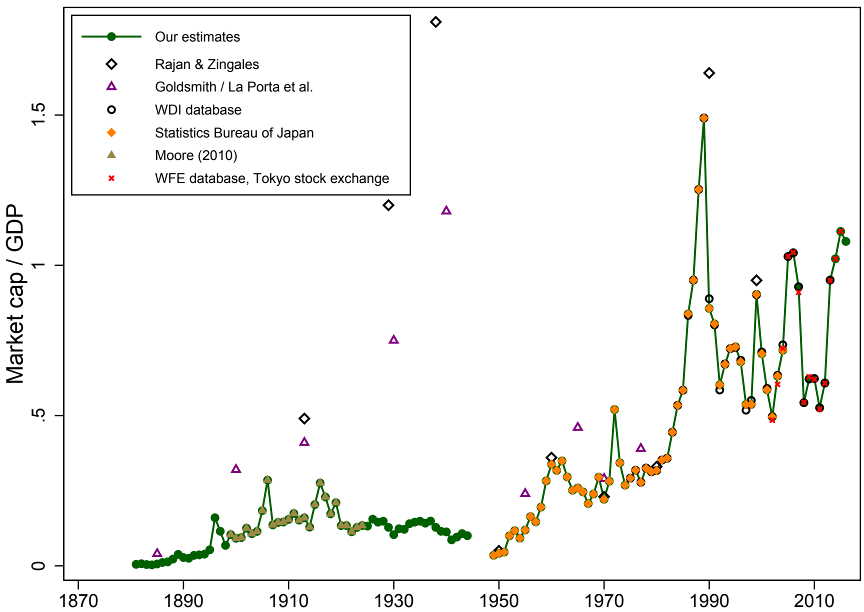


Table D.9 documents the sources of our stock market capitalization data for Japan, and Figure D.9 plots the resulting series alongside alternative existing estimates. For the early historical period, our main source are the Moore (2010b) estimates of the total capitalization of the Tokyo stock exchange. While these may somewhat understate the total capitalization of Japanese firms because they exclude regional exchanges, they may also overstate it via including foreign shares, with these two biases, to some extent, balancing against each other. The Moore (2010b) data cover the period 1899–1924. For the adjacent historical periods, we rely on a mixture of book capital data that covers both listed and unlisted businesses, from Bank of Japan (1966), and stock price data from Jordà, Knoll, et al. (2019). For each year in the period 1881–1898 and 1925–1945, we estimate the change in market cap as the stock price change multiplied by the change in the book capital of listed firms. This implicitly assumes that the share of the book capital of listed firms relative to that of all firms remains relatively stable. For the late 19th century period our data may, therefore, somewhat understate the growth in market cap – but the book capital statistics already capture the rapid growth in business equity in Japan during this period, with both book and market cap growing rapidly between 1881 and 1900. The ratio of market cap to GDP, and the relative importance of listed and unlisted firms seem to somewhat stabilise from 1910 onwards, even during the period for which we have the non-extrapolated market capitalization data.

Alternative estimates for the early period do exist, but they appear somewhat more noisy and less reliable than even our extrapolated data. The estimates of Rajan and Zingales (2003) and Goldsmith (1985) are not too far away from ours in the 1880s, but report much higher capitalization especially for the period of the 1930s and World War 2. These very high capitalization ratios are, however, may be somewhat difficult to justify in light of other available data. The implied stock market expansion in the 1930s goes far beyond both the book capital growth and the increase in the share price index, implying new listings that far exceed the data reported for other periods and countries in our sample. The post World War 2 data, where our estimates, based on the Statistics Bureau of Japan historical statistics, are more consistent with those of Rajan and Zingales (2003) and Goldsmith (1985), again suggest a drop in market size that far exceeds that suggested by the stock price data. Even though the Tokyo stock exchange was closed during years 1946–1947, the market capitalization ratios reported by Rajan and Zingales (2003) and Goldsmith (1985) in the 1940s are in the region of 1.2–1.8 of GDP, whereas those reported in the late 1940s and 1950s by both Statistics Bureau of Japan and Rajan and Zingales (2003) are closer to 0.035–0.05 of GDP. Even without taking the falls in GDP during this period into account, this implies a 30–50 fold drop company valuations during this short period of time, which seems unlikely. In light of these, we do not benchmark our series to the Rajan and Zingales (2003) and Goldsmith (1985) estimates, but without a doubt, there is likely to be some noise in this early period data, especially in the 1930s and 1940s.

For the recent period, we use the Statistics Bureau of Japan estimates of the Tokyo stock exchange capitalization (both the 1st and 2nd sections) during the period 1949–2004, which match up rather well with the total capitalization of all Japanese listed firms reported in the World Bank's *WDI Database*. For the latest period, we use the *World Federation of Exchanges* capitalization of Japanese firms listed on the Tokyo exchange, which is similar to the *WDI Database* estimates. Even though *WFE* also provide estimates for the Osaka exchange capitalization, a comparison with *WDI* data suggests a high degree of cross-listings among the two exchanges, therefore we use the Tokyo only series for the most recent years.

Netherlands

Table D.10. Data sources: Netherlands

Year	Data source
1899–1924	Total capitalization of the Amsterdam stock exchange from Moore (2010b), scaled down to proxy domestic firms only (using the proportion of domestic to foreign shares listed on the exchange in Moore, 2010b).
1938	Netherlands stock market cap estimate from Rajan and Zingales (2003).
1951–1974	Total capitalization of Dutch firms listed on the Amsterdam stock exchange from Central Bureau of Statistics (2010).
1975–1988	Total capitalization of Dutch firms' shares listed on Dutch exchanges, from World Bank's *WDI Database*.
1989–2017	Total capitalization of Dutch firms, shares listed in the Netherlands, from the *ECB Statistical Data Warehouse*, Security issues statistics. Spliced with WDI data for year 1989.

Figure D.10. Netherlands: alternative stock market cap estimates

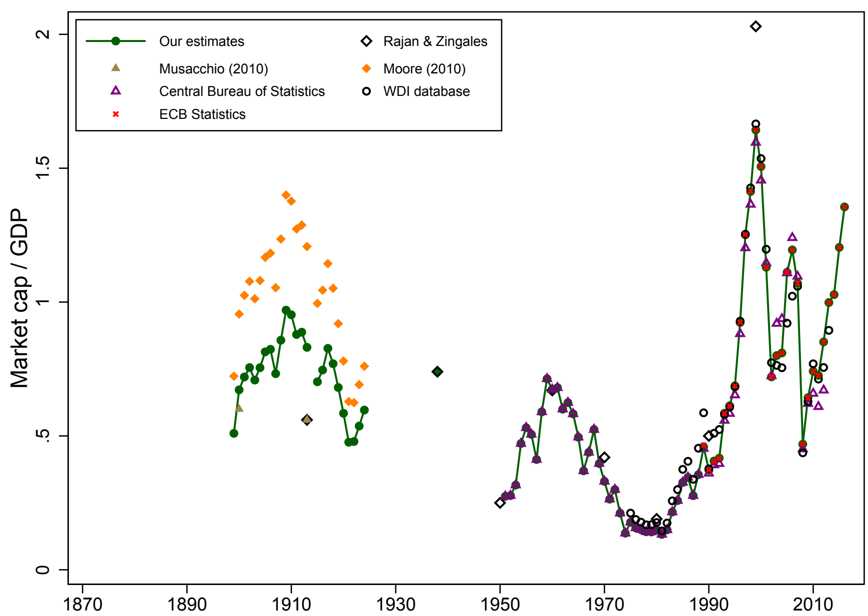


Table D.10 documents the sources of our stock market capitalization data for the Netherlands, and Figure D.10 plots the resulting series alongside alternative existing estimates. In the early period, our main source are the Moore (2010b) estimates of the total capitalization of the Amsterdam stock exchange. One issue, however, is that the Amsterdam exchange played an important role in the international financial system during this time period, and was used to trade many foreign as well as domestic stocks, as is clear from examining the stock exchange listings and the summary statistics on foreign and domestic listings in Moore (2010b). The total Amsterdam capitalization estimates in Moore (2010b) are, therefore, likely to substantially overstate the capitalization of Dutch firms. This also helps explain why the Moore (2010b) market cap estimates are higher than those of Rajan and Zingales (2003) and Musacchio (2010) for the early period, whereas our estimates for the later periods are broadly in line with those of Rajan and Zingales (2003). To adjust for this bias, we scale down the total capitalization of the Amsterdam exchange using the statistics on domestic and foreign shares listed in years 1899, 1909 and 1924 in Moore (2010b), calculating capitalization for this early period as total Amsterdam cap * number of Dutch shared listed / total number of shares listed. Depending on the relative size of the average capitalization of domestic and foreign shares, and the accuracy of estimates in-between the benchmark periods, these estimates could either somewhat over- or understate the total capitalization of Dutch firms.

For the post-1950 data, we rely on estimates of capitalization of Dutch firms listed on Dutch exchanges from three sources: the 111 year statistics Central Bureau of Statistics (2010), and the data from World Bank's *WDI database* and ECB's *Statistical Data Warehouse*. These estimates tend to be similar to each other, and to those of Rajan and Zingales (2003). In light of this data consistency among the different sources, we also make use of the Rajan and Zingales (2003) estimate of the 1938 Dutch stock market cap.

Norway

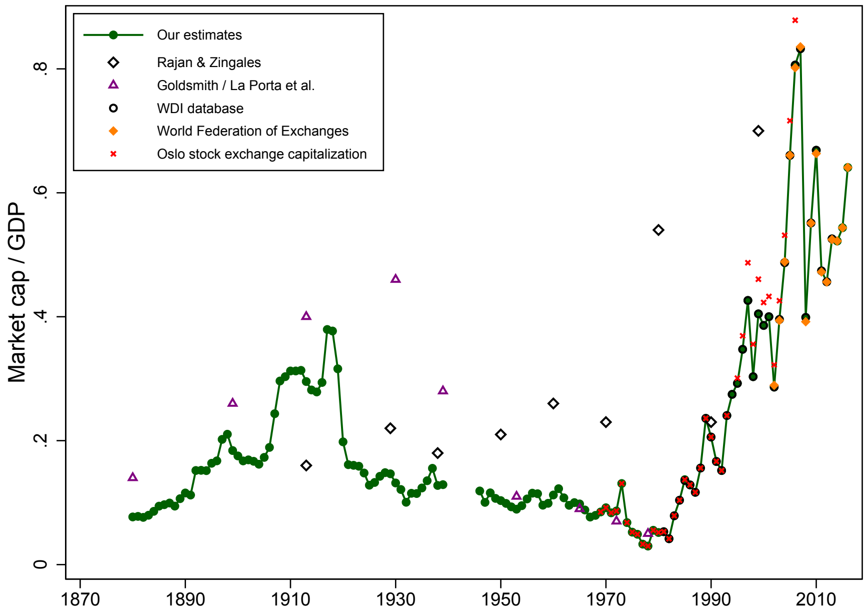
Table D.11. Data sources: Norway

Year	Data source
1880–1899	Total market capitalization of all Norwegian listed firms' ordinary shares, own estimates using individual stock data in the *Kierulf handbook* and Oslo stock exchange listings.
1900–1918	Market to book of listed firms times an estimate of listed book capital. Changes in listed book capital proxied using changes in total book capital for years 1900–1911 and 1911–1917. The data for 1912 and 1918 are direct measures of the total market capitalization of Norwegian firms, computed in the same way as for the period 1880–1899. Microdata sourced from *Kierulf handbook* and Oslo *Kurslisten*; aggregate book capital data sourced from the statistical yearbooks, various years.
1919–1968	Estimate of the total capitalization of Norwegian firms, computed as share capital of all Norwegian firms * proxy for share of listed firms * market-to-book of listed firms. The share of listed firms calculated as listed book capital relative to book capital of all firms in 1918, and as market capitalization of Oslo stock exchange relative to market value of all firm equity in 1969, and interpolated in-between (the 1918 and 1969 listed firm shares are very similar). Sources: *Kierulf handbook*, Oslo *Kurslisten*, statistical yearbooks, various years.
1969–1993	Total capitalization of the Oslo stock exchange, data kindly shared by Daniel Waldenström.
1994–2013	Total capitalization of Norwegian firms' shares listed in Norway, from World Bank's *WDI Database*.
2014–2016	Total capitalization of Norwegian firms' shares listed in Norway, from *World Federation of Exchanges* (WFE) reports, various years.

Table D.11 documents the sources of our stock market capitalization data for Norway, and Figure D.11 plots the resulting series alongside alternative existing estimates. For the early historical period, we construct our own estimates of stock market capitalization using data on individual stock prices and quantities, sourced from various issues of the *Kierulf handbook*, and the Oslo stock exchange listings. For the late 19th century, we compute market capitalization in this manner for each individual year, and for the early 20th century we compute capitalization at benchmark years and use changes in book capital of all companies and the market-to-book value of listed companies to calculate the year-on-year movements in market capitalization. Since the share of listed company capital relative to book capital of all companies varies little across the different benchmark years, this calculation ought to be fairly accurate. The data show a substantial stock market boom in the late 1910s, and the subsequent stock market crash of the early 1920s during which market capitalization more than halved.

For the modern period (1969 onwards), we start off by using the total capitalization of the Oslo stock exchange. Given the negligible presence of non-Norwegian

Figure D.11. Norway: alternative stock market cap estimates



companies on the exchange during this time period (which can be seen, for example, by comparing the *WDI* estimates for Norwegian firms with the Oslo exchange cap for overlapping years in Figure D.11), this acts as a good proxy for the total capitalization of Norwegian listed firms. In the 1990s and 2000s, we switch to using the *WDI* and *WFE* data, which focus on Norwegian firms only.

To link the 1969 and 1918 measures of stock market cap, we estimate market cap movements using changes in the book capital of all firms, the market-to-book value of listed firms, and a proxy for the proportion of the firms that are listed. The time between the 1920s bust and the 1980s marks a relatively stable period for the Norwegian stock market with, for example, the listed firm share growing by only 4 percentage points, from 30% in 1918 to 34% in 1969, which suggests that our estimates should have a relatively high degree of accuracy.

Taken together, our market capitalization estimates are substantially below those of Rajan and Zingales (2003). Somewhat surprisingly, the benchmark year Rajan and Zingales (2003) estimates for the early 20th century do not contain any evidence of the large boom-bust cycle that took place around 1920 and is evident both in the share price and our market capitalization data. The estimates of Goldsmith (1985) are above ours for the early to mid 20th century period, but similar to ours after 1950.

We would like to thank Jan Tore Klovland for helping us locate and interpret the historical sources for the Norwegian stock price data, and the staff at the Oslo Nasjonalbiblioteket in Oslo for their help in locating the sources.

Portugal

Table D.12. Data sources: Portugal

Year	Data source
1870–1987	Total market capitalization of all Portuguese firms listed in Lisbon, own estimates using individual stock data and company published accounts. Sourced from *Diario do Governo*, *Boletim da Bolsa* and individual company accounts, various years. For years 1900–1925, we use changes in book capital for a subset of listed firms to estimate the changes in book capital of all listed firms. Market capitalization during the Carnation revolution related stock market closure in 1975–1976 is interpolated linearly using the data for 1974 and 1977.
1988	Splice own estimates constructed from microdata in the *Boletim da Bolsa* and the ECB series, using the average of 1987 cap * price growth, and 1989 cap / price growth.
1989–2017	Total capitalization of Portuguese firms, shares listed in Portugal, from the *ECB Statistical Data Warehouse*, Security issues statistics.

Figure D.12. Portugal: alternative stock market cap estimates

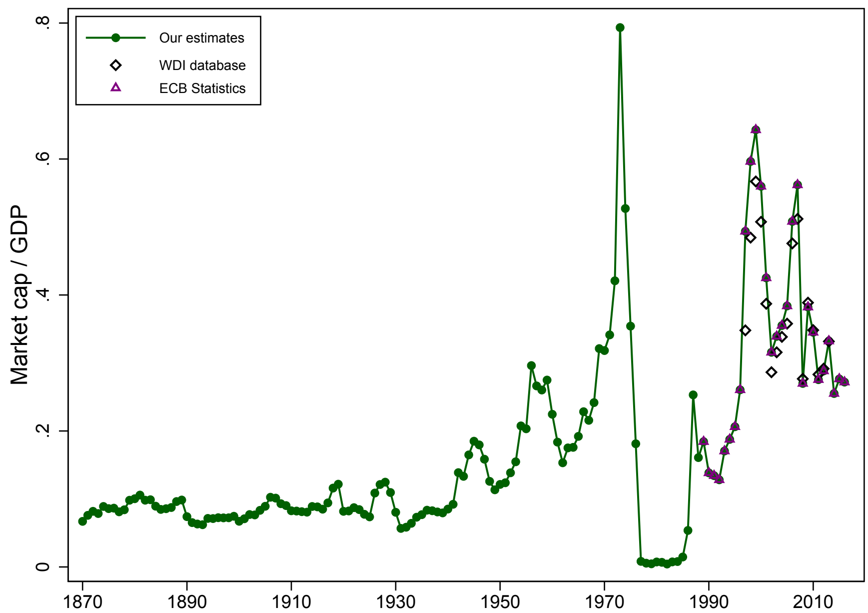


Table D.12 documents the sources of our stock market capitalization data for Portugal, and Figure D.12 plots the resulting series alongside alternative existing estimates. Very few estimates of the Portuguese market capitalization exist, particularly for the period before 1990. Therefore we construct our own data using the prices and quantities of each stock listed on the Lisbon stock exchange during this period, and aggregating the individual shares' market capitalization. Throughout, we exclude preference shares, foreign and colonial companies to arrive at a measure of domestic market capitalization. Even though a smaller stock exchange operated in Porto, data from the stock listings suggest that its size was very small relative to the Lisbon exchange; therefore our estimates provide a good measure of the total market capitalization of Portuguese listed firms.

Most of the early period data are sourced from the official stock exchange listing *Boletim da Bolsa*, available for years 1874 to 1987. This listing contains information on both stock prices and quantities. These data are complemented by stock listings and company balance sheets published in the government newspaper *Diario do Governo*, and balance sheet data in the published accounts of limited companies. These additional sources are particularly important for the period 1900–1925, during which the official *Boletim* stopped publishing share quantity data. For these years, we use a subset of listed companies, for which we have published accounts data, to estimate the changes in share quantities for the entire market. Another approximation is undertaken during years 1975–1976, when the stock exchange was closed in the aftermath of the Carnation revolution. Stock market capitalization dropped almost twenty-fold between 1974 and 1977, and we interpolate this drop across the years during which the stock exchange was closed, so that it this negative shock is not absent from our data. After the shock of the Carnation revolution, the market stagnated during the 1970s before recovering apidly in the late 1980s. Portugal is the only country in our sample that saw very high net issuance during this "big bang" period as new companies entered the market – this, however, is rather specific to the recovery of the market from the turmoil associated with the 1970s revolution.

The modern data are sourced from the World Bank *World Development Indicators* and ECB's *Statistical Data Warehouse*, and match up with our own estimated series, as well as each other, rather well.

We are grateful to Jose Rodrigues da Costa and Maria Eugenia Mata for help and advice in finding and interpreting the data sources for the historical Portuguese data. We are also grateful to staff at the Banco do Portugal archive for helpful advice and sharing data.

Spain

Table D.13. Data sources: Spain

Year	Data source
1900–1924	Total market capitalization of all Spanish firms listed in Madrid, own estimates using microdata helpfully shared by Lyndon Moore. See Moore (2010b) and Moore (2010a) for the original source. We scale up the series to match our own estimates using microdata from the Madrid stock exchange listings in 1925.
1925–1936; 1940	Total market capitalization of all Spanish firms listed in Madrid, own estimates using microdata from the Madrid stock exchange listings, *Boletin de Cotization Oficial*, various years.
1941–1988	Total capitalization of the major Spanish stock exchanges from López, Carreras, and Tafunell (2005). Between 1941 and 1971, data are provided at 5-year benchmarks, with the in-between changes in market cap estimated using the changes in the stock price index from Jordà, Knoll, et al. (2019), and changes in the total book capital of Spanish firms from López, Carreras, and Tafunell (2005).
1989–2017	Total capitalization of Spanish firms, shares listed in Spain, from the *ECB Statistical Data Warehouse*, Security issues statistics.

Figure D.13. Spain: alternative stock market cap estimates

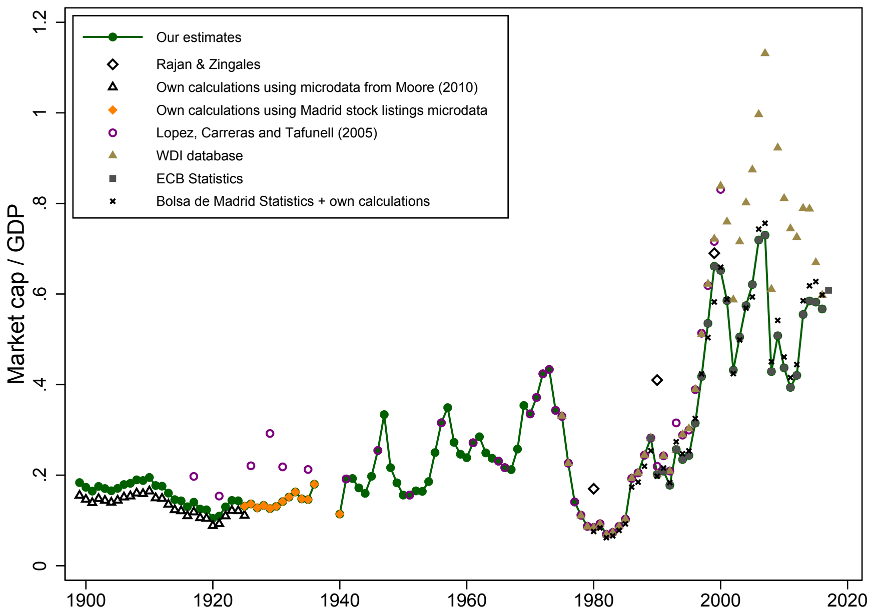


Table D.13 documents the sources of our stock market capitalization data for Spain, and Figure D.13 plots the resulting series alongside alternative existing estimates. For the early historical period, we construct estimates of total capitalization of ordinary shares of Spanish firms listed on the Madrid stock exchange by aggregating up the capitalizations of individual shares in the official Madrid stock list. The data on share prices and quantities for 1925–1941 were source directly from the official stock list, *Boletin de Cotization Oficial*. Microdata for the 1899–1924 period were helpfully shared with us by Lyndon Moore, and are a slightly updated version of the series in Moore (2010b), sourced from Moore (2010a). The 1899–1924 are missing some of the smaller securities listed on the exchange, and we scale up these series slightly using benchmark ratios from overlapping data in 1925. The data do not include the Barcelona stock exchange, as the listings in, for example, the *La Vanguardia* newspaper do not contain information on quantities. But the early 20th century Barcelona listings suggest that trading on that exchange mainly comprised of government and corporate bonds, with few shares listed on the Barcelona exchange. The bias from excluding this exchange is, therefore, likely to be small. During the Spanish civil war, the stock exchange was closed, hence the data for years 1937–1939 are missing. Given that the stock capitalization did not change dramatically over this period, and that the missing period covers several years, we choose not to interpolate the data for the civil war period.

From 1941 onwards, we use estimates for the total capitalization of the major Spanish exchanges – starting with Madrid, and later also including Barcelona, Bilbao and Valencia – provided by López, Carreras, and Tafunell (2005). Before 1970, these are only available at 5-year benchmark periods. To estimate the market cap movements between benchmark years, we estimate the year-to-year changes in capitalization as the stock price growth times the change in the capital of all Spanish firms, using data from Jordà, Knoll, et al. (2019) and López, Carreras, and Tafunell (2005) respectively, with the growth rates scaled up or down to match the capitalization at benchmark years. Accurate interpolation relies on the proportion of listed firms not fluctuating too much from year to year within the five-year benchmark periods. Given that book capital of listed firms does not vary dramatically from year to year in other time periods in the Spanish data, or in the data for other countries, the measurement error from this interpolation is unlikely to be large. From 1970 onwards, López, Carreras, and Tafunell (2005) provide annual estimates of Spanish listed firms' market capitalization. The *WDI Database*, Bolsa de Madrid and the ECB *Statistical Data Warehouse* provide alternative estimates for the modern period. The WDI estimates for Spain, unfortunately, seem to suffer from considerable measurement error (after liaison with the WDI database staff some of these were fixed, but some seem to remain in place given the difference between the WDI series and all other estimates in Figure D.13). The *Bolsa de Madrid Statistics* estimates are accurate, but the share of foreign firms had to be proxied by us before year 2001. In light of this, we use the ECB's series for the modern period, which are close to estimates provided by López, Carreras, and Tafunell (2005) and Bolsa de Madrid.

We would like to thank Lyndon Moore for sharing the microdata from the Madrid stock exchange for the early historical period as well as offering helpful advice, and Stefano Battilossi in helping locate the historical data sources.

Sweden

Table D.14. Data sources: Sweden

Year	Data source
1870–2012	Total market capitalization of Swedish firms from Waldenström (2014).
2013–2017	Total capitalization of Swedish firms, shares listed in Sweden, from the *ECB Statistical Data Warehouse*, Security issues statistics.

Figure D.14. Sweden: alternative stock market cap estimates

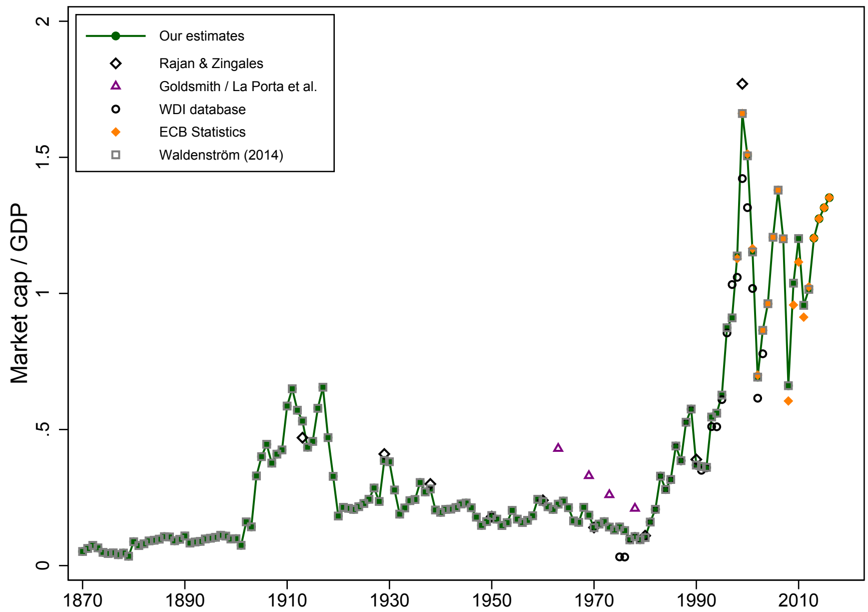


Table D.14 documents the sources of our stock market capitalization data for Sweden, and Figure D.14 plots the resulting series alongside alternative existing estimates. The main source for our series are the data compiled by Waldenström (2014), who put together a long-run series of Swedish stock market capitalization as part of a broader effort to document the evolution of returns and capitalization of the Swedish stock market, and the evolution of wealth in Sweden. For the modern period, the Waldenström (2014) series are very similar to the estimates in the World bank's *WDI Database* and the ECB's *Statistical Data Warehouse*. Because the WDI series contains what looks like typos in years 1975–1976, we use the ECB series to complement the Waldenström (2014) data for the modern period. Our data are close to the estimates of Rajan and Zingales (2003) for the selected benchmark years, and somewhat below the earlier Goldsmith (1985) series.

We are grateful to Daniel Waldenström for providing helpful advice in interpreting the historical Swedish data and sources.

Switzerland

Table D.15. Data sources: Switzerland

Year	Data source
1900–1925	Total market capitalization of all Swiss firms listed in Zurich, own estimates using microdata helpfully shared by Lyndon Moore. See Moore (2010b) and Moore (2010a) for the original source.
1926–1974	1925 stock market cap extrapolated forward using net issuance data from the Swiss National Bank *Capital Market statistics*, and stock returns, with growth rates adjusted down on average 0.15% per year to match the 1975 market cap value from the *WDI Database*. Net issuance before 1941 is estimated as fixed proportion of gross issuance.
1975–1979	Total capitalization of all Swiss listed firms, shares listed on Swiss exchanges. Source: World Bank *WDI database*.
1980–2017	Total capitalization of the Swiss and Liechtenstein firms listed on the SIX (Swiss stock exchange), from the SNB *Capital Market Statistics*.

Figure D.15. Switzerland: alternative stock market cap estimates

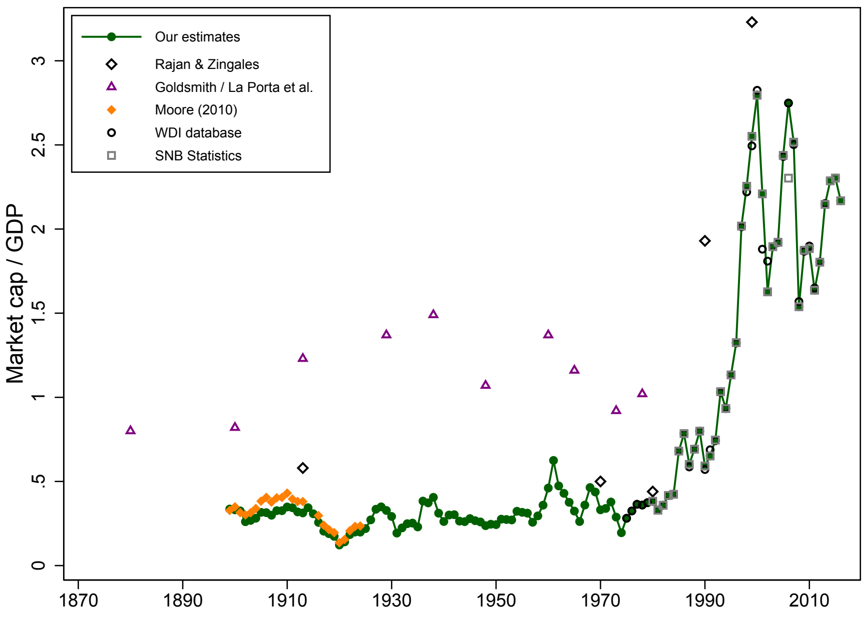


Table D.15 documents the sources of our stock market capitalization data for Switzerland, and Figure D.15 plots the resulting series alongside alternative existing estimates. The early estimates of the Swiss stock market capitalization are based on the Moore (2010b) data for the capitalization of the Zurich stock exchange. We use microdata helpfully shared with us by Lyndon Moore to construct our own estimate of the capitalization of Swiss firms listed in Zurich, with the data sourced from Moore (2010a), a sligthly updated version of Moore (2010b). The estimates are close to the Zurich total in Moore (2010b), but slightly below it due to the exclusion of foreign firms.

The modern data are based on the statistics in the World Bank's *WDI Database*, and the *Capital Market Statistics* of the Swiss National Bank, both of which aim to capture all Swiss firms listed in Switzerland. The two series are close to each other, and we use the WDI series for the early years, switching to the SNB data when these become available.

To link the capitalization estimates in 1925 and 1975, we use data on net issuance, provided by the SNB, and stock price data from Jordà, Knoll, et al. (2019). The net issuance data cover all publicly floated issues, and thus closely mirror the issuance of listed firms. Before 1944, we proxy net issuance as a fixed proportion of the gross issuance series. We calculate the capitalization in each year as the previous year's capitalization, times the stock capital gain, plus the net issuance times half the capital gain for the year (thus assuming that the issuance, on average, occurred in the middle of the year). Altogether, this proxy captures the two drivers of the movements in market capitalization, and hence should have a high degree of accuracy. Consistent with this, our estimate of the market capitalization in 1975, constructed by extrapolation using net issuance and capital gains over the period 1926–1975, is within 10% of the WDI stock market cap value in 1975, implying an average estimation error of less than 0.2% of market cap (or 0.06% of GDP) per year. We adjust the overall growth rate between 1926–1975 down slightly to match the 1975 benchmark.

Compared to other commonly used estimates, ours are substantially smaller than the early proxies from Goldsmith (1985), and are similar but slightly below the estimates of Rajan and Zingales (2003) at the corresponding benchmark years.

We would like to thank Lyndon Moore for sharing the microdata from the Zurich stock exchange and offering helpful advice, and to Carmen Hofmann and Rebekka Schefer for helping locate the historical sources.

United Kingdom

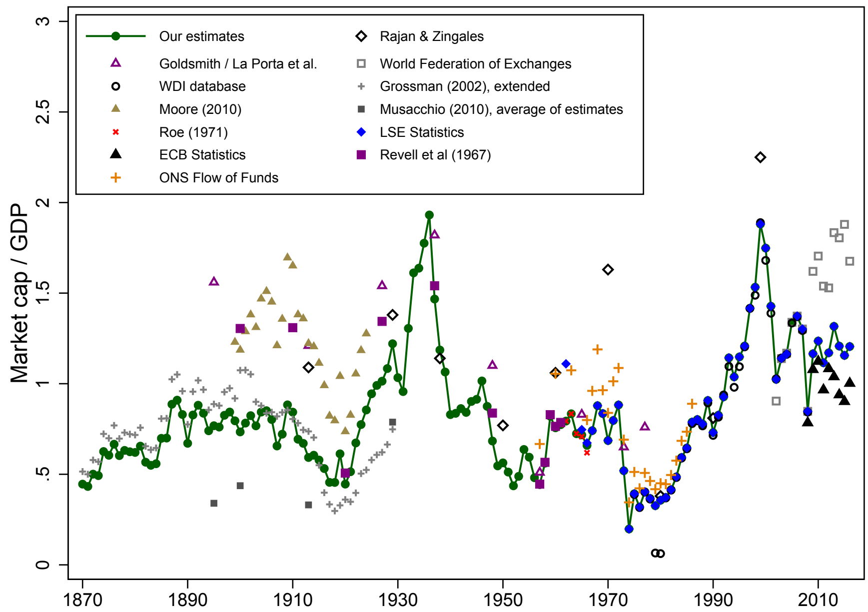
Table D.16. Data sources: United Kingdom

Year	Data source
1870–1898	The 1899 capitalization extrapolated back using annual changes in the market capitalization of all UK firms shared with us by Richard Grossman (see Grossman, 2002, for a description of the data).
1899–1924	Estimate of total market capitalization of all UK firms from the total capitalization of the London Stock Exchange computed by Moore (2010b). We scale down the Moore (2010b) estimates to proxy UK-only firms using data on the share of domestic firms in the listings from Musacchio (2010), and scale it up to proxy non-London exchanges by using data on the share of regional exchanges from Campbell, Rogers, and Turner (2016).
1925–1929	The 1924 capitalization extrapolated forward using annual changes in the market capitalization of all UK firms shared with us by Richard Grossman (see Grossman, 2002, for a description of the data).
1930–1956	Market value of equity of all UK firms (listed and unlisted) from Solomou and Weale (1997), scaled down to proxy listed firms only, using overlapping data with our estimates in the 1920s, and the market value of quoted shares estimated by Roe (1971) in the 1950s.
1957–1964	Total value of quoted UK ordinary shares, from Roe (1971).
1965–1994	Marked value of all UK and Irish companies listed on the London Stock Exchange, from *LSE Historical Statistics*. Spliced with the Roe (1971) data over the period 1965–1967.
1995–2004	Marked value of all UK companies listed on the London Stock Exchange, from *LSE Historical Statistics*.
2005–2006	Total capitalization of the UK firms listed at the London Stock Exchange, from the *World Federation of Exchanges* (WFE) reports, various years.
2007–2017	Marked capitalization of all UK listed firms, from the London Stock Exchange *Main Market Factsheets*, various years.

Table D.16 documents the sources of our stock market capitalization data for the United Kingdom, and Figure D.16 plots the resulting series alongside alternative existing estimates. The main difficulty in estimating the UK's stock market capitalization comes about from two sources. First, since London has been an active financial center throughout the historical period considered, with an especially active role in the 19th and early 20th centuries, many stocks listed in London are those of foreign companies and need to be excluded from the total. Second, especially in the 19th century, the UK had a number of active regional exchanges (Campbell, Rogers, and Turner, 2016), whose capitalization needs to be added to the total.

For the early years in our sample, Grossman (2002) provides an estimate of UK market capitalization that fits our desired definition: the total cap of UK ordinary shares listed in London and other UK exchanges, using data from the *Investor Monthly Manual*, that covers UK and foreign stock listed on all UK exchanges. We

Figure D.16. United Kingdom: alternative stock market cap estimates



would like to thank Richard Grossman for sharing his market capitalization estimates with us, in an extended version of the Grossman (2002) dataset that covers years 1869–1929. The accuracy of these data is, however, subject to recent debate, with Hannah (2018) pointing out a number of potential irregularities in the series when compared to other sources. While current debate remains active around the quality of these early data, we use the estimates of Moore (2010b), instead, as our main source. The Moore (2010b) capitalization data, however, are for all London shares, and need to be adjusted to exclude foreign shares, and include shares listed on other exchanges. We do this in two steps. First, we scale the series down to exclude foreign stocks, using Musacchio (2010) estimates of domestic and foreign capitalization on the LSE. Musacchio (2010) provides a range of estimates covering benchmark years 1895, 1900, 1913 and 1929, and we use the average of his estimates interpolated between these benchmark years to proxy the domestic share (which remains close to 60% throughout this period). Second, we scale the domestic series up using estimates in Campbell, Rogers, and Turner (2016) of the London capitalization compared to other UK and Irish exchanges, using the share of London relative to UK and Ireland minus Dublin at 10-year benchmarks. The Campbell, Rogers, and Turner (2016) data include preference shares and debt as well as ordinary shares, so we cannot use their estimates directly, and instead use them to scale the Moore (2010b) data, which cover ordinary shares only.

The resulting early-period series, green line in Figure D.16, are below the Moore (2010b) estimates of total London market cap, because the foreign share is much larger than the contribution of provincial and regional exchanges to the total. The series is reasonably close to the estimates of Grossman (2002), and we use the changes in the Grossman (2002) series to extrapolate movements in stock market cap beyond the years 1899–1924 that are covered by the adjusted Moore (2010b) data. Our estimates are substantially below those of Goldsmith (1985) and Rajan and Zingales (2003), whose proxies are much closer to the London total, unadjusted to exclude foreign shares, and above the average of estimates in Musacchio (2010).[26]

For the mid-20th century, we rely on estimates of the national wealth of the UK, published in a variety of sources, and in particular the part of wealth that is attributed to quoted UK shares. The early data are sourced from Solomou and Weale (1997), who publish a combined figure that includes the market value of both listed and unlisted UK firms. We scale this down to proxy the capitalization of listed firms only, using overlapping data with listed-only series in the 1920s (our estimates based on Grossman, 2002; Moore, 2010b) and 1950s (the data from Roe, 1971). In the 1950s, we switch to Roe (1971)'s estimated of the value of all quoted UK shares. What stands out in these data is the UK stock market boom in the 1930s which saw market capitalization rise to as high as 2 times GDP – a value similar to that observed at the height of the dot-com boom in the late 1990s. The growth in market capitalization in the 1930s was almost entirely driven by rising stock prices – consistent with evidence reported in Section 3.4 of this paper – and dissipated close to the onset of World War 2. The only reason why this boom was not apparent in earlier estimates of Goldsmith (1985) and Rajan and Zingales (2003) is presumably the benchmark-year nature of their data – for example, the boom is apparent in the total listed and unlisted equity wealth estimates provided by Solomou and Weale (1997) (not shown in Figure D.16, but available from authors upon request).

For the second half of the 20th century and 21st century, we rely on official estimates of the capitalization of UK, or UK and Irish firms, provided by the London Stock Exchange. We use the UK and Irish capitalization provided in the *LSE Historical Statistics* between the 1960s and 1994. For the early 1960s, we stick to the Roe (1971) data, given that the LSE statistics estimate for 1962 seems to be an outlier, making us doubt its correctness (see Figure D.16). For 1995 onwards, we use data for UK firms only, with data before 2005 taken from the *LSE Historical Statistics*, and data after 2007 – from the LSE *Main Market Factsheets*, with the 2005–2006 gap plugged using the UK firms' London capitalization estimates provided by the *World Federation of Exchanges* (WFE) in their monthly statistical reports. A number of alternative estimates for this later period are shown in Figure D.16. These include national wealth estimates from the *Office for National Statistics*, World Bank's *WDI Database*, WFE reports and ECB's *Statistical Data Warehouse* data. These are generally close to our data and the estimates from the LSE, but overall seem somewhat

26. Musacchio (2010) recognises the difficulty of estimating the UK stock market capitalization precisely, and offers a range of estimates.

less accurate, with outliers such as the WDI data for 1975–1976 making us prefer the LSE data overall. Our estimates of the capitalization for the 1980s are similar to those of Rajan and Zingales (2003), while those at the height of the dot-com boom in 1999 are somewhat below theirs.

The diversity of the UK market, its large size, and the need to account for foreign shares and regional exchanges, make estimating the UK's market capitalization a tricky task, illustrated by the large variety of alternative estimates in Figure D.16. The ability to draw on all this previous work, however, means that we are able to select those estimates that best fit a consistent definition of UK firms' listed market cap, and provide a historical series that maps the evolution of the size of the UK equity market with a reasonable degree of accuracy. We are grateful to Richard Grossman for providing helpful advice and sharing data, and to Leslie Hannah and John Turner for offering helpful feedback on the data and historical sources.

United States

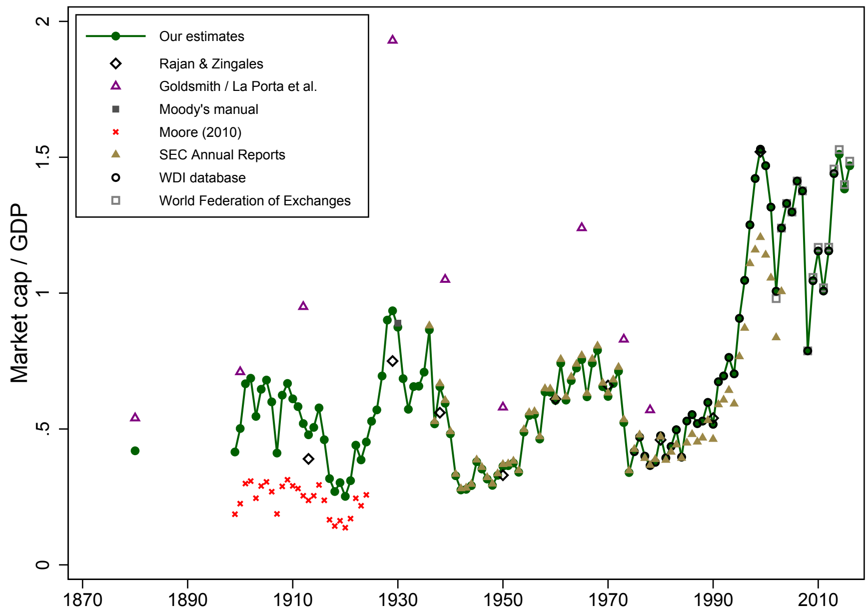
Table D.17. Data sources: United States

Year	Data source
1880	Goldsmith (1985) estimate of total equity wealth, scaled down to proxy the market capitalization of US listed firms, using the ratio of overlapping data for 1900 as the scaling factor.
1899–1924	Total NYSE market capitalization scaled up to reflect all exchanges, and scaled down to exclude foreign stocks. NYSE data from Moore (2010b). Scaling done using the data on relative importance of the NYSE and other exchanges helpfully shared by Leslie Hannah, and the ratio of NYSE to total cap in the Moody's manual. Share of foreign firms calculated using *NYX* historical data.
1925–1935	Total equity wealth of US firms scaled down to capture listed shares only. Equity wealth data from Piketty, Saez, and Zucman (2018). Scaling done by benchmarking to our pre-1925 estimates, to Moody's total US capitalization in 1930, and to SEC's data on capitalization of all US exchanges in 1936.
1936–1974	Total market capitalization of all US exchanges, from the SEC's *Annual Reports*, scaled down slightly to exclude foreign firms. Share of foreign firms calculated using *NYSE Historical Statistics*, and by comparing the SEC and WDI data for the 1970s.
1975–2013	Total capitalization of all US listed firms, shares listed on US stock exchanges, from the World Bank's *WDI database*.
2014–2016	Total capitalization of all US listed firms, shares listed on US stock exchanges, from *World Federation of Exchanges* (WFE) reports, various years.

Table D.17 documents the sources of our stock market capitalization data for the United States, and Figure D.17 plots the resulting series alongside alternative existing estimates. Most of the widely available estimates of US stock market capitalization refer to the New York stock exchange only, so the main challenge here reflects obtaining capitalization estimates that cover not only NYSE, but also other stock exchanges, and also adjusting estimates to exclude any foreign listings. Inclusion of non-NYSE stock exchanges is especially important for the early US data, with much of the trading taking place on the curb exchange and regional markets, as suggested in Sylla (2006)'s critique of Rajan and Zingales (2003) data.

Our early data use the Moore (2010b) estimates of the NYSE cap, scale these up to also account for other stock exchanges, and scale down to exclude foreign listings. We rely on a number of benchmark year estimates to approximate the relative importance of the NYSE. The 1906 NYSE share was helpfully shared with us by Leslie Hannah, and amounts to just over 40% in terms of book cap. Put differently, the New York Stock exchange accounted for less than half of total US capitalization in the early 20th century. By 1930, comparison of the total capitalization of US firms in Moody's manua to the NYSE capitalization estimates indicates that the NYSE share reached more than 60%, and by late 1930s that share was larger than 80%,

Figure D.17. United States: alternative stock market cap estimates



as suggested by data in the *SEC Annual Reports*. These broad trends are also consistent with turnover statistics of the different stock exchanges reported in O'Sullivan (2007). We interpolate the NYSE share in-between these benchmark years to obtain an annual proxy. As for the foreign share, based on the data from *NYX Historical Statistics*, this amounted to little over 2% in the mid 1920s. A similarly small foreign share is obtained by comparing the *SEC Annual Reports* and *WDI Database* estimates for the 1970s. Based on this, we adjust the Moore (2010b) NYSE-only estimates up substantially to approximate the inclusion of other exchanges, and account for the gradually increasing importance of the NYSE, and adjust them down slightly to proxy the exclusion of foreign ordinary shares. As a result, our market capitalization estimates in Figure D.17 are substantially above the NYSE capitalization in Moore (2010b), and are also higher than the Rajan and Zingales (2003) estimates which include regional exchanges but do not include the curb exchange, which was the largest non-NYSE market during this early period. We also use a market capitalization proxy for 1880, obtained from scaling down the Goldsmith (1985) data, which contain both listed and unlisted shares.

From mid-1930s onwards, estimates of total US market capitalization are available from the *SEC Annual Reports*. These include NYSE, Amex and regional exchanges. We adjust the estimates down very slightly to proxy the exclusion of foreign firms, and link the SEC series to the *WDI* data in the mid 1970s. For the modern period, we rely on a mixture of the *WDI* and *WFE* (World Federation of Exchanges) data, whose definition more precisely fits what we are after – namely, including all

US company shares listed on US stock exchanges. To fill a small gap in the 1920s and 1930s, we use annual growth in the capitalization of all US firms (listed and unlisted), provided by Piketty, Saez, and Zucman (2018), to estimate market capitalization growth in-between benchmark years.

Taken together, our US market capitalization estimates are much smaller than the early data from Goldsmith (1985), which includes a mixture of listed and unlisted shares. They are above the estimates of Rajan and Zingales (2003) for the early period, thanks to our inclusion of the curb exchange, and similar to the Rajan and Zingales (2003) estimates for the more recent period.

We would like to thank Leslie Hannah for sharing data and helping us locate and interpret the various historical sources.

References

Abildgren, Kim. 2006. "Monetary Trends and Business Cycles in Denmark 1875–2005: New Evidence Using the Framework of Financial Accounts for Organising Historical Financial Statistics." Danmarks Nationalbank Working Paper 43.

Alvaredo, Facundo, Anthony B. Atkinson, and Salvatore Morelli. 2018. "Top Wealth Shares in the UK over more than a Century." *Journal of Public Economics* 162: 26–47.

Annaert, Jan, Frans Buelens, and Marc De Ceuster. 2012. "New Belgian Stock Market Returns: 1832–1914." *Explorations in Economic History* 49 (2): 189–204.

Arbulu, Pedro. 1998. "Le marché Parisien des Actions au XIXe Siècle: Performance et Efficience d'un Marché Émergent." Doctoral dissertation. Orléans.

Atje, Raymond, and Boyan Jovanovic. 1993. "Stock Markets and Development." *European Economic Review* 37 (2): 632–640.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2016. "Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy." CEPR Discussion Paper 11734.

Baker, Malcolm, and Jeffrey Wurgler. 2000. "The Equity Share in New Issues and Aggregate Stock Returns." *Journal of Finance* 55 (5): 2219–2257.

Bank of Japan. 1966. *Hundred-Year Statistics of the Japanese Economy*.

Bernanke, Ben S. 2005. "The Global Saving Glut and the US Current Account Deficit." remarks by Governor Ben S. Bernanke at the Homer Jones Lecture, St. Louis, Missouri, The Federal Reserve Board of Governors.

Bianchi, Francesco, Martin Lettau, and Sydney C. Ludvigson. 2016. "Monetary Policy and Asset Valuation." NBER Working Paper 22572.

Black, Susan, Joshua Kirkwood, Thomas Williams, and Alan Rai. 2013. "A History of Australian Corporate Bonds." *Australian Economic History Review* 53 (3): 292–317.

Bozio, Antoine. 2002. "La Capitalisation Boursière en France au XXe Siècle." Master's thesis, Ecole Normale Supérieure.

Buckley, Kenneth Arthur Haig, and Malcolm Charles Urquhart. 1993. *Historical Statistics of Canada*. Statistics Canada.

Buffett, Warren, and Carol Loomis. 2001. "Warren Buffet on the Stock Market." Fortune Magazine, December 10th 2001.

Burdett, Henry C. 1882. *Burdett's Official Intelligence for 1882*. London: Spottiswoode.

Campbell, Gareth, Meeghan Rogers, and John D. Turner. 2016. "The Rise and Decline of the UK's Provincial Stock Markets, 1869-1929." QUCEH Working Paper 2016-03.

Campbell, John Y., and Robert J. Shiller. 1988. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1 (3): 195–228.

Carpentier, Cécile, Jean-François L'Her, and Jean-Marc Suret. 2009. "On the Competitiveness of the Canadian Stock Market." *Banking & Finance Law Review* 24 (2): 287.

Central Bureau of Statistics. 2010. *111 Jaar Statistiek in Tijdreeksen, 1899–2010*. Den Haag: Statistics Netherlands.

Chen, Long. 2009. "On the Reversal of Return and Dividend Growth Predictability: A Tale of Two Periods." *Journal of Financial Economics* 92 (1): 128–151.

Chen, Long, Zhi Da, and Richard Priestley. 2012. "Dividend Smoothing and Predictability." *Management Science* 58 (10): 1834–1853.

Cochrane, John H. 2008. "The Dog That Did Not Bark: A Defense of Return Predictability." *Review of Financial Studies* 21 (4): 1533–1575.

Cochrane, John H. 2011. “Presidential Address: Discount Rates.” *Journal of Finance* 66 (4): 1047–1108.

Cooper, Ilan, and Richard Priestley. 2008. “Time-varying Risk Premiums and the Output Gap.” *Review of Financial Studies* 22 (7): 2801–2833.

De Loecker, Jan, and Jan Eeckhout. 2017. “The Rise of Market Power and the Macroeconomic Implications.” NBER Working Paper 23687.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2019. “Global Trends in Interest Rates.” *Journal of International Economics* forthcoming:

Deutsche Bundesbank. 1976. *Deutsches Geld-Und Bankwesen in Zahlen, 1876-1975*. Knapp.

Diez, Federico, Daniel Leigh, and Suchanan Tambunlertchai. 2018. “Global Market Power and its Macroeconomic Implications.” IMF Working Paper 18137.

Eggertsson, Gauti B., Jacob A. Robbins, and Ella Getz Wold. 2018. “Kaldor and Piketty’s Facts: The Rise of Monopoly Power in the United States.” NBER Working Paper 24287.

Erichsen, Christian. 1902. “Green’s Danske Fonds og Aktier 1902. Udgivet og Redigeret af Holmer Green og Hendrik Stein.” NATIONALØKONOMISK TIDSSKRIFT 3.

Farhi, Emmanuel, and François Gourio. 2018. “Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia.” NBER Working Paper 25282.

Foster, Rob A. 1996. “Australian Economic Statistics 1949–50 to 1994–95.” Reserve Bank of Australia Occasional Paper 8.

Goldsmith, Raymond W. 1985. *Comparative National Balance Sheets: A Study of Twenty Countries, 1688–1978*. eng. Chicago, Ill.: University of Chicago Press.

Green, Theodor. 1887. “Fondsbørsen.” Nationaløkonomisk Tidsskrift 5.

Greenwald, Daniel L., Martin Lettau, and Sydney C. Ludvigson. 2014. “Origins of Stock Market Fluctuations.” NBER Working Paper 19818.

Greenwood, Robin, and Samuel G. Hanson. 2013. “Issuer Quality and Corporate Bond Returns.” *Review of Financial Studies* 26 (6): 1483–1525.

Greenwood, Robin, Andrei Shleifer, and Yang You. 2018. “Bubbles for Fama.” *Journal of Financial Economics* forthcoming:

Grossman, Richard S. 2002. “New Indices of British Equity Prices, 1870–1913.” *Journal of Economic History* 62 (1): 121–146.

Gutierrez, German. 2017. “Investigating Global Labor and Profit Shares.” Working paper.

Hall, Alan Ross. 1968. *The Stock Exchange of Melbourne and the Victorian Economy 1852–1900*. Canberra, ACT: Australian National University Press.

Hannah, Leslie. 2018. “The London Stock Exchange, 1869—1929: New Statistics for Old?” *Economic History Review* 71 (4): 1349–1356.

Hansen, Svend Aage, and Knud Erik Svendsen. 1968. *Dansk pengehistorie. 1: 1700–1914*. København: Danmarks Nationalbank.

Hoffmann, Walther G. 1965. *Das Wachstum der Deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts*. Berlin: Springer.

Holston, Kathryn, Thomas Laubach, and John C. Williams. 2017. “Measuring the Natural Rate of Interest: International Trends and Determinants.” *Journal of International Economics* 108 (S1): 59–75.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. “The Rate of Return on Everything, 1870–2015.” *Quarterly Journal of Economics* forthcoming:

Jordà, Òscar, Björn Richter, Moritz Schularick, and Alan M. Taylor. 2017. “Bank Capital Redux: Solvency, Liquidity, and Crisis.” NBER Working Paper 23287.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." In *NBER Macroeconomics Annual 2016, Volume 31*. Edited by Jonathan A. Parker Martin Eichenbaum. Chicago, Ill.: University of Chicago Press, pp. 213–263.

Jordà, Òscar, Moritz Schularick, Alan M. Taylor, and Felix P. Ward. 2018. "Global Financial Cycles and Risk Premiums." NBER Working Paper 24677.

Karabarbounis, Loukas, and Brent Neiman. 2018. "Accounting for Factorless Income." NBER Working Paper 24404.

King, Robert G., and Ross Levine. 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics* 108 (3): 717–737.

Knoll, Katharina, Moritz Schularick, and Thomas M. Steger. 2017. "No Price like Home: Global House Prices, 1870–2012." *American Economic Review* 107 (2): 331–352.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2017. "Income and Wealth Inequality in America , 1949–2013." CEPR Discussion Paper 20547.

Kuvshinov, Dmitry. 2018. "The Time Varying Risk Puzzle." unpublished.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46 (2): 285–332.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1997. "Legal Determinants of External Finance." *Journal of Finance*, 1131–1150.

Lamberton, Donald McLean. 1958. *Share Price Indices in Australia: An Examination of the Measurement and Significance of Changes in Share Prices, Together with New Indices for the Period July, 1936-December, 1957*. Law Book Company of Australia.

Le Bris, David. 2012. "Wars, Inflation and Stock Market Returns in France, 1870–1945." *Financial History Review* 19 (3): 337–361.

Le Bris, David, and Pierre-Cyrille Hautcoeur. 2010. "A Challenge to Triumphant Optimists? A Blue Chips Index for the Paris Stock Exchange, 1854–2007." *Financial History Review* 17 (2): 141–183.

Lettau, Martin, and Sydney Ludvigson. 2002. "Consumption, Aggregate Wealth, and Expected Stock Returns." *Journal of Finance* 56 (3): 815–849.

Lettau, Martin, Sydney C. Ludvigson, and Jessica A. Wachter. 2008. "The Declining Equity Premium: What Role Does Macroeconomic Risk Play?" *Review of Financial Studies* 21 (4): 1653–1687.

Levine, Ross, and Sara Zervos. 1996. "Stock Market Development and Long-run Growth." *World Bank Economic Review* 10 (2): 323–339.

López, Carlos Barciela, Albert Carreras, and Xavier Tafunell. 2005. *Estadıústicas Históricas De España: Siglos XIX–XX*. Madrid: Fundacion BBVA.

McGrattan, Ellen R., and Edward C. Prescott. 2005. "Taxes, Regulations, and the Value of U.S. and U.K. Corporations." *Review of Economic Studies* 72 (3): 767–796.

Mediobanca. Various years. "Indici e Dati Relativi ad Investimenti in Titoli Quotati."

Michie, Ranald C. 1988. "The Canadian Securities Market, 1850–1914." *Business History Review* 62 (1): 35–73.

Moore, Lyndon. 2010a. "Financial Market Liquidity, Returns and Market Growth: Evidence from Bolsa and Börse, 1902-–1925." *Financial History Review* 17 (1): 73–98.

Moore, Lyndon. 2010b. "World Financial Markets 1900–25." Working Paper.

Musacchio, Aldo. 2010. "Law and Finance c. 1900." NBER Working Paper 16216.

Nelson, William. 1999. "The Aggregate Change in Shares and the Level of Stock Prices." Federal Reserve Board Finance and Economics Discussion Series 1999-08.

Nyberg, Peter M., and Mika Vaihekoski. 2011. "Descriptive Analysis of Finnish Equity, Bond and Money Market Returns." BANK OF FINLAND DISCUSSION PAPER SERIES 14/2011.

Nyberg, Peter M., and Mika Vaihekoski. 2014a. “Descriptive Analysis of the Finnish Stock market: Part II.” BANK OF FINLAND DISCUSSION PAPER SERIES 10/2014.

Nyberg, Peter M., and Mika Vaihekoski. 2014b. “Equity Premium in Finland and Long-Term Performance of the Finnish Equity and Money Markets.” *Cliometrica* 8 (2): 241–269.

O’Sullivan, Mary. 2007. “The Expansion of the U.S. Stock Market, 1885–1930: Historical Facts and Theoretical Fashions.” *Enterprise & Society* 8 (3): 489–542.

Parum, Claus. 1997. “Det Danske Organiserede Aktiemarked i Historisk Belysning.” Copenhagen Business School Department of Finance Working Paper 97-6.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Harvard University Press.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. “Distributional National Accounts: Methods and Estimates for the United States.” *Quarterly Journal of Economics* 133 (2): 553–609.

Piketty, Thomas, and Gabriel Zucman. 2014. “Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010.” *Quarterly Journal of Economics* 129 (3): 1255–1310.

Rajan, Raghuram G., and Luigi Zingales. 2003. “The Great Reversals: the Politics of Financial Development in the Twentieth Century.” *Journal of Financial Economics* 69 (1): 5–50.

Rangvid, Jesper. 2006. “Output and Expected Returns.” *Journal of Financial Economics* 81 (3): 595–624.

Roe, Alan. 1971. *The Financial Interdependence of the Economy, 1957-1966*. London: Chapman & Hall.

Roine, Jesper, and Daniel Waldenström. 2012. “On the Role of Capital Gains in Swedish Income Inequality.” *Review of Income and Wealth* 58 (3): 569–587.

Ronge, Ulrich. 2002. *Die Langfristige Rendite Deutscher Standardaktien: Konstruktion eines Historischen Aktienindex ab Ultimo 1870 bis Ultimo 1959*. Frankfurt am Main: Lang.

Saez, Emmanuel, and Gabriel Zucman. 2016. “Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data.” *Quarterly Journal of Economics* 131 (2): 519–578.

Saint-Marc, Michèle. 1983. *Histoire Monétaire de la France (1800–1980)*. Paris: Presses Universitaires de France.

Schularick, Moritz, and Alan M. Taylor. 2012. “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008.” *American Economic Review* 102 (2): 1029.

Shiller, Robert J. 2015. *Irrational Exuberance*. Princeton, N.J.: Princeton University Press.

Solomou, Solomos, and Martin Weale. 1997. “Personal Sector Wealth in the United Kingdom, 1920–1956.” *Review of Income and Wealth* 43 (3): 297–318.

Sylla, Richard. 2006. “Schumpeter Redux: a Review of Raghuram G. Rajan and Luigi Zingales’s Saving Capitalism from the Capitalists.” *Journal of Economic Literature* 44 (2): 391–404.

Waldenström, Daniel. 2014. “Swedish Stock and Bond Returns, 1856–2012.” In *Historical Monetary and Financial Statistics for Sweden, Volume 2: House Prices, Stock Returns, National Accounts and the Riksbank Balance Sheet, 1860-2012*. Edited by Tor Jacobson Rodney Edvinsson and Daniel Waldenström. Stockholm: Sveriges Riksbank, and Ekerlids förlag, pp. 223–293.

Zucman, Gabriel. 2014. “Taxing Across Borders: Tracking Personal Wealth and Corporate Profits.” *Journal of Economic Perspectives* 28 (4): 121–48.

Chapter 4

Sovereigns Going Bust: Estimating the Cost of Default*

Joint with Kaspar Zimmermann

4.1 Introduction

In the summer of 2015, the Greek prime minister Alexis Tsipras had to decide whether to default on the country's sovereign debt or accept the conditions set by Greece's creditors. The decision was greatly complicated by the lack of agreement about what the economic consequences of a default would be. This lack of information points to a fundamental issue at the heart of economic models of sovereign debt. Because sovereign debt contracts are not directly enforceable, the existence of sovereign debt markets hinges on an indirect punishment mechanism in the form of default costs. And yet our empirical knowledge of these costs remains limited.

The gaps in empirical knowledge come from two main sources. First, there is little agreement on how costly, in general, sovereign default is. Defaulting countries experience a very wide range of economic outcomes. And existing empirical studies place the default cost anywhere between zero (Levy-Yeyati and Panizza, 2011) and a fifth of a country's output (De Paoli, Hoggarth, and Saporta, 2009; Furceri and Zdzienicka, 2012).[1] Second, we do not yet know what exactly generates the default

* We are grateful for helpful comments from Christian Bayer, Tobias Berg, Benjamin Born, Yao Chen, Òscar Jordà, Florian Kirsch, Gernot Müller, Moritz Schularick, Alan Taylor, Felix Ward, Mark Wright, and to Christoph Trebesch for sharing data. We would also like to thank two anonymous referees, the editor Eric Leeper and seminar participants at the University of Bonn, the European Central Bank, RGS Doctoral Conference and ADEMU Toulouse Summer School. This work is part of a larger project kindly supported by a research grant from the Bundesministerium für Bildung und Forschung (BMBF). Kuvshinov: University of Bonn, email: dmitry.kuvshinov@uni-bonn.de. Zimmermann: University of Bonn, email: kaspar.zimmermann@uni-bonn.de.

1. See Section 4.2 for a more detailed review of the existing literature. The Furceri and Zdzienicka (2012) estimate refers to the medium-term cost of a sovereign crisis occurring in isolation, which is larger than their baseline estimate (10% of output).

cost. A number of mechanisms operate at a microeconomic level, but it is not clear which ones are important in generating the macroeconomic cost. Our paper seeks to address both of these issues. We estimate the overall cost of default using up-to-date econometric methods and data, and investigate which transmission channels are most important in generating and amplifying this cost.

The disagreement between empirical cost estimates can largely be traced back to differences in the method and data used. Because the decision to default is taken contingent on the country's economic conditions, naive cost estimates which do not account for this endogeneity may be biased. But given the lack of instruments and other measures of exogenous variation in defaults, the cost estimate will generally vary depending on how well the estimation method accounts for such endogenous selection into defaulters. In addition to this, default can be defined in several different ways and, being rare events, the data used in the estimation can suffer from a small sample problem. This means that the cost estimate will also be sensitive to the quality and representativeness of the sample data.

Our first key contribution is to provide a conclusive best-practice estimate of the macroeconomic cost of default which relies on up-to-date comprehensive methods and data. To deal with endogeneity, we introduce a novel econometric method – the "inverse propensity score weighted regression adjustment" (IPSWRA) of Jordà and Taylor (2016) – to the literature of default costs. This two-step procedure first rebalances the sample of defaulters and non-defaulters to mimic a situation where these were selected at random, and then applies local projections to the rebalanced sample to estimate the default cost over a horizon of 10 years. To make sure our results are not biased by the data we use, we apply this method to a new dataset which combines and extends 5 alternative default definitions most commonly used in the literature. These annual data span the period 1970 to 2010, encompassing 112 countries and 92 external defaults in our preferred specification.

IPSWRA offers several advantages relative to other methods, largely owing to the lack of restrictions it places on the data. It allows for non-linearities in selection and time response of GDP to default, is "doubly-robust" to misspecification, and enables us to compute both short, medium and long run default costs. IPSWRA's flexibility is especially important when dealing with sovereign default, because defaults are rare events accompanied by sudden shifts in economic outcomes, with negotiations often taking years and costs playing out over a prolonged period of time. The methodology relies on one key identifying assumption – "selection on observables" – which requires that our control set reflects the policymakers' information on the eve of default. To ensure this, we consult a broad range of sources to constuct a comprehensive set of economic, financial, political and crisis variables which can affect the default decision, and complement these with credit ratings and IMF forecast data in a series of robustness tests.

Our second key contribution is to assess which factors are empirically more important in determining and amplifying the cost of default. To do this, we first look at how much the cost is amplified by other crisis events, such as banking and currency crises. We then investigate which types of economic activity – for instance, consumption, trade or investment – are most affected by the default. Here again, the flexibility of the IPSWRA method comes into play. By simply redefining the treatment or outcome variable definition, we can compute the state-contingent impact of default on different parts of the economy in a manner that is consistent with our overall macroeconomic cost estimate.[2] The end result is a data-driven estimate of the sovereign default cost, its amplification and transmission channels, all computed within the same doubly-robust semi-parametric econometric framework.

Our first key finding is that the sovereign default cost is sizeable and persistent, but not permanent. Default reduces GDP by 2.7% on impact and continues to drag down output over the subsequent years. During the first five years after default, the cost gradually increases, peaking at 3.7% of GDP, but it largely disappears by year 10. This stands in contrast to much of the emerging markets literature that finds largely permanent costs of default and other crises (Cerra and Saxena, 2008; Furceri and Zdzienicka, 2012). Making use of our comprehensive cross-country dataset, we make sure that this finding is robust to using alternative default definitions, and after controlling for expectations encompassed in forward-looking variables.

Our second key finding is that sovereign-banking spillovers, trade frictions and financial autarky play a key role in generating the cost of default. The cost doubles if the sovereign default is followed by a systemic banking crisis. In this case GDP drops by 9.5% after the first three years alone. The bulk of the default cost is driven by sharp declines in investment and credit, which are particularly stark during joint sovereign-banking crises. Consistent with the importance of the banking channel, we also find that countries with more developed financial systems experience higher default costs. Defaulters also undergo a sizeable and rapid external adjustment. After a default gross trade collapses, with imports in particular falling sharply as the country reduces its external dependence by increasing net exports. Both the size of the adjustment and the magnitude of the output cost, as well as the pre-default current account imbalances, are much higher under pegged exchange rate regimes. These results point to high output costs of financial autarky, especially when the necessary external adjustment is difficult to attain.

These findings have important implications for the understanding of sovereign default and its aftermath. First, we show that even after endogenising the decision to default there is still a significant and persistent – but not permanent – sovereign default cost. Second, this cost is accompanied by a substantial reallocation of resources within the economy which, in presence of adjustment frictions, could generate the observed output cost. Third, the magnitude of the cost is largely contingent on two

2. The flexibility of local projections has also made them attractive to the literature on fiscal multipliers (see Auerbach and Gorodnichenko, 2012; Owyang, Ramey, and Zubairy, 2013; Ramey and Zubairy, 2014).

factors: banking system conditions and the feasibility of the necessary external adjustment. While defaults under flexible exchange rates incur little trade disruption and carry a near-zero cost similar to that found by Levy-Yeyati and Panizza (2011), the cost of default followed by systemic banking crises exceeds that of most other "extreme events" in emerging and advanced economies (Cerra and Saxena, 2008; Jordà, Schularick, and Taylor, 2013).

How do our results map into the theory literature on sovereign risk? The default cost estimate is higher than the temporary 2% endowment penalty typically assumed in the literature (see, for example Aguiar and Gopinath, 2006; Yue, 2010), but lower than the output cost attributed to the endogenous reinforcement mechanism in Mendoza and Yue (2012). The increase in net exports and the collapse in gross trade indicate that autarky costs – the key mechanism in most sovereign default models – do play an important role in explaining the cost of default. However, our findings suggest that banking distress acts as a key amplifier and propagator of default costs. A better understanding of this second mechanism and its interaction with autarky costs would enhance both the intuitive appeal and the applicability of sovereign default models.

This paper is structured as follows: the next section reviews the theoretical and empirical literature on default costs. We then describe the methodology and data used in our estimation, and present our results. A final section concludes.

4.2 What we know about sovereign default costs

Theoretical economic models assume that sovereign default is costly. Because sovereign debt contracts are not enforceable, defaulters have to face a credible punishment in order to ensure debt repayment and facilitate sovereign borrowing in the first place. The classic analysis in Eaton and Gersovitz (1981) assumes that this punishment takes the form of a permanent exclusion from international borrowing markets, or autarky. But even though autarky is sufficient to sustain sovereign borrowing in theory, it carries no direct output costs and only affects the government's ability to smooth consumption over time. This limits the amount of punishment in the canonical model and results in very low levels of sustainable debt (Aguiar and Gopinath, 2006). It also predicts that countries would tend to default during times of good economic performance or high productivity, which is the opposite of what we tend to observe empirically (Tomz and Wright, 2007).

To get around these problems, theoretical models have introduced a number of modifications that make default more costly. A number of papers – for example Aguiar and Gopinath (2006) and Yue (2010) – add a direct output cost, typically parametrised at 2% of the country's economic potential, in order to achieve higher sustainable debt levels. Some – such as Arellano (2008) – further assume that this direct cost increases with output, which reduces the incentive to default during good times. More recent work has suggested ways to microfound this direct output cost.

Table 4.2.1. Existing estimates of the cost of sovereign default

Paper	Default cost, % GDP		Method
	First year	Medium term	
Historical unconditional estimates:			
Reinhart and Rogoff (2011a)	3–4%†	5%†	Average path of GDP
Tomz and Wright (2007)	1.6%	1.4%	Deviation from HP trend
Conditional estimates using more recent data:			
De Paoli, Hoggarth, and Saporta (2009)	5.5 – 10.5% †† per year		Fixed effects panel + counter-factual comparison. Defaults with high arrears only.
Furceri and Zdzienicka (2012)	5.6%	10%	Two-stage GMM panel. Also local projections. Sovereign crises only.
Borensztein and Panizza (2008)	2.6%	not sig.	Fixed effects panel + controls
Levy-Yeyati and Panizza (2011)	not sig.	not sig.	Fixed effects panel, quarterly data

Notes: Not sig. means "not significant". All estimates are based on annual data unless otherwise specified.
† We use the estimates determined by Reinhart and Rogoff (2011a) for GDP growth after a default on external debt, and subtract a 2% annual GDP growth trend to arrive at the estimate in the table.
† † De Paoli, Hoggarth, and Saporta (2009) median cost estimates, baseline results. The average cost is higher (12 – 13% GDP).

Mendoza and Yue (2012) show that post-default autarky can harm firms' production capabilities because they would not be able to import the necessary intermediate inputs, whilst Gennaioli, Martin, and Rossi (2014), Bocola (2016), Perez (2015) and Sosa-Padilla (2018) show that default can inflict damage on the country's banking system, either via write-offs on sovereign bonds held by banks, or contagion to bank funding markets. This in turn reduces bank lending, investment and output. Quantitative theoretical models of sovereign default leave two main open questions for the empirical literature: first, what is the overall cost – or direct penalty – of sovereign default; and second, what are the channels through which a default affects economic performance.

Empirical studies focussing on the channels through which default affects the economy tend to find some evidence in support of autarky. Cruces and Trebesch (2013) show that defaulters subsequently experience higher credit spreads and outright capital market exclusion, and that the penalties are higher, the less favourable the default is for creditors. Turning to the direct trade channel, Rose (2005) and Borensztein and Panizza (2010) document a negative impact of default on exports and export-oriented firms. Hébert and Schreger (2017) exploit legal rulings in a sovereign debt case to estimate the causal response of Argentinian equity prices to rising default probabilities. They show that equity prices fall on average after court rulings and find stronger effects for export-oriented firms and banks. Gennaioli, Martin, and Rossi (2018), Acharya, Eisert, Eufinger, and Hirsch (2018) and Andrade and Chhaochharia (2017) also find evidence in support of the banking channel: at time of sovereign stress, domestic banks with larger sovereign debt exposures tend to reduce lending, while firms reliant on these banks lower investment and sales, and experience drops in stock prices. Borensztein and Panizza (2008) find an increased likelihood of systemic banking crises after sovereign default, while Reinhart and Rogoff (2011) show that sovereign debt crises are often preceded by banking crises in historical data.

A number of studies have also examined potential channels for amplification of the default cost, but these have almost exclusively focused on the negotiation process itself. Overall, the evidence presented by Trebesch and Zabel (2017) and Asonuma and Trebesch (2016) suggests that a pre-emptive, or more collaborative approach to negotiation results in lower default costs. More recent studies following our paper have further looked into these links. Asonuma, Chamon, and Sasahara (2016) show that pre-emptive renegotiation attenuates the impact of default on trade and output and Balteanu and Erce (2018) provide further evidence on the interactions of debt and banking crises.

Studies of overall default costs have tackled this problem in a number of ways, with the results summarised in Table 4.2.1. Whilst historical studies (see Reinhart and Rogoff, 2011a; Tomz and Wright, 2007) have documented a general negative correlation between default and output growth, other studies based on more recent data have attempted to disentangle the effect of sovereign default from that of other observed confounders. The range of these conditional sovereign default cost estimates, however, is extremely broad. At one end, there are the estimates of De Paoli, Hoggarth, and Saporta (2009) and Furceri and Zdzienicka (2012) who find sovereign default costs of 6% or more of a country's GDP on impact, and a permanent cost upwards of 10% GDP in the longer term. At the other end, Levy-Yeyati and Panizza (2011) who base their findings on quarterly data, find no default cost at all. Lying between these two extremes is Borensztein and Panizza (2008)'s estimate of a 2.6% GDP cost on impact. This dispersion among individual estimates, taken together with the wide variety of methods and data used, makes it difficult to make inferences about the size of the default penalty.

Our study complements the existing literature in two ways. First, we apply an up-to-date econometric method to a comprehensive dataset in order to provide a more conclusive estimate of the overall default cost. Second, we study the different channels that may transmit the sovereign default cost through the economy within the same empirical framework, which allows us to bridge some of the gaps between the literature on overall default costs and that on the individual transmission channels.

Table 4.3.1. Characteristics of treatment and control groups

	Treatment (defaulters)	Control (non-defaulters)	Difference significant?
GDP growth	-1.76	1.70	Yes(1% level)
External public debt/GDP	43.68	47.10	No
Inflation	24.75	17.13	Yes(5% level)
Openness	62.92	79.89	Yes(1% level)
Governance quality score (Polity)	-1.80	-0.02	Yes(5% level)
Banking crisis probability	0.10	0.05	Yes(5% level)
Currency crisis probability	0.12	0.08	Yes(10% level)
War intensity (scale 0 – 20)	1.02	0.96	No
Coup probability	0.09	0.06	No

Notes: All values refer to the year preceding default, and in the case of banking and currency crisis probabilities, to two years before default. Openness is the ratio of gross imports and exports to GDP. Governance quality is scored on a scale from -10 to 10, with a higher score meaning better governance. All ratios are presented as percentage points, all growth rates in percent. The third column tests the equality of the respective means between the treatment and the control group. GDP growth and inflation are winsorized at the 2% level to exclude outliers.

4.3 Estimating sovereign default costs

To calculate the cost of sovereign default, we need to compare two counterfactual scenarios: one where the representative country in our sample defaulted and the other where it did not. If the default decision was random – or exogenous – it would be sufficient to compare the average performance of defaulters to that of non-defaulters. But countries do not default at random. Table 4.3.1 shows that the decision to default is endogenous to a number of observable variables: for example, defaulters tend to have higher debt and lower growth, with many still recovering from another crisis – all factors that could suppress future economic performance. A simple means comparison would therefore conflate the impact of these confounding factors with that of the default itself.

To negotiate this problem, we need to capture the exogenous variation of default decisions. We cannot do this by means of an experiment; moreover there are no apparent historical natural experiments or plausible exogenous instruments when it comes to analysing sovereign defaults. Therefore this analysis proceeds in a different direction: we accept that the default decisions in our dataset are endogenous, but we seek to explicitly model this endogenous decision process and account for it in our estimation. Modelling the default decision allows us to effectively reverse-engineer it and rebalance the sample "as if" it were taken at random. To do this, we use the inverse propensity score weighting methodology developed by Angrist, Jordà, and Kuersteiner (2017) and Jordà and Taylor (2016), described in the following section.

4.3.1 Estimation procedure

The inverse propensity score weighting (IPSW) estimation proceeds in two stages. The first stage models the default decision by estimating a policy propensity score for each observation in our sample. This score is simply the likelihood of default predicted by a logit model, as follows:

$$\widehat{PD}_{i,t} = \Lambda(X^P_{t-1}, \tilde{X}^P_{t-1}, \tilde{X}^P_{t-2}, \hat{\beta}) \tag{4.3.1}$$

Here $\widehat{PD}_{i,t}$ is the predicted default probability for country i at time t, conditional on a set of predictor variables $\{X^P, \tilde{X^P}\}$; some (X^P) included with one lag and others $(\tilde{X^P})$ with two. Λ is the logistic distribution function.

The second stage rebalances the sample to mimic a setting where the default decision was random. Compared to a random sample, our group of defaulters contains too many cases where countries defaulted for endogenous reasons such as having low growth or high debt. The control group, on the contrary, will contain countries with very good debt fundamentals and a low likelihood of default. We can estimate the extent of this non-random selection using the logit in equation (4.3.1). A highly endogenous default would be forecastable based on observables, and attain a high predicted default probability $\widehat{PD}_{i,t}$ in the logit. A highly endogenous control group observation would, on the contrary, have a low probability of default. To correct for this non-randomness, the IPSW procedure rebalances the sample by giving the more endogenous observations in each group a low weight in the estimation. For this purpose, the weights – or inverse propensity scores – are $1/\widehat{PD}_{i,t}$ for the defaulter (treatment) group, and $1/(1-\widehat{PD}_{i,t})$ for the non-defaulter (control) group.

Once the sample is rebalanced, the cost of default is measured as its "average treatment effect": the average difference in potential outcomes of defaulters and non-defaulters across the sample. Potential outcomes are computed using a conditional local projection forecast over a horizon of 10 years (Jordà, 2005):

$$\Delta y_{i,t+h} = \alpha_i + \theta_h \delta_{i,t} + \Gamma_{h,1} X^C_{i,t-1} + \tilde{\Gamma}_{h,1} \tilde{X}^C_{i,t-1} + \tilde{\Gamma}_{h,2} \tilde{X}^C_{i,t-2} + \varepsilon_{i,t} \quad h \in \{0, \ldots, 9\} \tag{4.3.2}$$

Here $\Delta y_{i,t+h} = y_{i,t+h} - y_{i,t}$ is the conditional forecast of the cumulative growth in the outcome variable (GDP), for years t to $t+9$. α_i are country fixed effects, $\delta_{i,t}$ is the treatment variable – in our case a simple 0/1 sovereign default dummy – and $X^C, \tilde{X}^C$ are the control variables, again included up to two lags. We follow Jordà and Taylor (2016) and use a richer set of predictors in stage 1 (4.3.1) than controls in stage 2 (4.3.2), hence $X^C \subset X^P$ and $\tilde{X}^C \subset \tilde{X}^P$. $\varepsilon_{i,t}$ is the constant-variance zero-mean error term. Standard errors are clustered by country. Accounting for correlation across time and in the cross-section as in Driscoll and Kraay (1998) would reduce the local projection standard errors and as a result the errors reported in the tables should be viewed collectively as an upper bound.[3]

The estimation in (4.3.2) runs 10 separate regressions, one for each horizon h, and uses them to compute counterfactual forecasts of future GDP in the event of default or continued repayment, for each observation in our sample. The average treatment effect of sovereign default is then the weighted difference between these potential outcomes, computed on the rebalanced sample where each observation is weighted by the inverse of its propensity score:

$$ATE_h(\delta)^{\text{IPSWRA}} = \frac{1}{n_{\text{Def}}} \sum_i \sum_t \frac{\Delta\hat{y}_{i,t+h} * \delta_{i,t}}{\widehat{PD_{i,t}}} - \frac{1}{n_{\text{NoDef}}} \sum_i \sum_t \frac{\Delta\hat{y}_{i,t+h} * (1-\delta_{i,t})}{1-\widehat{PD}_{i,t}} \qquad (4.3.3)$$

Here, $\Delta\hat{y}_{i,t+h}$ is the forecast obtained by estimating (4.3.2), $\delta_{i,t}$ is the default dummy used to separate observations into the treatment and control groups (defaulters and non-defaulters), and $(1/\widehat{PD}_{i,t})$ and $1/(1-\widehat{PD}_{i,t})$ are the inverse propensity score weights for the two groups. We truncate the weights at 10 as recommended by Imbens (2004). $ATE_h(\delta)^{\text{IPSWRA}}$ is the average treatment effect of default, again computed over the ten-year horizon. In our setting, the treatment δ is a dummy variable. In this case, the treatment effect $ATE_h(\delta)$ equals the regression coefficient θ_h^w on the sovereign default dummy δ in a weighted local projection regression, where the weights IPW correspond to the inverse propensity scores:

$$\begin{aligned}\Delta y_{i,t+h} = \alpha_i + \theta_h^w \delta_{i,t} * IPW_{i,t} + \Gamma_{h,1} X_{i,t-1}^C * IPW_{i,t} + \tilde{\Gamma}_{h,1} \tilde{X}_{i,t-1}^C * IPW_{i,t} + \\ \tilde{\Gamma}_{h,2} \tilde{X}_{i,t-2}^C * IPW_{i,t} + \varepsilon_{i,t}, \quad (4.3.4)\end{aligned}$$

with $IPW_{i,t} = 1/\widehat{PD_{i,t}}$ for defaulters and $IPW_{i,t} = 1/(1-\widehat{PD_{i,t}})$ for non-defaulters.

Combining the local projection methodology with inverse propensity score weighting gives us the inverse propensity score weighted regression-adjusted (IPSWRA) estimator introduced by Jordà and Taylor (2016). This estimator has a number of advantages compared to other methods used in the literature. Most of existing estimates of sovereign default costs rely on OLS or GLS, while some recent papers have also used LPs without the IPSW adjustment. Compared to the OLS and GLS estimates, the IPSWRA framework produces a direct unbiased estimate of the medium- and long-run cost of default, i.e. the “average treatment effect” from t to $t+h$.[4] This is crucial, since the fallout from macroeconomic crises in emerging markets tends

3. The Driscoll and Kraay (1998) procedure is not well specified for an IPSWRA estimator. We therefore report country-clustered errors for all specifications to ease comparability.

4. One potential alternative is to include lagged default dummies in an OLS, as, for example in (Borensztein and Panizza, 2008). Unlike LPs, this does not allow us to estimate the ATE of sovereign default, because each lagged treatment dummy is conditioned on contemporaneous controls at t, and not lagged controls at $t-h$, which confounds the impact of past default on current control variables, and current outcome variable.

to be quite persistent, both in sovereign default models and in the data. This dynamic estimate is also robust to misspecification, for a number of reasons. The local projection in (4.3.2) imposes little structure on the data, and allows the response $ATE_h(\delta)^{\mathrm{IPSWRA}}$ to vary in a non-linear manner over the forecast horizon h – unlike, say, a VAR which carries a linear structure of the form $ATE_h(\delta)^{\mathrm{VAR}} = F * ATE_{h-1}(\delta)$, where F is some coefficient matrix. The propensity score weighting additionally allows the selection into defaulters to be a non-linear function of predictors and controls. With both the default decision and the cost likely subject to a multitude of threshold effects, accounting for non-linearities in both stages of the estimation is crucial for obtaining an unbiased sovereign default cost estimate.

The combination of local projections and propensity score weighting makes the estimator "doubly robust" to regression misspecification: it is unbiased as long as at least one of the regression stages (4.3.1) and (4.3.2) is specified correctly. This regression framework is also highly flexible, and allows us to account for a number of state dependencies, types of treatment and outcomes within the same empirical framework. Looking into these state dependencies is key to understanding what, ultimately, drives the cost of default. For example, to see if the banking and trade channels are important, we can estimate the export cost of defaults that are followed by systemic banking crises by changing the definition of δ to a "default and banking crisis" scenario, and the definition of y – to exports rather than GDP. Taken together, the IPSWRA methodology offers a data driven, flexible, robust semi-parametric approach that gives us a best-practice estimate of sovereign default cost, and helps shed light on its underlying determinants.

4.3.2 Identification

A causal interpretation of our estimates relies on one crucial assumption: selection on observables (see, for example Imbens, 2004; Jordà and Taylor, 2016; Angrist, Jordà, and Kuersteiner, 2017). Conditional on the propensity score predictors in (4.3.1) and local projection controls in (4.3.2), the decision to default δ_t should be independent of potential outcomes – denoted here as $\Delta y_{i,t+h}(\delta)$ – which capture counterfactual future GDP growth in the event of default ($\delta_t = 1$) or continued repayment ($\delta_t = 0$), for the horizon of 10 years. This assumption can be summarised as

$$\Delta y_{i,t+h}(\delta) \perp\!\!\!\perp \delta_{i,t} \mid X^P_{t-1}, \tilde{X}^P_{t-1}, \tilde{X}^P_{t-2}, \beta \quad h \in \{0, \ldots, 9\}, \tag{4.3.5}$$

where $X^P_{t-1}, \tilde{X}^P_{t-1}, \tilde{X}^P_{t-2}$ and β are the policy score predictors and parameters from (4.3.1).

In practical terms, "selection on observables" means that our control and predictor set should be be rich enough to explain the variation in default decisions that is endogenous to future growth prospects, such that any remaining variation is independent of growth outcomes. The main advantage of this identifying assumption is that it does not rely on any form of exclusion restrictions. Put differently, all the variables in our dataset can be endogenous, from the decision to default to export growth and other crisis events. What matters is that we, in a sense, capture the full information set of the policymaker: conditional on all the endogenous variables we can observe at time t, there should be no systematic deviations in default decisions that are correlated with future GDP in periods t to $t+h$. This non-reliance on exclusion restrictions makes IPSWRA ideal for estimating the cost of default in a broad macroeconomic setting. Even though it may be possible to find credible exogenous instruments for sovereign defaults in individual case studies – as, for example, argued by Hébert and Schreger (2017) for the legal disputes surrounding Argentina's default in 2001 – this is not an option in a richer cross-country setting. This means that any cross-country analysis of sovereign default costs, including the popular OLS and GLS methods, has to rely on some form of selection on observables. IPSWRA is simply the most robust method of extracting default cost information from a set of observable endogenous variables.

The "selection on observables" assumption is demanding, and difficult to satisfy completely: even if our dataset included all observable data on the eve of default, policymakers may still have access to private information about future economic prospects that influences their decisions. That being said, there are several things we can do to ensure that our estimates come as close as possible to fulfilling this assumption and identifying the causal impact of sovereign default. These basically come down to definitions of δ and X in equation (4.3.5). To allow for selection on observables, the decision to default δ has to be as exogenous as possible, and X should come close to capturing the full information set of the policymaker. We take several steps to ensure these conditions are met. First, our X is constructed to capture all the main determinants of sovereign default identified in the existing literature (see, for example Manasse and Roubini, 2009) – from macroeconomic to international, macro-financial and political factors. In a series of robustness checks, we further extend X to include "softer" forward-looking information on sovereign credit ratings and GDP growth forecasts. For δ, we use the default definition that is least dependent on the country's current and future economic performance. We discuss the choice in detail in Section 4.4 but, in brief, we include all instances where a sovereign debt payment was missed and recognised as such by the rating agencies, and do not focus on extreme or crisis events, which are likely to be more endogenous. We also investigate whether our results hold across a number of different definitions of δ, and for different data subsamples and country groups, which acts as a check for variation in unobservable characteristics of the treatment and control groups.

4.4 A dataset of sovereign defaults and their drivers

To compute the default cost estimate, we require data on the sovereign default decisions δ, their economic outcomes y, and the conditioning set X that informs us about the state of the economy before the default takes place. The first challenge lies in

Table 4.4.1. Alternative definitions of sovereign default

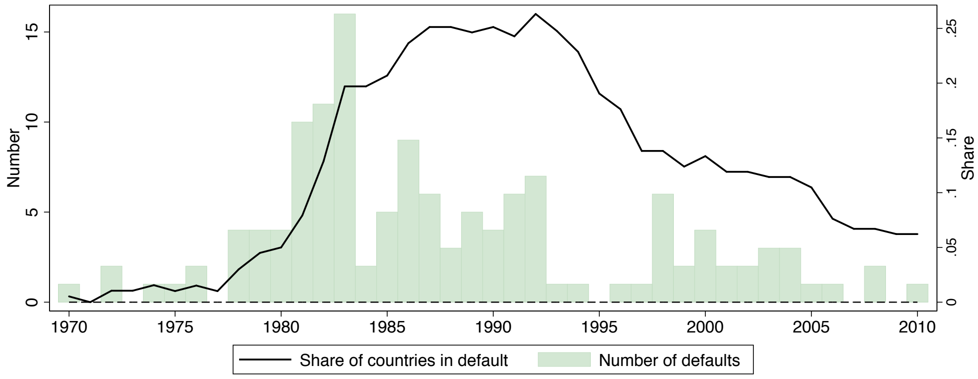
Source	Definition	Key criteria
Standard & Poor's (baseline)	Failure to make a payment; distressed restructurings	Legal
Reinhart and Rogoff (2011b)	Similar to *Standard & Poor's*, but using slightly different sources	Legal
Beim and Calomiris (2000)	As above, but group spells of defaults less than 5 years apart together, and ignore political defaults	Legal + duration
Laeven and Valencia (2008)	Failure to make a payment; distressed restructurings; case-by-case selection of crisis episodes	Legal + extent of crisis
Detragiache and Spilimbergo (2001)	Non-payment arrears > 5% of total debt, and distressed restructurings	Legal + size of arrears

defining what constitutes a sovereign default. The literature has proposed several such definitions, which are summarised in Table 4.4.1. The simplest way to define default is in strict legal terms, as a failure to honour the original conditions of the sovereign debt contract. This involves either missing a payment, or changing the contractual terms as part of a distressed restructuring. *Standard & Poor's* (Beers and Chambers, 2006), Reinhart and Rogoff (2011b) and Reinhart and Trebesch (2016) broadly follow this default definition.

A number of authors have proposed modifications to the simple legalistic definition which effectively make it more stringent. The *Standard & Poor's* definition attaches the same significance to short repayment delays that are relatively unsubstantial and defaults that involve large financial distress for debtors and creditors. To exclude these less substantial default episodes, Beim and Calomiris (2000) only count repayment delays of six months or more, and combine default spells that occur within five years of each other. Beim and Calomiris (2000) also exclude defaults that occurred for political motives. Laeven and Valencia (2008) use a somewhat less precise case-by-case approach to select only those, more severe, defaults that can be classified as "debt crises". Detragiache and Spilimbergo (2001) only count those repayment delays where a country accumulated arrears amounting to 5% or more of their external public debt or when there is a restructuring agreement with commercial creditors listed in the Global Development Finance.

To fulfil the "selection on observables" assumption, our default definition has to be as neutral or exogenous as possible. The definitions that focus on more severe defaults will by their nature be more endogenous, and are likely to select those events that have relatively less favourable economic outcomes. We therefore use the simple legal definition, as categorised by *Standard & Poor's*, as baseline. Because the *Standard & Poor's* data beyond 2006 and before 1975 are not systematically available, we complement these with estimates of Reinhart and Rogoff (2011b) and

Figure 4.4.1. Frequency of sovereign defaults since 1970



Notes: Data are based on our baseline default definition, which follows *Standard & Poor's*. Share of countries in default is relative to all countries in our sample, including advanced economies but excluding countries classified by Reinhart and Trebesch (2016) as not being independent at the time.

Reinhart and Trebesch (2016) for the corresponding years. Throughout, we only consider defaults on external debt. Even though domestic debt is important from a broader perspective (Reinhart and Rogoff, 2011a), defaults on these obligations are more difficult to define, and are more likely to be endogenous to the country's economic conditions.[5] Our baseline default definition therefore consists of all instances of missed repayments and distressed restructurings of external government debt to private creditors that took place between 1970 and 2010.

To gain a comprehensive picture of sovereign default costs, we also apply our baseline estimation to the four alternative default definitions listed in Table 4.4.1 (See Section 4.5.2 and Appendix Section 4.B.4.1). To do this, we extend each of these definitions to cover the period 1970 – 2010, using data from *Standard & Poor's*, Reinhart and Rogoff (2011b), Reinhart and Trebesch (2016) and Beers and Nadeau (2015). Appendix Section 4.A.2 and Table 4.A.1 provide further details on the data construction, and Appendix Figure 4.A.1 provides a timeline of sovereign defaults under the five alternative definitions for each country in our sample.

Figure 4.4.1 shows the frequency of sovereign default events from 1970 to today under our baseline definition. The teal bars show the number of defaults in each year, and the solid line – the share of countries in default. The in-default share is the ratio of countries that have newly defaulted, or a still negotiating a past default, to all independent countries.[6] Appendix Figure 4.A.1 shows the corresponding trends

5. For example, domestic defaults can take form of high inflation as well as outright debt repudiation, which creates difficulties both in terms of definition and the endogeneity of such events.

6. We use the classification of Reinhart and Trebesch (2016) to exclude all countries that are not independent in a certain year, but we additionally include countries that are not covered in the Reinhart and Trebesch (2016) dataset as part of the total, which is why our in-default share is somewhat lower than that of Reinhart and Trebesch (2016). See Appendix 4.A.2 and Figure 4.A.1 for further detail.

for the four alternative default definitions, which paint a similar picture to Figure 4.4.1. Defaults peak at 10–15 per year during the 1980s Latin American debt crisis. Many of these defaulters continue the distressed debt negotiations until well into the 1990s, such that between 1985 and 1995, around one in every four countries is in default on its external debt obligations. Defaults become less frequent after the early 1990s, averaging less than 5 per year, and the in-default share falls to 5–10 percent. Over the whole sample, sovereign default is a relatively regular occurrence: at any point in time, between 5% and 25% of countries are in default on their external debt obligations. Yet, there is little consensus on the economic costs of these events and their underlying drivers. To approach these two questions, we turn to the IPSWRA regression framework described in Section 4.3.

The data sources for each variable in our regression are described in Appendix Table 4.A.1. The treatment variable δ is the sovereign default dummy, set to equal 1 in the first year of a default and zero otherwise. The outcome variable y is equal to cumulative GDP growth and its components – i.e. consumption, investment, government spending and net exports. We use a consistent sample throughout our estimation, which means that for every default, we have data on economic outcomes 10 years ahead, and the full set of the conditioning variables. This reduces the number of defaults considered relative to Figure 4.4.1, and means that we only include defaults up to 2001, which uses the data on outcomes up to year 2010. In line with *Standard & Poor's*, we treat each default or distressed resturucturing as a new event, even if it is part of a serial default spell: for example, the repeated debt restructurings by Uruguay during the 1980s are recorded as three separate default events with $\delta = 1$ in 1983, 1987 and 1990. Finally, we allow the 10-year treatment windows to overlap across default: for example, the treatment effect of the 1983 Uruguay default will include years 1987 and 1990 within the 10-year spell. This precludes us from making judgements about potential default outcomes which could violate the "selection on observables" assumption. The Beim and Calomiris (2000) definition of δ minimises such potential for overlap by requiring a minimum 5 year distance between the end of one and the beginning of another default episode. Appendix Table 4.A.2 lists the defaults included in our sample (92 in total), and Appendix Table 4.A.4 lists the countries and years included in the regression.

To choose the set of control variables, we follow existing literature in Jordà and Taylor (2016) and Imbens (2004), which suggests a rich set of predictors in Stage 1, complemented by a smaller set of controls in Stage 2. The Stage 1 predictors include all variables that help forecast sovereign default, as established, for example by Manasse and Roubini (2009) and Manasse, Roubini, and Schimmelpfennig (2003). Stage 2 controls are those variables that both help predict defaults and are likely to affect future economic outcomes. Given the large number of control variables relative to default occurrences, we limit the number of lags to 2, and only include 1 lag for those variables where the second lag is insignificant in statistical and economic terms. Increasing the number of lags does not generally affect the size of the estimated coefficients, but reduces their precision. We summarise these conditioning variables below.

Controls and predictors:. X^C. These variables enter both the logit in Stage 1, and the LP in Stage 2. *Macroeconomic controls* capture the fact that defaulting countries tend to have low growth, and often accumulate external imbalances, hence we include GDP growth, level and deviation from trend, inflation and a host of trade-related variables such as terms of trade and the current account balance. *Debt controls* capture the fact that defaulting countries tend to have high levels of debt, and there is also evidence linking debt levels to future growth. *Political controls* capture the quality and changes in governance, which should affect long-run growth through institutions, and the default decision through policymaker preferences and constraints. *Crisis controls* allow us to condition on systemic banking, currency or political crises occurring prior to default, which tend to both trigger poor GDP growth and increase the default probability. *Soft information* on sovereign credit ratings and growth forecasts reduces our sample size, but provides an important robustness check that better captures sovereign distress and growth prospects.

Predictors only:. X^P not in X^C. These capture additional default predictors that are connected to financial rather than macroeconomic conditions, and global financial factors. *Debt and financing conditions* include country-specific short-term refinancing needs, and measures of global risk appetite and funding costs. Because the logit regression does not include country fixed effects, we complement these with *country-specific factors* which affect the likelihood of default, such as default history and continent dummies.

The set of controls is substantially broader than that used in the existing literature, which typically relies on a subset of our macroeconomic controls, with some papers also conditioning on a preceding banking crisis (Borensztein and Panizza, 2008; De Paoli, Hoggarth, and Saporta, 2009; Levy-Yeyati and Panizza, 2011; Furceri and Zdzienicka, 2012). This study expands the typical conditioning set by adding additional debt, political and crisis controls to the LP in stage 2, utilising the power of the full set of macro-financial predictors from the literature on predicting sovereign debt crises (Manasse, Roubini, and Schimmelpfennig, 2003; Manasse and Roubini, 2009) in the stage 1 logit, and performing additional robustness checks using forward-looking rating and forecast variables. A more detailed discussion of why we include each variable, and its expected impact on the likelihood of default and future growth is provided in Appendix Table 4.A.3. Taken together, the relatively neutral default definition and the rich conditioning set help us fulfil the "selection on observables" assumption and ensure that our empirical analysis provides a robust and reliable estimate of the cost of sovereign default.

4.5 The cost of default

What is the cost of sovereign default? Table 4.5.1 (bottom row) and Figure 4.5.1 (solid black line) show our baseline IPSWRA default cost estimate, computed by applying the methodology described in Section 4.3 to the dataset described in Section 4.4. To put our findings in perspective, we also present an unconditional cost estimate that does not make any adjustment for endogeneity beyond controlling for time-invariant cross-country differences in GDP growth rates, and a conditional cost estimate which includes the full set of controls X^C, but does not rebalance the sample using propensity score weights. For each set of estimates, the cost is calculated as the difference in cumulative GDP growth between two counterfactual scenarios: one where a country defaults in year 1, and one where it does not (see Appendix Section 4.B.1 for further detail).[7]

Without controlling for endogeneity, sovereign default appears very costly. The unconditional sovereign default cost is 3.3% of GDP in year 1, 4.9% in year 2, and 5.6% in year 10. After a sovereign default, there seems to be no economic recovery. But if we, instead, account for observable economic, political and financial conditions before default, the cost becomes much smaller and less persistent. Controlling for observables in a local projection (Table 4.5.1 middle row, Figure 4.5.1 solid grey line) reduces the cost estimate to 2.7% of GDP in year 1, 4% in year 2 and a statistically insignificant 3% in year 10. Accounting for non-linearities in selection through IPSWRA further reduces the cost, such that it is below 4% of GDP at all horizons and statistically insignificant beyond year 6, with a point estimate of 1.7% in year 10.

Two key findings emerge from this analysis. First, sovereign default is costly. Even after controlling for endogeneity and non-linearities in outcomes and selection, sovereign default reduces output by 2.7% on impact and 3.7% at peak in year 5. Second, controlling for endogeneity matters. The IPSWRA estimate is roughly half the size of the unconditional cost, and the output paths under these two types of estimation are statistically different.[8]

The reason for this lower cost is that our Stage 1 logit and Stage 2 IPSWRA account for non-random selection into defaulters, and the co-dependence between default and future GDP growth. Appendix Figure 4.B.2, Table 4.B.2 and Table 4.B.3

7. As shown in equation (4.3.4), this difference is also equal to the coefficient θ_h^w on the default dummy in a weighted least squared regression, where the weights correspond to the inverse propensity scores. All the tables in the main text show the average treatment effect only. Coefficients on predictors and controls, and R^2 statistics at different horizons are shown in Appendix Tables 4.B.2 and 4.B.3.

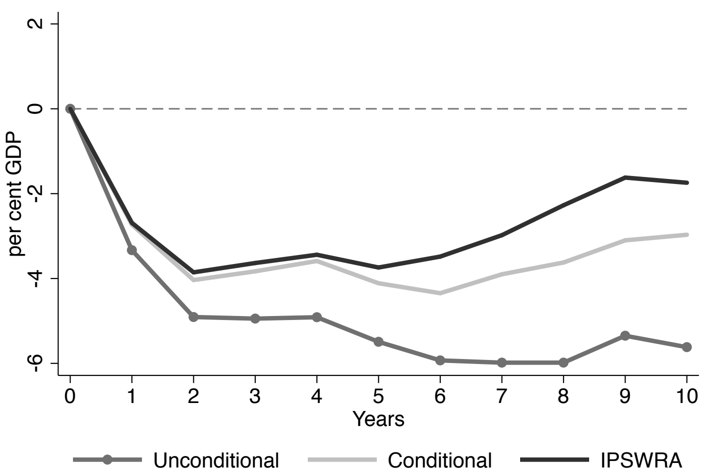
8. Using a "sandwich" estimator, we find that the conditional and unconditional paths are significantly different, at 10% level over the full horizon, with higher significance levels for individual years 1, 7, 8, 9 and 10. We cannot test for the difference between the IPSWRA and unconditional paths because these specifications are not nested, but since the IPSWRA cost is smaller than the conditional LP cost, this test acts as a more conservative lower bound for the difference between the two specifications.

Table 4.5.1. Impact of sovereign default on GDP

Year	1	2	3	4	5	6	7	8	9	10
Unconditional	-3.33***	-4.91***	-4.95***	-4.91***	-5.49***	-5.93***	-5.98***	-5.98***	-5.35***	-5.62**
	(0.64)	(0.97)	(1.06)	(1.24)	(1.34)	(1.53)	(1.70)	(1.88)	(2.14)	(2.48)
Conditional	-2.73***	-4.04***	-3.83***	-3.59***	-4.11***	-4.35***	-3.90***	-3.62**	-3.10	-2.97
	(0.57)	(0.92)	(1.01)	(1.14)	(1.27)	(1.45)	(1.56)	(1.78)	(2.06)	(2.30)
IPSWRA	-2.69***	-3.85***	-3.63***	-3.44***	-3.74***	-3.48**	-2.98	-2.27	-1.62	-1.74
	(0.60)	(1.01)	(1.16)	(1.34)	(1.54)	(1.77)	(1.92)	(2.18)	(2.51)	(2.84)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	92	92	92	92	92	92	92	92	92	92

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth. Clustered standard errors in parentheses. The unconditional local projection controls for country fixed effects only. The conditional local projection controls for country fixed effects and all the varialbes listed in Table 4.A.1. The IPSWRA uses all predictors listed in Table 4.A.1 in the first stage, and all controls in Table 4.A.1 plus country fixed effects in the second stage. *, **, ***: Significant at 10%, 5% and 1% levels respectively

Figure 4.5.1. Impact of sovereign default on GDP



Notes: Cumulative treatment effect, GDP per capita growth. Unconditional specification controls for country fixed effects only. Conditional and IPSWRA specifications control for country fixed effects and the full list of variables in Table 4.A.1.

present the outcomes of the Stage 1 and 2 logit and LP regressions. The Stage 1 logit does well at predicting defaults, with a ROC of 0.84 substantially higher than the naive prediction benchmark ROC of 0.5.[9] Appendix Table 4.B.1 shows that the resulting sample rebalancing helps make our control and treatment groups more similar along a number of observable characteristics, bringing our data closer to that selected at random. Stage 2 controls help forecast GDP growth both at long and short horizons. The Stage 2 LP, in turn, is able to explain much of the endogenous variation in GDP growth, especially at long horizons, with R^2 statistics of 28% in year 1 and 74% in year 10.

Consistent with the existing literature (Manasse, Roubini, and Schimmelpfennig, 2003; Manasse and Roubini, 2009), both debt and macroeconomic variables help in predicting sovereign default, with higher debt service and low growth making default more likely (Appendix Table 4.B.2). We find that debt levels are somewhat less important, perhaps because countries with good growth prospects can and want to borrow more, but are also less likely to default. We also find an important role for global factors such as commodity prices and interest rates, which have so far received relatively little attention in the literature. The Stage 2 LP shows that some of these variables also help forecast future GDP: for example, consistent with existing studies (Borensztein and Panizza, 2008), low GDP growth means both that default is more likely and that future GDP growth will also be low (Appendix Table 4.B.3). Extending the set of controls beyond the usual macro variables shows that debt and global factors also affect future GDP, with high debt or increasing commodity prices predicting high future growth. Political and crisis variables, in turn, matter a lot for long-run growth, with most variables which reduce future GDP – such as banking crises and coups – also increasing the default probability.

The IPSWRA default cost estimate is somewhat lower than most of those in existing literature, particularly at longer time horizons. Our short-run cost estimate is similar to that in Borensztein and Panizza (2008), but above the zero cost found by Levy-Yeyati and Panizza (2011). Our long-run cost estimate is a fraction of those in Furceri and Zdzienicka (2012) and De Paoli, Hoggarth, and Saporta (2009), who find magnitudes of close to 10–15% of GDP. In the bigger picture, the cost of sovereign default appears to be somewhat lower than that of other emerging market crises (Cerra and Saxena, 2008), but above that of a "normal" recession in advanced economies (Jordà, Schularick, and Taylor, 2013). The dynamic path of our cost estimate also stands apart from most studies of sovereign default and other emerging market crises, which find either a very small cost at all horizons, or large costs both in the short and medium to long term (see Table 4.2.1 and Cerra and Saxena, 2008).

9. ROC, or the "receiver operating characteristic" is a relative comparison of true positive and false positive rates, bounded between 0 and 1, with 0.5 corresponding to naive or uninformed prediction and 1 – to a perfectly accurate forecast. Schularick and Taylor (2012) provide a more detailed description of the methodology when applied to rare economic crisis events.

We, on the contrary, find a sizeable short-run cost, but a very low or zero long-run cost.

The differences between ours and other existing estimates of sovereign default cost come about from two sources. First, our comprehensive sample of defaulting countries, complemented by a consistent best-practice default definition (see Section 4.4) ensures that even unconditionally, the cost estimate is sizeable but not overly large. Studies which find a zero or very high cost (De Paoli, Hoggarth, and Saporta, 2009; Levy-Yeyati and Panizza, 2011) generally rely on much more restrictive samples of countries and defaults. Second, the conditioning on observables in the two stages of the IPSWRA attenuates this cost estimate, especially at longer horizons. Figure 4.5.1 shows that the distance between the unconditional and IPSWRA cost estimates increases with the horizon, and the same is true for the R^2 of the stage 2 explanatory regression shown in the Appendix Table 4.B.3. This means that we are able to attribute much of the long-run GDP variation to endogenous factors rather than sovereign default.

To further delineate the contribution of our method – including both the extensive control set and the IPSWRA estimation – Figure 4.B.3 compares our methodology to that used in two other prominent sovereign default cost studies, by Borensztein and Panizza (2008) and Furceri and Zdzienicka (2012). We estimate the default cost by applying the methodology of these two papers to our sample and default definition, thereby abstracting from any differences in sample coverage and data choices. Figure 4.B.3 shows that under these alternative specifications, the cost estimate is close to our unconditional results, and considerably larger than both our conditional and IPSWRA estimates. The cost difference attributable to the method becomes larger at longer horizons. This suggests that the broad set of controls in the LP combined with the IPSWRA sample rebalancing play an important role in explaining the difference between our cost estimate and those in other studies, a finding that also emerges from the more detailed analysis in Sections 4.5.1 and 4.5.2.

Taken together, our baseline results offer both good and bad news for the existing empirical literature on sovereign default costs. On the one hand, at shorter horizons, more naive conditional and unconditional correlations between GDP growth and default – such as those in the historical studies of Reinhart and Rogoff (2011a) and Tomz and Wright (2007) – are likely to have some causal meaning. But when it comes to estimating the full long-run impact of sovereign default, controlling for endogenous selection makes a big difference. The existing empirical consensus is that emerging market crises impose costs that are largely permanent (Aguiar and Gopinath, 2007; Cerra and Saxena, 2008; Furceri and Zdzienicka, 2012; Gornemann, 2014). But our results show that for one specific type of emerging market crisis – sovereign default – controlling for endogeneity in selection and economic outcomes makes most of the long-term cost disappear. Assessing whether this is also the case for other crisis events is a worthy goal for future research.

The size and duration of the default cost fits well with the assumptions made in most current theoretical models. It is higher than the typically assumed 2% temporary endowment penalty (see, for example Aguiar and Gopinath, 2006; Yue, 2010), but lower than the 6% output cost attributed to the endogenous reinforcement mechanism in Mendoza and Yue (2012).[10] The estimate is similar to the 5% default cost assumed by Cole and Kehoe (1996) for the Mexican 1994–95 debt crisis.[11]

We have argued that the difference between our findings and those in the existing literature comes down to a more up-to-date method combined with a comprehensive sample of defaults and control variables. It is, however, still possible that our results are affected by endogenous selection into defaulters and certain choices we make about the data. The next two sections explore whether this could be the case by, first, utilising additional controls and predictors in the IPSWRA and, second, estimating the cost under various alternative data definitions.

4.5.1 Dealing with endogeneity

Section 4.3.2 makes it clear that a causal interpretation of our results relies on a rich conditioning set X and a neutral, or exogenous default definition δ. The choice of variables for the baseline specification, described in Section 4.4, tries to ensure that this is the case. Here we go further by expanding the conditioning set X by including information on sovereign credit ratings and GDP forecasts, which contain soft information on default probabilities and expected economic outcomes which is of direct relevance to the "selection on observables" assumption in Section 4.3.2.

We use country credit ratings provided by the *Institutional Investor Magazine*, which have much broader coverage than those of other agencies. The ratings enter the regression in both levels and first differences. For GDP growth forecasts, we use the dataset provided in the IMF's Historical WEO Forecasts Database. These forecasts were made by the IMF's individual country units, and cover horizons of up to 5 years ahead. The use of both of these datasets substantially reduces our estimation sample, effectively restricting it to defaults that took place in the 1990s. To improve comparability with the baseline specification, we also construct a synthetic credit rating proxy which covers the full sample by predicting ratings out of sample using the methodology of Cantor and Packer (1996) (see Appendix 4.A.1 for further detail).

Table 4.5.2 presents the results. Panel (a) limits the regressions to a smaller sample – effectively, the 1990s – but uses the more accurate raw data on credit

10. The Mendoza and Yue (2012) calibration is based on the Argentinian default of 2001, which is more severe than the representative default in our sample.

11. In models of self-fulfilling sovereign crisis such as Cole and Kehoe (1996) and Cole and Kehoe (2000), the cost of default is not brought about by the government deciding to default *per se*, but by investors that stop rolling over the debt. Still, even in these models there is an element of government discretion, since a prudent government can rule out defaults by keeping its debt levels below the "crisis region".

ratings and forecasts. Panel (b) uses the less accurate synthetic ratings data, but extends the sample to match that in our baseline estimation. Table 4.5.2 panel (a) shows that the credit ratings and GDP forecasts contain little additional information relative to our baseline set of observables. The estimation results in panel (a) top row, which do not account for ratings and growth forecasts, are very similar to those in the bottom row, which include the additional information. Similarly to our baseline estimation in Table 4.5.1, default is costly but the cost is not persistent, even though the estimated size of the cost is different because of the smaller sample size. In line with this intuition, Table 4.5.2 panel (b) shows that adding synthetic ratings to the control and predictor set in the full sample specification makes almost no difference to the size and significance of the estimated regression coefficients.

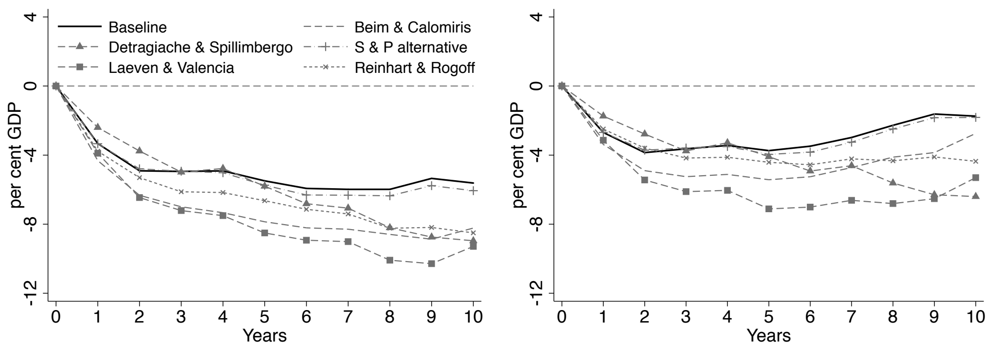
Taken together, the results in this section suggest that if anything, further controlling for default endogeneity strengthens our baseline findings: sovereign default is costly, but the cost is attenuated by conditioning on observables, particularly in the long run. Still, our observable data themselves, including the default definitions

Table 4.5.2. Controlling for sovereign credit ratings and growth expectations

Year	1	2	3	4	5	6	7	8	9	10
(a) Small sample: 17 defaults; 927 observations)										
Baseline	-4.50***	-5.39***	-4.87**	-4.69*	-4.38	-4.01	-1.83	-1.16	-1.09	2.29
	(1.23)	(2.28)	(2.41)	(2.67)	(2.68)	(2.97)	(2.91)	(3.16)	(3.04)	(3.11)
Ratings and Forecasts	-4.67***	-5.52**	-5.11**	-5.17*	-5.00*	-4.80	-2.59	-1.95	-1.69	1.63
	(1.25)	(2.41)	(2.57)	(2.86)	(2.94)	(3.26)	(3.19)	(3.48)	(3.25)	(3.22)
(b) Large sample: 92 defaults; 2546 observations										
Baseline	-2.70***	-3.78***	-3.53***	-3.23***	-3.53**	-3.30*	-2.77	-2.05	-1.42	-1.57
	(0.59)	(1.02)	(1.17)	(1.34)	(1.55)	(1.77)	(1.93)	(2.20)	(2.52)	(2.83)
Synthetic Ratings	-2.66***	-3.72***	-3.44***	-3.12***	-3.35**	-3.10*	-2.55	-1.78	-1.08	-1.15
	(0.60)	(1.00)	(1.14)	(1.32)	(1.51)	(1.74)	(1.90)	(2.18)	(2.51)	(2.80)

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth. IPSWRA estimates using country fixed effects. Clustered standard errors in parentheses. Panel (a) is based on a smaller sample for which data on ratings and forecasts are available. The baseline specification in panel (a) includes the full set of controls and predictors from Appendix Table 4.A.1 and the IPSWRA specification in Table 4.5.1, apart from the second lag of the banking crisis dummy. The specification additionally includes *Institutional Investor Magazine* ratings and GDP forecasts for years 1 to 5 from the Historical WEO Forecasts Database in the control and predictor set. Panel (b) is based on a larger sample consistent with the main results in Table 4.5.1. The baseline specification in panel (b) includes the full list of control and predictors in Appendix Table 4.A.1. The specification additionally includes synthetic ratings constructed in accordance with Cantor and Packer (1996) in the control and predictor set. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Figure 4.5.2. Cost estimates for different definitions of sovereign default



(a) Unconditional **(b)** IPSWRA

Notes: The baseline definition uses *Standard & Poor's* data, Beim & Calomiris group default spells less than 5 years apart together, Detragiache & Spillimbergo definition is based on arrears to total debt, Laeven & Valencia focus on sovereign crises, Reinhart & Rogoff takes the data on defaults and distressed restructurings from Reinhart and Rogoff (2011b), and S & P alternative drops defaults which occur when the country is still in default on another type of debt.

and variable choices, are potentially subject to further selection biases which we examine next.

4.5.2 Other considerations

Our baseline estimate relies on the *Standard & Poor's* default definition, and does not consider the effects of different types of default. This means that first, our results could be driven by the definition we use, and second, that the estimate may not be representative. The cost may be driven by a small subset of costly defaults – for example, those that have high magnitude or those that happen when the country is already experiencing substandard economic performance. We briefly examine each of these concerns to check whether our estimate provides a sufficiently representative and accurate picture of the default cost, with further details provided in Appendix 4.B.4.

Default definition. Figure 4.5.2 shows default cost estimates under the five alternative definitions listed in Section 4.4 Table 4.4.1. It also includes an alternative *Standard & Poor's* based definition, which excludes defaults that happen while the country is still negotiating terms on a previous default on a different type of debt. Figure 4.5.2a shows the unconditional estimates with country fixed effects only, and Figure 4.5.2b presents our preferred IPSWRA specification. Appendix Section 4.B.4.1 and Table 4.B.5 provide a more detailed discussion, as well as point estimates and confidence intervals for each regression specification.

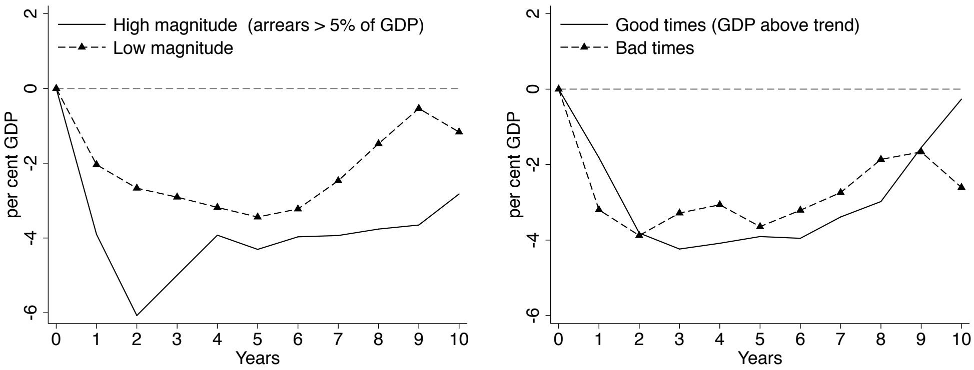
Our two key results continue to hold under these alternative default definitions. First, sovereign default is costly: the impact of default on GDP is negative, sizeable and significant at short to medium term horizons, for all six default definitions, both unconditionally and under IPSWRA. Second, conditioning on observables reduces the cost, especially at long horizons. Compared to the unconditional estimates, the IPSWRA cost is 1–2 percentage points smaller at short to medium horizons, and 4–6 percentage points smaller at long horizons, across the different definitions. The Laeven and Valencia (2012) and Detragiache and Spilimbergo (2001) default definitions which focus on more severe default events, and are hence likely to be more endogenous, result in higher costs, which peak at 6–7% of GDP and persist until year 10 of the regression horizon. The costs under the baseline definition, the S & P alternative, and those of Beim and Calomiris (2000) and Reinhart and Rogoff (2011b) are broadly similar. The higher costs under the arrears-based definition of Detragiache and Spilimbergo (2001) suggest that the size of the default may play a role in determining the cost, which we examine next.

Magnitude. We measure magnitude as the total in-default sovereign debt obligations to private creditors during the first year of default in proportion to nominal GDP. The debt in default data come from the Bank of Canada CRAG database (Beers and Nadeau, 2015). Figure 4.5.3a contrasts the IPSWRA estimates of the cost of high- and low-magnitude defauts. Further discussion and point estimates are provided in Appendix 4.B.4.2 and Table 4.B.7. As intuition would suggest, high-magnitude defaults are more costly, particularly in the short term. While the cost of low-magnitude defaults is around 2–3 % of GDP, close to that of our baseline estimates, that of high-magnitude defaults peaks at 6% of GDP in year 2.

The higher cost of large defaults is most likely driven by a less creditor-friendly negotiation process, which in turn results in higher economic uncertainty and more severe punishment from the creditors. Our findings are thus in line with those of Trebesch and Zabel (2017), who find that "hard" defaults accompanied by more coerciveness towards creditors tend to result in higher output costs.[12] Asonuma and Trebesch (2016) also show that pre-emptive debt restructurings, which differ from outright defaults and are negotiated in a creditor-friendly manner, impose very little cost on the economy. Going back to our findings, even low-mangitude defaults are likely to involve some coerciveness towards creditors, which explains why the costs for these types of events remain sizeable, and shows that our baseline results are not driven by a subsample of high-cost, high-magnitude defaults. However, the cost could still be driven by a small subsample of countries with bad economic fundamentals, both before and after default.

12. Whereas Trebesch and Zabel (2017) provide a detailed measure of the ex-post negotiation outcome during the entirety of the default process, we only capture the debt defaulted in the first year of default, because including any information on post-default outcomes would violate the "selection on observables" assumption.

Figure 4.5.3. Impact of default magnitude and economic situation on the cost of default



(a) Default magnitude

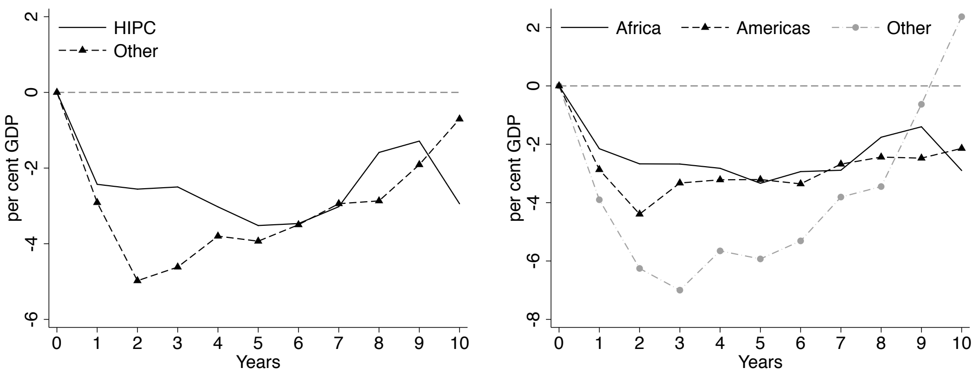
(b) Defaulting in good and bad times

Notes: IPSWRA estimates. Left panel shows the cost of high-magnitude compared to low-magnitude defaults. Default is classified as high-magnitude if debt arrears to private creditors exceed 5% of GDP in the year of default. Right panel shows the cost of defaulting during times of good, or bad economic performance. Good economic performance means that GDP growth is on average above trend during the three years preceding default.

Defaulting in good and bad times. We check whether the cost of default depends on the country's economic fundamentals. To this end, we compare the cost of default for countries growing above and below their HP-filtered trend – what we label "good" and "bad" times. Figure 4.5.3b compares the IPSWRA cost estimates for bad- and good-time defaults (solid and dashed lines respectively), which are classified based on growth in the three years preceding default. We find that defaulting during good times is still costly, which further allays the potential concerns about endogeneity discussed earlier. We also find that defaulting during good times is no more costly than defaulting during bad times, contrary to what is assumed in a number of theoretical models of sovereign default (see the discussion in Section 4.2). Appendix Section 4.B.4.3 and Table 4.5.3b provide further details.

Default cost among different groups of countries. Figure 4.5.3b shows that the stage of the economic cycle seems to have little bearing on the default cost. But some countries tend to suffer from persistently bad economic outcomes, and the cost of default for these economies may be much higher. Figure 4.5.4a compares the cost of default in heavily indebted poor countries (HIPCs), as classified by the World Bank, to that in more developed economies. The costs across these two country groups are similar and, if anything, the shorter term costs for HIPCs are slightly lower, perhaps reflecting the fact that these countries already have poor economic prospects, and defaulting makes a relatively smaller additional difference. Figure 4.5.4b estimates the default costs across different continents and again finds that these are similar, with defaults in Asian and European countries, included in the "Other" group, having a somewhat higher cost.

Figure 4.5.4. Default costs for different country groups



(a) Heavily Indebted Poor Countries

(b) Estimates by continent

Notes: IPSWRA estimates. The HIPC sample split is based on the World Bank classification of heavily indebted poor countries. For continents, Americas includes both North and South America, and other includes Europe, Asia and Oceania.

Alternative regression specifications. In Appendix Section 4.B.4.5, we show that our results are robust to different specifications of the method, such as using a different IPW truncation threshold, different weighting assumptions, a larger control set in the local projection Stage 2, or excluding countries still negotiating a past default from the control group. We also relax the consistent sample assumption and include the extra default predictors from Stage 1 in the Stage 2 estimation.

Our analysis shows that the significance and size of the default cost remains relatively stable and robust across a wide variety of definitions, additional controls and treatments. This stability of the baseline cost estimate raises an altogether different question: are there any factors which we have not examined so far, that systematically amplify default costs? We consider this in the next section by analysing how the costs of default vary with concurrence of other crises.

4.6 Amplification of the cost

Existing research shows that sovereign defaults frequently coincide with banking and currency crises (see, for example Reinhart and Rogoff, 2011b; Morais and Wright, 2008), and the inherently political nature of sovereign decisions means that they also often coincide with political crises such as coups and wars (see Appendix 4.A.3). Such crisis events can – at least in theory – serve to amplify the economic fallout from a sovereign default. To again use the 2015 Greek crisis as an example, Alexis Tispras' government was facing not just the danger of default, but three other significant risks. First, the Greek banking system was highly vulnerable and heavily reliant on the central bank (and ECB) for support. Second, a default would raise the prospects of a severe currency crisis, with Greece likely being forced out of the euro altogether. And third, the prolonged economic slump was accompanied by a tense political climate, frequent street protests and a general disillusionment with the mainstream political parties. In such a situation, we assess whether the cost of default is amplified by a concurrence of a banking, currency or political crisis, to offer more precise guidance to both policymakers and theoretical models of sovereign default.

4.6.1 Defaults and systemic banking crises

The recent Eurozone sovereign crisis has reminded us of the dangerous links between the health of the sovereign and the banking sector (see, for example Reinhart and Rogoff, 2011b; Gennaioli, Martin, and Rossi, 2014; Bocola, 2016; Jordà, Schularick, and Taylor, 2016; Gennaioli, Martin, and Rossi, 2018). Banking crises often require bailouts and activate automatic fiscal stabilisers, increasing the pressure on government finances. Sovereign risk, in turn, spills over to the banking sector through direct write-offs, higher funding costs and liquidity shortages, and a loss of an effective lender of last resort.

In light of this, we pose the following question: is sovereign default more costly when it leads to a systemic banking crisis? Despite its relevance, this issue has received little attention in existing literature. De Paoli, Hoggarth, and Saporta (2009) provide some evidence suggesting that a combination of default and a currency or banking crisis is associated with higher GDP cost. But their study is an outlier in terms of its very small sample size (only three "standalone" sovereign defaults are considered), default and outcome variable definition, and does very little conditioning on observables, which makes it difficult to draw general conclusions. In this section, we use the IPSWRA methodology and our rich conditioning set to estimate the cost of sovereign defaults which are followed by systemic banking crises.

As with our baseline specification, the first task is to define what constitutes a joint default and sovereign crisis event. To do this, we identify sovereign defaults using our baseline *Standard & Poor's* definition, and use the list of systemic banking crisis compiled by Laeven and Valencia (2012). In classifying the joint events, we want to exclude those occasions where a sovereign default was caused by problems originating in the banking sector. We therefore only include those events where a sovereign default occurred 1 or 2 years *before* the banking crisis, or where the two occurred in the same year but problems in the sovereign sector preceded, or were not related to, banking distress.[13] This leaves us with 11 joint banking-sovereign de-

13. To do this, we undertook a narrative examination of each joint default and banking crisis event, and excluded all those where banking system problems seemed to be the main cause for the sovereign default, or preceded sovereign distress. Because our sample mainly consists of emerging markets with relatively undeveloped financial systems, the line of causation from banking crisis to default is relatively rare. We exclude two joint events – Ecuador 1982 and Indonesia 1983 – where the banking sector problems predated default, and keep 8 other joint events.

fault events. The list includes both those defaults where sovereign distress directly triggered the banking panic, such as those of Russia 1998 and Argentina 2001, and those where both crises were triggered by a third, unrelated factor such as the collapse in the price of uranium in the early 1980s, which triggered the banking and sovereign default of Niger in 1983. Appendix Table 4.A.1 contains a short description of each joint sovereign and banking crisis.

The second task lies in defining the appropriate set of control variables. Because sovereign and banking crises often have common causes (Reinhart and Rogoff, 2011b), and we focus on events where sovereign distress was the primary driver of the joint crisis, our list of controls and predictors from the baseline estimation in Section 4.5 is generally sufficient. Still, banking and sovereign distress generally have somewhat different causes, and high banking sector vulnerability may make certain countries more likely to experience a joint sovereign and banking crisis event. Existing literature suggests that build-ups in credit, and funding imbalances of the financial system are the two key predictors of banking crises (Schularick and Taylor, 2012; Jordà, Richter, Schularick, and Taylor, 2017). We therefore add two additional variables to our control and predictor set in order to capture these channels: the growth in the credit-to-GDP ratio, and the ratio of loans to deposits, both sourced from World Bank *Financial Development and Structure Database* (Beck, Demirgüç-Kunt, and Levine, 2010). Appendix Figure 4.B.1 shows that our first stage prediction does a good job at forecasting both standalone and joint sovereign-banking crises, generating high predicted probabilities for both of these events. This makes the IPSWRA specification well-suited to controlling for endogenous selection into such crises.

Table 4.6.1 and Figure 4.6.1 present the IPSWRA estimates of the cost for those defaults that are followed by a systemic banking crises (Table 4.6.1 bottom row, Figure 4.6.1b), compared to those which are not (Table 4.6.1 top row, Figure 4.6.1a). One key result stands out: sovereign defaults are significantly more costly when followed by a systemic banking crisis. While the cost of standalone sovereign defaults is similar to that in our baseline specification, the onset of a banking crisis roughly doubles the short- to medium-term fallout from default. Under the twin crisis scenario, the cost reaches 4.4% of GDP in the first year, and peaks at 9.5% of GDP in year 3. The default costs under the standalone default and joint crisis scenarios are statistically different in years 2 and 3, and the full paths of the response are significantly different at 1% level.[14] The GDP cost of sovereign-banking crises is higher than that of most other crisis events examined in the literature, including financial recessions in Jordà, Schularick, and Taylor (2013), and systemic banking crises or

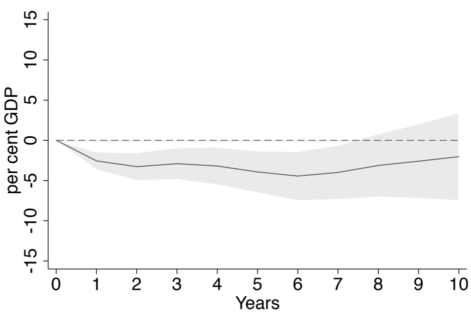
14. As in Section 4.5, we use a "sandwich" estimator to test for joint difference in the treatment effect estimates for standalone defaults vs sovereign-banking crisis, in all of the years 1–10.

Table 4.6.1. Cost of sovereign default and systemic banking crises

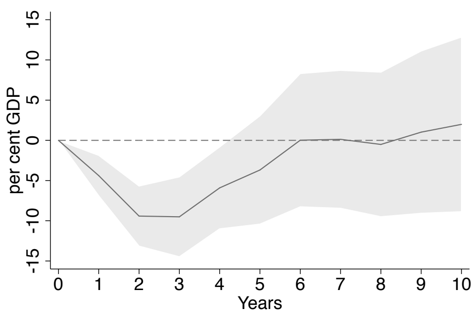
Year	1	2	3	4	5	6	7	8	9	10
Default + no Crisis	-2.56***	-3.27***	-2.90***	-3.19**	-3.93***	-4.44***	-3.99*	-3.12	-2.59	-2.03
(no. defaults = 72)	(0.69)	(1.07)	(1.21)	(1.44)	(1.60)	(1.87)	(2.07)	(2.41)	(2.83)	(3.34)
Default + Crisis	-4.39***	-9.42***	-9.51***	-5.91*	-3.69	0.02	0.13	-0.51	1.02	1.99
(no. defaults = 11)	(1.54)	(2.28)	(3.03)	(3.11)	(4.11)	(5.06)	(5.23)	(5.48)	(6.16)	(6.62)
Observations	2245	2245	2245	2245	2245	2245	2245	2245	2245	2245
p-value:										
crisis = no crisis	0.25	0.01	0.05	0.42	0.96	0.41	0.46	0.65	0.59	0.57

Notes: Average treatment effect on cumulative real GDP per capita growth: defaults that are followed, or not followed by a systemic banking crisis. All defaults that are followed by a banking crisis in the next two years are classified as Default + Crisis events. Banking crises occurring prior to default, even within the same year, are excluded. Treatments are based on a simple sample split of our baseline default definition. All figures are IPSWRA estimates controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Figure 4.6.1. Cost of sovereign default and systemic banking crises



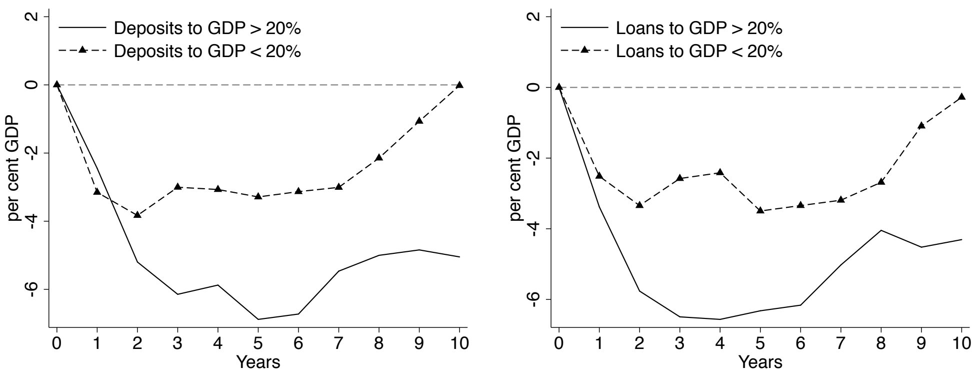
(a) Default and no systemic banking crisis



(b) Default and systemic banking crisis

Notes: Cumulative treatment effect, GDP per capita growth. Shaded bands indicate 90% confidence intervals. Sample split based on defaults followed by a systemic banking crisis within two years. IPSWRA estimates using country fixed effects and the full list of variables in Table 4.A.1.

Figure 4.6.2. Financial Development and default



(a) Deposits to GDP **(b)** Loans to GDP

Notes: IPSWRA estimates of the cost of default. The sample split is based on the loan-to-GDP and deposit-to-GDP ratios in the year before default.

civil wars in Cerra and Saxena (2008).[15] The much higher cost of joint sovereign-banking crisis episodes persists under a variety of alternative joint event definitions. Appendix Section 4.B.5 Figure 4.B.2 and Table 4.B.1 show that sovereign-banking crises remain costly regardless of whether the default happens after a banking crisis, in the same year, or the banking crisis precedes the default event.

The link between banking distress and sovereign default costs stretches beyond the analysis of joint crisis events discussed above. Figure 4.6.2 suggests that higher levels of financial development tend to amplify the costs of sovereign default, regardless of whether these defaults are followed by a banking crisis or not. Countries with higher deposits or loans relative to GDP incur default costs that are roughly double those of financially undeveloped countries, particularly over the medium to long term.[16] This indicates that the cost of sovereign default is amplified by financial sector distress, and that impairment of the banking system has an important role in generating the costs in the first place.

This analysis makes clear that the policymakers would be right to worry about the potential impact of default on the domestic banking system. But should they also be concerned about the potential currency crisis, and the economic costs of any political fallout from default?

15. We also estimate the cost of a third scenario – a banking crisis that is not preceded by a default – and find that these are substantially lower than those of the joint default-crisis events. Results are available from authors upon request.

16. The thresholds are chosen to correspond to the mean loan and deposit to GDP ratios in the sample of defaulters. Results are robust to using different thresholds; additional results are available from authors upon request.

4.6.2 Currency and political crises

Both currency and political crises represent significant risks during the time of sovereign default. Since emerging-market sovereigns tend to denominate their external debt in foreign currency, a sharp devaluation may make the debt unsustainable. Equivalently, a sovereign default may reduce the confidence in the currency, triggering a self-fulfilling currency panic. Indeed, the strong link between sovereign default and currency crises has been well-documented in the existing literature (see for example, Reinhart and Rogoff, 2011b; Kaminsky, 2006; De Paoli, Hoggarth, and Saporta, 2009). The political environment around the time of default has not been studied as systematically, but it should not come as a surprise that defaults often coincide with times of political turmoil, as documented in the Appendix Table 4.A.1.

In light of these facts, we examine how the cost of default changes if the default coincides, is preceded or followed by a political or currency crisis, using a one-year joint event window. We follow the Laeven and Valencia (2012) definition of a currency crisis, and define a political crisis as a high-intensity war, a coup or a political transition. The results are reported in Appendix Tables 4.B.2 and 4.B.3. As with banking crises, the no-crisis results are similar to our baseline estimates. The costs of joint crisis events are slightly higher than those of "standalone" defaults, especially on impact in year 1. But the differences are small (1 – 2% of GDP), and the costs lie far below those of joint sovereign and systemic banking crises (Figure 4.6.1b). We therefore conclude that unlike banking crises, currency and political crises do not strongly amplify the cost of sovereign default.

Overall, the state-contingent effects of sovereign default can be summarised in one simple sentence: to paraphrase Bill Clinton's famous slogan, "it's the banks, stupid". It turns out that the costs of sovereign default, even though substantial, can be reasonably contained as long as the banking system remains operational – even if the country is experiencing a currency or a political crisis. Should the banks fail, however, the defaulting country ought to brace itself for a severe economic downturn.

Why is it that sovereign-banking crises, and defaults in financially developed countries are so costly? There are two main channels through which a sovereign default can transmit through the financial system, related to banking sector solvency and liquidity. The solvency channel hurts banks through write-offs or lower valuations of sovereign debt holdings on their balance sheets (as in Gennaioli, Martin, and Rossi, 2014). While this channel may be important in some cases, it is unlikely to be the main driving force behind our results because we only consider defaults on external debt, little of which tends to be held by domestic banks in emerging market economies. Liquidity-based explanations are, therefore, likely to be important. A sovereign default may result in higher funding costs, or outright exclusion of domestic banks from international funding markets (Cruces and Trebesch, 2013, provide evidence that such an exclusion does take place for sovereigns). In such an event, banks are likely to struggle to replace lost foreign funding at short notice, especially when the banking system is vulnerable and domestic deposits – scarce. This liquidity drain might then force much of the banking system into insolvency, or significantly impair its functioning, and create negative knock-on effects on the real economy.

Even though the above reasoning helps explain why sovereign and banking default events may occur together and amplify each other, it only offers limited insights into which precise transmission channels translate the sovereign and banking distress into a cost for real economic activity. We aim to shed more light on this in the next section.

4.7 Decompostion of the cost

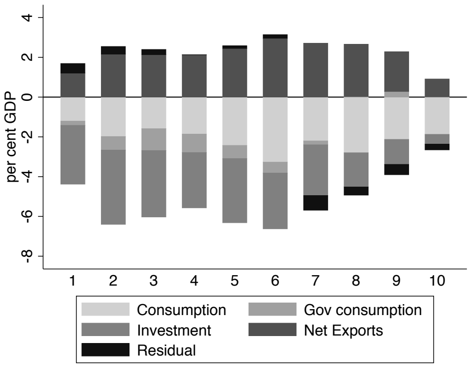
Theoretical models have proposed a number of channels though which sovereign default can harm the economy (see Section 4.2). These transmission channels generally have an asymmetric impact on different economic sectors: for example, banking disruption may disproportionately affect investment, while trade sanctions should reduce exports and imports. In this section, we decompose our aggregate default cost estimate into individual components of GDP – consumption, investment, government spending, exports and imports – to gain insights into how the sovereign default cost comes about in the first place.

Figure 4.7.1a breaks down the aggregate GDP cost in Table 4.5.1 into individual GDP components. The bars show the contribution of each component to the total GDP treatment effect, which can be either positive (bar above zero line) or negative (bar below zero line). For example, the cumulative treatment effect in Year 1 is around –2.7% of GDP. Of that, investment contributes –3% of GDP, consumption –1.2% of GDP and government spending –0.2% of GDP (all shown by negative bars). In contrast net exports exert a positive contribution of +2% of GDP.[17] Table 4.7.1 lists the point estimates and standard errors underlying Figure 4.7.1. In order to interpret these figures one has to take into account the share of each component in GDP (Appendix Table 4.B.1). In other words, all else equal those components with the largest shares are also expected to make the largest contributions to GDP.

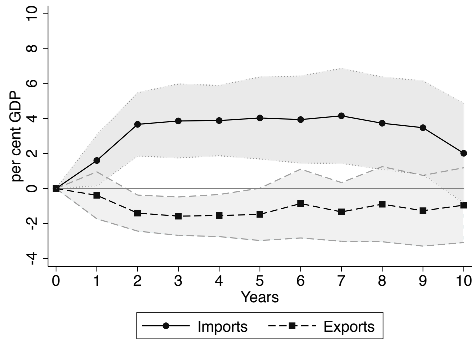
Sovereign default brings about a rapid and sizeable reallocation of resources within the economy. After a default most GDP components fall, but they do not fall equally. Investment experiences the most pronounced decline: it falls by 3% in the year of default and continues to drag down GDP by 3.8% in the year after. Given the small GDP share of this component – about 18% for defaulters (Appendix Table 4.B.1) – this represents a drop of more than one-fifth in relative terms. The fall in private consumption is modest, especially given its GDP share of 70%. In

17. The sum of all components will not exactly equal the total GDP treatment effect due to a small residual (dark bar), in the case of Year 1 roughly 0.5% of GDP

Figure 4.7.1. The impact of default on components of GDP



(a) Decomposition of the treatment effect



(b) Exports and Imports

Notes: Cumulative contribution of individual components to GDP after a sovereign default. Calculated as the absolute change in a GDP component between t and t + h, scaled by the GDP level at t. Here t is the year of the default, and h is the horizon, plotted on the x-axis. Shaded bands indicate 90% confidence intervals. IPSWRA estimates using country fixed effects and the full list of variables in Table 4.A.1.

contrast to private demand, the drop in government consumption is much smaller and more gradual, perhaps because reneging on sovereign debt obligations frees up the resources for other expenditures.

Another sharp adjustment takes part on the external side of the economy. Defaulters tend to sharply reduce external dependence and increase net exports by around 3% of GDP in the medium term. But they cannot achieve this by simply increasing exports – as documented in previous studies by Rose (2005) and Borensztein and Panizza (2010), sovereign default tends to harm exporting firms. We also find a 1–2% GDP drop in exports. The required increase in net exports can then only be achieved via a rapid and sharp reduction in imports, which peaks around 4% of GDP in years 5–7, and persists into years 8 and 9 even as the total GDP cost becomes insignificant. The drop in imports represents a decline of around one-sixth in relative terms. We now turn to examine the underlying mechanisms behind the sharp declines in gross trade and investment observed after the default.

4.7.1 Understanding the decline in gross trade

To gain further insight into what drives the sharp post-default drop in trade, we assess whether the cost of default varies according to the exchange rate regime. Pegged countries tend to run up larger current account deficits prior to defaulting, and have less scope for an orderly external adjustment because their exchange rate is fixed. If external imbalances and the associated adjustment frictions are important in generating the default cost, we would expect this cost to be higher under pegged exchange rates. To do this, we split the sample of defaulters into countries with

Table 4.7.1. The impact of default on components of GDP

Year	1	2	3	4	5	6	7	8	9	10
Investment	-2.96***	-3.73***	-3.34***	-2.78***	-3.22***	-2.80***	-2.56**	-1.72	-1.26	-0.49
	(0.78)	(1.01)	(1.04)	(1.08)	(1.17)	(1.19)	(1.19)	(1.10)	(1.03)	(1.22)
Consumption	-1.22	-1.98**	-1.59	-1.86	-2.43**	-3.27***	-2.20	-2.80*	-2.13	-1.87
	(0.87)	(0.94)	(1.27)	(1.18)	(1.19)	(1.34)	(1.71)	(1.66)	(1.98)	(2.08)
Government	-0.20	-0.68**	-1.10***	-0.93***	-0.67*	-0.55	-0.20	0.03	0.28	0.01
Consumption	(0.17)	(0.33)	(0.34)	(0.36)	(0.34)	(0.34)	(0.44)	(0.49)	(0.48)	(0.46)
Exports	-0.39	-1.41**	-1.58***	-1.55**	-1.48	-0.86	-1.34	-0.90	-1.27	-0.95
	(0.82)	(0.63)	(0.67)	(0.73)	(0.91)	(1.20)	(1.02)	(1.31)	(1.23)	(1.31)
Imports	1.60*	3.67***	3.87***	3.89***	4.04***	3.94***	4.16***	3.73**	3.48**	2.02
	(0.89)	(1.10)	(1.29)	(1.22)	(1.43)	(1.52)	(1.65)	(1.61)	(1.63)	(1.74)
Real GDP	-2.69***	-3.85***	-3.63***	-3.44***	-3.74***	-3.48**	-2.98	-2.27	-1.62	-1.74
(total)	(0.60)	(1.01)	(1.16)	(1.34)	(1.54)	(1.77)	(1.92)	(2.18)	(2.51)	(2.84)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	92	92	92	92	92	92	92	92	92	92

Notes: Average treatment effect of sovereign default on individual components of GDP. The outcome variable is the absolute change in a GDP component between t and t + h, scaled by the GDP level at t. Here t is the year before default, and h is the horizon. IPSWRA specification, controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. Effects do not sum exactly to the treatment effect on GDP; small residual. *, **, ***: Significant at 10%, 5% and 1% levels respectively

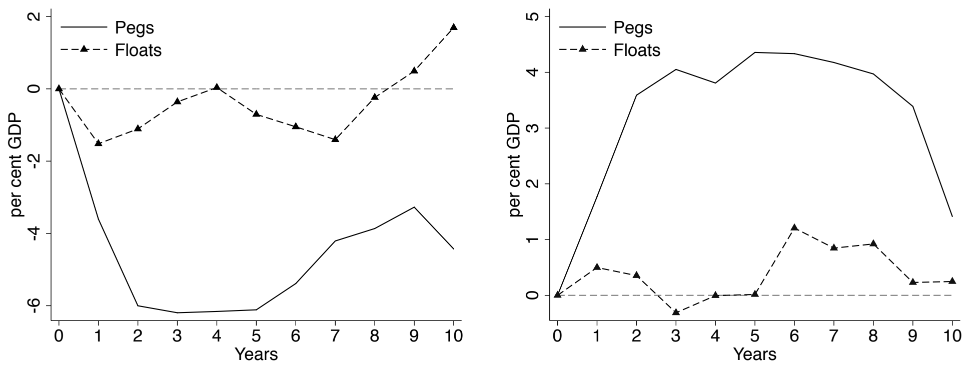
pegged and floating exchange rate regimes in the year before default, using the classification in Ilzetzki, Reinhart, and Rogoff (2017).[18]

Figure 4.7.2a presents the IPSWRA default cost estimates under pegged and floating exchange rate regimes. The point estimates and standard errors are reported in Appendix Table 4.B.2. The cost of default varies substantially according to the exchange rate regime, in sharp contrast to the other sample splits analysed in Section 4.5.2 of this paper. Almost all of the sovereign default cost is incurred under pegged exchange rates. For countries with floating exchange rates the GDP cost is close to zero, whereas pegged countries suffer GDP losses of close to 6% in the medium run, and 4% in year 10.

Figure 4.7.2b helps us understand why the costs under pegged exchange rates are so high. Pegged countries tend to run large current account deficits of near 8% of GDP before default, and have to undertake a rapid rebalancing in its aftermath, increasing their net exports by 4% of GDP for nearly a decade. It is difficult to undertake such a rapid adjustment via increases in exports, especially when the nominal

18. We classify all countries with no separate legal tender, hard pegs, crawling pegs and narrow exchange rate corridors as pegs, and both managed floating and floating exchange rates as floats.

Figure 4.7.2. Default costs under pegged and floating exchange rate regimes



(a) GDP

(b) Net Exports

Notes: IPSWRA estimates of the GDP cost of default (panel (a)) and changes in net exports relative to GDP in year 0 (panel (b)). Pegged exchange rates include countries with no separte legal tender, hard pegs, crawling pegs and narrow exchange rate corridors. Standard errors and point estimates for pegs are reported in the Appendix Table 4.B.2.

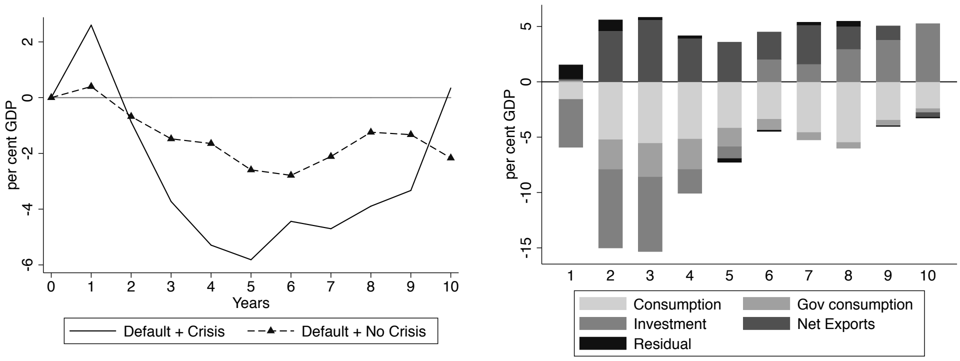
exchange rate is fixed. In fact, exports actually decline after the default, and all of the external adjustment in Figure 4.7.2b takes the form of lower imports (see Appendix Figure 4.B.1). This adjustment seems to impose substantial costs on the economy in the form of lower GDP.

The evidence for rapid external adjustment and high costs for pegged exchange rates is consistent with the main theoretical mechanism underlying most sovereign default models – that of financial autarky. But this transmission mechanism generates much higher costs than the standard default model à la Eaton and Gersovitz (1981), where autarky increases consumption volatility but has no effect on the level of GDP. This suggests that economic frictions, which make it difficult to reallocate resources between sectors and firms, play an important role in generating the output cost. The patterns in the data are consistent with the model of Mendoza and Yue (2012), where firms face constraints on working capital, and struggle to finance imports of intermediary inputs after a default, which reduces production efficiency and hence output. Na, Schmitt-Grohé, Uribe, and Yue (2018) provide an alternative mechanism which can explain our empirical findings. In their model, defaulting countries need to undertake a large relative price adjustment in order to stabilise the economy. A pegged exchange rate limits the scope for such changes in relative prices, meaning that defaults under a peg should be accompanied by large and persistent increases in involuntary unemployment.

4.7.2 Understanding the decline in investment

Investment projects are typically long-term and reliant on bank financing, especially in emerging markets where non-bank financial intermediation is relatively undeveloped. The sharp investment decline in Figure 4.7.1a may, therefore, be directly con-

Figure 4.7.3. Credit and GDP components in the aftermath of sovereign-banking crises



(a) Credit to GDP **(b)** GDP components after a sovereign-banking crisis

Notes: IPSWRA estimates using country fixed effects and the full list of variables in Table 4.A.1. Left-hand panel shows the treatment effect on the credit to GDP ratio, for two groups of defaults: those that are, and are not followed by a systemic banking crisis. Right-hand panel shows the cumulative contribution of individual components to GDP after a sovereign default which is followed by a systemic banking crisis within two years. Calculated as the absolute change in a GDP component between t and t + h, scaled by the GDP level at t. Here t is the year of the default, and h is the horizon, plotted on the x-axis.

nected to the health of the banking system, and to the high output cost of sovereign-banking crises documented in Section 4.6. To further test for presence of these connections in the data, we investigate whether sovereign defaults are accompanied by declines in bank credit, and whether the declines in credit and investment are larger for those defaults which are followed by systemic banking crisis events.

Figure 4.7.3a presents the IPSWRA estimates of the impact of sovereign default on credit to GDP. These use the methodology from Section 4.3, but replace the outcome variable y with the credit to GDP ratio. We calculate the impact separately for standalone sovereign defaults (dashed line), and those followed by a systemic banking crisis (solid line). Consistent with the hypothesised importance of the banking channel, credit declines after both standalone and twin sovereign-banking crisis defaults. In the absence of a banking crisis, the fall is already sizeable and amounts to 2% of GDP, around one-tenth of the average credit to GDP ratio of 20% in the defaulter sample. But after dual sovereign-banking crises, credit to GDP declines by 6 percentage points, or roughly one-third of the sample average.

Figure 4.7.3b shows that this large credit decline during sovereign-banking crises is accompanied by sharp falls in investment. The figure provides a component decomposition of the total sovereign-banking crisis GDP cost in Table 4.6.1 in the same way that Figure 4.7.1a does for our baseline estimates. Appendix Table 4.B.3 shows the underlying point estimates and standard errors. After a sovereign-banking crisis, investment declines by 4.3% of GDP in year 1 and 7.1% of GDP in year 2. Given the average pre-crisis investment to GDP ratio of 18%, this represents a fall of more than one-third in relative terms. Figure 4.7.3b also shows that the trade channel continues to play an important role: net exports increase by roughly

6% of GDP by year 3, with this adjustment, again, driven by reductions in imports. In relative terms, imports fall by around one-third during the first three years after a sovereign-banking crisis. This suggests that international autarky and banking sector distress interact and amplify each other in important ways.

Our findings are consistent with those in Acharya et al. (2018), who show that high sovereign risk in the euro area crisis led to worse borrowing conditions and lower investment, employment and sales for affected firms. The results in this section suggest that this credit distress channel may also be an important driver of sovereign default costs. Outright default and exclusion from international financial markets, however, brings a new twist to the story: when the banking system breaks down in the presence of financial autarky, firms may lose access to trade and investment credit both at home and abroad. Investment collapses, and together with it so do the imports of investment goods, trade credit and domestic production. The combination of these factors helps explain the marked output contraction observed following the sovereign-banking crisis events.

The evidence on the importance of different transmission channels has direct implications for theoretical models of sovereign default. Our results confirm that financial autarky plays an important role in generating sovereign default costs. At the same time, the impact of autarky seems to go far beyond a simple increase in consumption volatility: output declines, investment contracts and gross trade collapses. This asymmetric impact of default is also quite different from a standard endowment penalty assumed in the literature, which would impact all components of GDP proportionately. Even though a number of mechanisms could underly this, our analysis suggests that the impact of default on the banking sector, and its interaction with autarky costs, is particularly important. Incorporating the banking sector and the interplay between sovereign and banking distress into sovereign default models offers a natural way to microfound the output cost of default. This could provide an endogenous mechanism that amplifies the cost of autarky and facilitate both stronger creditor punishment and higher levels of sovereign debt.

4.8 Conclusion

This paper provides a new best-practice sovereign default cost estimate by applying novel econometric methods to a comprehensive panel dataset of sovereign defaults and their determinants. We find that sovereign default is costly: its impact on GDP is negative, statistically significant and highly persistent – but not permanent. Accounting for endogenous selection attenuates the cost, but its magnitude remains higher than that of a normal recession, and comparable to that of other crisis events, as well as the costs assumed in a variety of theoretical models. This helps to explain why defaults – even though they do happen occasionally – are still considered extreme events rather than regular occurrences, at least for most countries.

What is it that makes default costly? The impact of default on trade, and the high costs incurred under pegged exchange rates, point to the importance of autarky costs in the transmission mechanism. However, the high cost of defaults followed by systemic banking crises, and the sharp drops in investment and credit observed for all types of default, suggest that banking sector distress is equally important. Theoretical models of sovereign default should, therefore, benefit from focussing on sovereign-banking spillovers and their interaction with autarky costs. When it comes to making policy decisions, it may be tempting to focus entirely on the negotiations and the potential retaliation from the country's creditors. But when a country's sovereign is going bust, it pays to keep a close eye on domestic banks.

Appendix 4.A Data Appendix

4.A.1 Data sources and summary statistics

The first part of the Appendix describes the construction of the dataset. Table 4.A.1 lists the sources used to construct each variable in our baseline regression, divided into outcomes y, treatments δ, and controls X. Control variables are split into three groups: those that enter both the Stage 1 logit and Stage 2 local projection; those that enter the Stage 1 logit only; and those which are only available for a subsample of our data, and are used for extra robustness checks in Section 4.5.2.

Table 4.A.3 provides a rationale for the inclusion of each control and predictor variable, as well as their expected impact on the two outcomes of interest – sovereign default and GDP. Our conditioning set contains information on the country's debt position, macroeconomic and political environment, different types of ongoing crises, short-term liquidity needs and global financial conditions. Table 4.A.4 summarises the sample coverage by listing the countries and years for which we have data on the full set of outcome, treatment control and predictor variables.

The set of controls is substantially broader than that used in the existing literature. Most studies of the cost of default focus on macroeconomic controls, and on variables that affect growth or development more generally such as past GDP or investment share; with several international or openness related variables also typically included (Borensztein and Panizza, 2008; De Paoli, Hoggarth, and Saporta, 2009; Levy-Yeyati and Panizza, 2011; Furceri and Zdzienicka, 2012). The literature on predicting sovereign debt crises makes the full use of debt data and often links to other crisis events, but does not generally connect the results to future GDP outcomes (Reinhart and Rogoff, 2011b; Manasse, Roubini, and Schimmelpfennig, 2003; Manasse and Roubini, 2009). Measures of political situation and distress are not typically conditioned on. By utilising the full predictive power of macro-financial variables, and including a broad set of information on the macroeconomic, political and financial situation of the country as controls in the LP, we attain a substantially broader conditioning set than that used in the existing literature. Section 4.B.3 shows that the inclusion of these additional controls substantially reduces the long-run cost default cost estimate (effectively, this can be gauged by comparing our conditional LP estimate with an LP mimicing the empirical specification in Furceri and Zdzienicka, 2012).

We generally use the raw data with few modifications, with the exception of three variables. For our political crisis measure, we combine information on wars, coups d'etat and political transitions from the Polity datasets, and define a political crisis event as any one of these events taking place, with the war intensity threshold set to 4 out of 20 to isolate the more severe events.

For the endogeneity robustness checks in Section 4.5.1, we construct a synthetic sovereign credit rating variable. To do this, we follow Cantor and Packer (1996) and predict ratings out of sample using real GDP growth (2 lags), GDP level, inflation, external debt to GDP and the number of past defaults, with the T-bill rate and continent dummies added to proxy for global financing conditions and levels of economic development, respectively.

Table 4.A.1. Data sources and variables used in main regressions

Variable	Source	Description
Dependent variables		
GDP growth	Penn World Tables (PWT)	Percentage change in real GDP per capita
GDP components	PWT	Growth of investment, consumption, government spending and net exports relative to GDP
Treatments		
External default	Beers and Chambers (2006); *Standard & Poor's* reports	Failure to repay or a distressed restructuring of external debt. Dummy variable equal to 1 in the first year in default and 0 otherwise. *Standard & Poor's* data are complemented with defaults in Reinhart and Rogoff (2011b) and Reinhart and Trebesch (2016) before 1975 and after 2006.
B & C defaults	Beim and Calomiris (2000), extended using baseline definition	Equals 1 for the first year in default and 0 otherwise.
D & S defaults	Detragiache and Spilimbergo (2001), extended using arrears data	Equals 1 for year of default and 0 otherwise.
L & V defaults	Laeven and Valencia (2012)	Equals 1 for the first year in default and 0 otherwise.
R & R defaults	Reinhart and Rogoff (2011b), Reinhart and Trebesch (2016)	Equals 1 for the first year in default and 0 otherwise.
Default magnitude	Beers and Nadeau (2015)	Private creditor debt in default relative to GDP

Table 4.A.1. Data sources and variables used in main regressions (continued)

Variable	Source	Description
Controls & predictors: used in both Stage 1 (logit) and Stage 2 (local projection)		
Public external debt	World Bank GDF (2012) & IDS (2014)	Ratio to GDP
Total external debt	as above	Ratio to GDP
Real GDP level	PWT	GDP per capita
GDP cyclical component	PWT	Relative deviation of real per-capita GDP from HP-filtered trend
Inflation rate	PWT	Change in GDP deflator
Terms of trade	PWT	Change in terms of trade
Current account	PWT	Ratio to GDP
Openness	PWT	(Imports+Exports)/GDP
Government size	PWT	Government consumption/GDP
Commodity Index, CCI	Thomson Reuters	Equally weighted index
Banking crisis	Laeven and Valencia (2012)	Equals 1 if a systemic banking crisis starts that year, 0 otherwise
Currency crisis	Laeven and Valencia (2012)	Equals 1 if a currency crisis starts that year, 0 otherwise
War	Marshall (2014) MEPV database	Sum of war intensities across all types of conflict ≥ 4
Coup	Marshall and Marshall (2014)	Dummy for coup or attempted coup
Political transition	Marshall, Gurr, and Jaggers (2014) Polity IV	Equals 1 in the first year of transition, 0 otherwise
Political crisis	MEPV and Polity IV	1 if dummy for war, coup or political transition equals to 1, and 0 otherwise
Governance quality	Polity IV	Revised combined Polity score
Predictors used in Stage 1 (logit) only:		
Short-term external debt	World Bank GDF (2012) & IDS (2014)	Ratio to GDP
Equity return over bills	Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019)	16 advanced economies, GDP weighted
Equity dividend yield	Jordà, Knoll, et al. (2019)	16 advanced economies, GDP weighted

Table 4.A.1. Data sources and variables used in main regressions (continued)

Variable	Source	Description
Predictors used in Stage 1 (logit) only, continued:		
Interest payments on external debt	World Bank GDF (2012) & IDS (2014)	Ratio to GDP
US T-Bill rate	Federal Reserve	1-year constant maturity rate
Number of past defaults	Standard & Poor's	Defaults since 1950
Continent	geonames.org	Continent dummies
Additional controls & predictors used for robustness purposes:		
Sovereign Credit Ratings	Institutional Investor Magazine	100-point scale, from 0 (highest credit risk) to 100 (lowest credit risk)
Synthetic sovereign ratings	Predicted sovereign rating following Cantor and Packer (1996)	Prediction uses data on GDP, inflation, debt, T-bill rate and continent dummies. In-sample estimates using IIM ratings are used to construct synthetic ratings outside of the IIM sample.
Growth Forecasts	Historical WEO Forecasts Database	GDP growth forecasts for the next 5 years
Additional variables used for banking-sovereign crisis prediction and sample splits:		
Credit to GDP	Beck, Demirgüç-Kunt, and Levine (2010)	Nominal credit divided by nominal GDP from PWT. We add these as controls and predictors to the sovereign-banking crisis IP-SWRA, in levels and changes
Deposits to GDP	Beck, Demirgüç-Kunt, and Levine (2010)	Total bank deposits divided by nominal GDP
Loans to deposits ratio	Beck, Demirgüç-Kunt, and Levine (2010)	Aggregate credit relative to total bank deposits
Other variables used to determine sample splits:		
Pegged exchange rate	Ilzetzki, Reinhart, and Rogoff (2017)	Peg dummy equals 1 if the country has no independent currency, a hard peg, a crawling peg, or a narrow exchange rate corridor. This corresponds to to regimes 1–11 on the scale of Ilzetzki, Reinhart, and Rogoff (2017). Both managed floats and free floats are classified as floating regimes.

Table 4.A.2. Defaults in the baseline sample

Argentina: 1982, 1989, 2001	Burkina Faso: 1983
Bulgaria: 1990	Bolivia: 1980, 1986, 1989
Brazil: 1983	Central African Republic: 1981, 1983
Chile: 1983	Cote dIvoire: 1983, 2000
Cameroon: 1985	Congo, Dem. Rep.: 1976
Congo, Republic of: 1983	Costa Rica: 1981
Dominican Republic: 1982	Algeria: 1991
Ecuador: 1982, 1999	Gabon: 1986, 1999
Ghana: 1987	Guinea: 1986, 1991
Gambia: 1986	Guinea-Bissau: 1983
Guatemala: 1986, 1989	Guyana: 1979, 1982
Honduras: 1981	Haiti: 1982
Indonesia: 1998, 2002	Jamaica: 1978, 1981, 1987
Jordan: 1989	Kenya: 1994, 2000
Liberia: 1981	Morocco: 1983, 1986
Moldova: 1998, 2002	Madagascar: 1981
Mexico: 1982	Myanmar: 1997
Mauritania: 1992	Malawi: 1982, 1988
Niger: 1983	Nigeria: 1982, 2001
Nicaragua: 1979	Pakistan: 1998
Panama: 1983, 1987	Peru: 1976, 1978, 1980, 1984
Philippines: 1983	Paraguay: 1986
Romania: 1981, 1986	Russia: 1998
Sudan: 1979	Senegal: 1981, 1990, 1992
Sierra Leone: 1983, 1986	Togo: 1979, 1982, 1988, 1991
Turkey: 1978, 1982	Tanzania: 1984
Uganda: 1980	Ukraine: 1998
Uruguay: 1983, 1987, 1990	Venezuela: 1983, 1990
Zambia: 1983	Zimbabwe: 2000

Table 4.A.3. Controls and predictor variables: rationale for inclusion and expected effects

Variable	Rationale for inclusion	Expected effect on	
		Default	GDP
External debt (public and total)	Indebted countries have more to gain from default, and high debt can either slow down future growth or be taken on in anticipation of higher growth.	+	±
GDP growth and cyclical component	Poor growth may make it relatively costly to meet debt repayments, and signal poor economic prospects	–	–
GDP level	Poor countries may be more likely to default. Effect on growth is ambiguous: positive if there is convergence, and negative if the country is stuck in a development trap	–	±
Inflation rate	High inflation may make it more difficult to repay foreign currency debts, and harm future growth prospects	+	–
Terms of trade	Deteriorating terms of trade (increase in the index) may make it more difficult to repay foreign currency debts and harm economic growth by imposing adjustment costs and putting a strain on corporate balance sheets	+	±
Current account	External dependence (a current account deficit) may make countries more vulnerable to foreign funding shocks, but could also deter default because of higher potential for creditor punishment, and could be harmful to future growth	±	+
Openness	More open economies may be more exposed to economic and external shocks, but also have more to lose from default making default less likely, and should grow faster over the long run	±	+
Commodity prices	Higher commodity prices are benefit commodity exporters and hurt importers, both in terms of ability to repay debts and economic growth	±	±
Banking, currency and political crises	These should all reduce future growth prospects, and make the country more likely to default, either because it has less resources for repaying debts or because the government changes	+	–
Government size	A large government sector may make default more costly (if financing is cut off), and less likely. A large inefficient government sector may also be a drag on future growth.	–	–

Table 4.A.3. Controls and predictor variables: rationale for inclusion and expected effects (continued)

Variable	Rationale for inclusion	Expected effect on	
		Default	
Governance quality	Erratic policymaking may make default more likely and hinder future growth	+	−
Short-term external debt size and interest payments	High short-term refinancing needs, which may trigger a default	+	
Equity excess return and dividend yield	Measures of ex post and ex ante risk premiums in advanced economies, which act as a proxy of investor risk appetite, and desire to fund risky emerging country credits.	+	
US T-Bill rate	A proxy for global funding conditions, and hence the desire to finance foreign governments. High rates mean unfavourable funding conditions and higher default probability.	+	
Number of past defaults	A proxy for unobservable country characteristics which made countries default in the past, and also more likely to default in the future	+	
Continent	A coarser substitute for country fixed effects in the logit	±	
Sovereign Credit Ratings	Higher rating is a proxy for low default probability.	−	
Growth Forecasts	Higher future growth, and hence better ability to repay the debt	−	

Table 4.A.4. List of observations included in the baseline regression

Albania	1989–2011	Algeria	1975–2011	Angola	1987–2011
Argentina	1970–2011	Armenia	1991–2011	Azerbaijan	1992–2011
Bangladesh	1972–2011	Belarus	1991–2011	Benin	1970–2011
Bhutan	1987–2011	Bolivia	1970–2011	Botswana	1975–2011
Brazil	1970–2011	Bulgaria	1983–2011	Burkina Faso	1975–2011
Burundi	1975–2011	Cambodia	1984–2011	Cameroon	1970–2011
Cape Verde	1979–2011	Central African Republic	1970–2011	Chad	1970–2011
Chile	1970–2011	China	1980–2011	Colombia	1970–2011
Comoros	1975–2011	Congo, Dem. Rep.	1970–2011	Congo, Republic of	1970–2011
Costa Rica	1970–2011	Cote d'Ivoire	1970–2011	Djibouti	1982–2011
Dominican Republic	1970–2011	Ecuador	1970–2011	Egypt	1970–2011
El Salvador	1975–2011	Eritrea	1998–2011	Ethiopia	1989–2011
Fiji	1974–2011	Gabon	1970–2011	Gambia, The	1975–2011
Georgia	1991–2011	Ghana	1970–2011	Guatemala	1970–2011
Guinea	1970–2011	Guinea-Bissau	1974–2011	Guyana	1971–2011
Haiti	1973–2011	Honduras	1975–2011	Hungary	1980–2011
India	1970–2011	Indonesia	1970–2011	Iran	1978–2011
Jamaica	1971–2011	Jordan	1970–2011	Kazakhstan	1991–2011
Kenya	1970–2011	Kyrgyzstan	1991–2011	Laos	1972–2011
Latvia	1991–2011	Lebanon	1975–2011	Lesotho	1976–2011
Liberia	1970–2011	Lithuania	1991–2011	Macedonia	1992–2011
Madagascar	1972–2011	Malawi	1975–2011	Malaysia	1970–2011
Mali	1970–2011	Mauritania	1970–2011	Mauritius	1975–2006
Mexico	1970–2011	Moldova	1991–2011	Mongolia	1990–2011
Morocco	1970–2011	Mozambique	1982–2011	Myanmar	1975–2011
Nepal	1975–2011	Nicaragua	1970–2011	Niger	1970–2011
Nigeria	1970–2011	Pakistan	1970–2011	Panama	1970–2011
Papua New Guinea	1975–2011	Paraguay	1970–2011	Peru	1970–2011
Philippines	1970–2011	Romania	1978–2011	Russia	1992–2011
Rwanda	1974–2011	Senegal	1970–2011	Serbia	1991–2011
Sierra Leone	1972–2011	South Africa	1992–2011	Sri Lanka	1970–2011
Sudan	1971–2011	Swaziland	1975–2011	Syria	1970–2011
Tajikistan	1991–2011	Tanzania	1970–2011	Thailand	1970–2011
Togo	1970–2011	Tunisia	1970–2011	Turkey	1970–2011
Turkmenistan	1992–2011	Uganda	1970–2011	Ukraine	1991–2011
Uruguay	1970–2011	Uzbekistan	1991–2011	Venezuela	1970–2011
Vietnam	1985–2011	Yemen	1990–2011	Zambia	1970–2011
Zimbabwe	1970–2011				

List of countries and years that are included in either treatment or control group in the baseline regressions. Sample coverage of the other specifications are available upon request. A few observations for the in-between years are missing.

4.A.2 Data on alternative default definitions

Table 4.A.1. Alternative default definitions

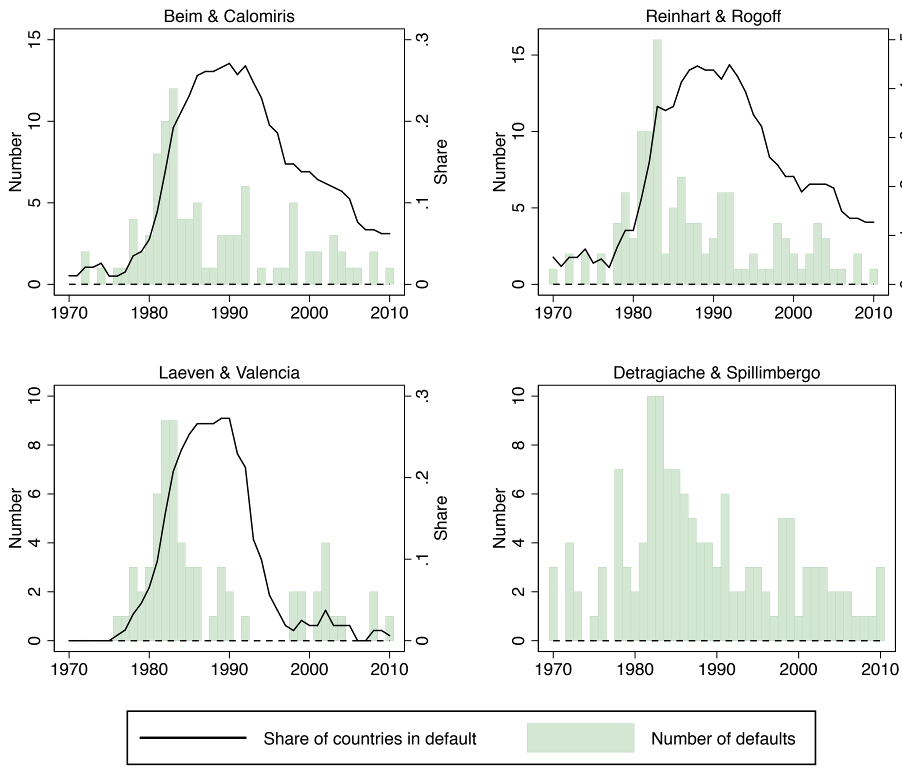
Source	Original definiton	Extension
Reinhart and Rogoff (2011b)	Failure to make a payment; distressed restructurings	Use original data, extended by Reinhart and Trebesch (2016)
Beim and Calomiris (2000)	Failure to make a payment over > 6 months, no political defaults, group default spells less than 5 years apart together, exclude voluntary refinancings	After 1992, and for countries not covered in Beim and Calomiris (2000), we add the defaults from our baseline definition, and group them together if the in-default periods are less than 5 years apart
Laeven and Valencia (2008)	Failure to make a payment; distressed restructurings; narrative selection of crisis episodes	Use original definition
Detragiache and Spilimbergo (2001)	Non-payment arrears > 5% of total debt; distressed restructurings	After 1998, and for countries not covered by Detragiache and Spilimbergo (2001), we classify all instances where arrears to private creditors are > 5% of total external debt, and arrears were below 5% for the preceding two years as a default. Arrears data come from Beers and Nadeau (2015).

Table 4.A.1 details the construction of each alternative default definition variable in our sample. The Reinhart and Rogoff (2011b) and Laeven and Valencia (2012) definitions are up-to-date and cover a broad selection of countries, hence we simply use the original definition provided by these authors. The Beim and Calomiris (2000) and Detragiache and Spilimbergo (2001) original definitions cover a lower number of countries, and years up to 1992 and 1998 respectively. For each of these datasets, we construct our own proxy of their definition for countries with no defaults in the original data, and for years beyond 1992 and 1998 respectively. For Beim and Calomiris (2000), we do this by merging together all *Standard & Poor's* default spells which start less than 5 years after the end of another negotiation. For Detragiache and Spilimbergo (2001), we add all instances where private debt arrears exceed 5% of total external public debt, using the arrears data in the Bank of Canada CRAG database, 2018 update (see Beers and Nadeau, 2015, for further detail). To exclude instances of countries repeatedly dipping below and above the 5% threshold, we only include those defaults where the level of arrears was below 5% for two consecutive years prior to the default date. Our version of these default definitions is in some ways cruder than the originals: for the Beim and Calomiris (2000) dates, we do not have data on repayment delays and case-by-case political narratives, so our extension may include political defaults or those with repayment

delays of less than 6 months. For Detragiache and Spilimbergo (2001), our extension will exclude any distressed restructurings that did not generate large arrears. In other ways, our data may be somewhat more accurate than those in the original studies: the default estimates of Beers and Chambers (2006), Reinhart and Rogoff (2011b) and Reinhart and Trebesch (2016) that we use to extend the series are relatively more up-to-date and accurate, as are the arrears data in Beers and Nadeau (2015).

For the first three definitions in Table 4.A.1, we construct two variables: the in-default dummy, which equals 1 whenever the country defaults or is negotiating a past default, and the default dummy, which only equals 1 in the first year of the default, and is equivalent to the δ we use in our empirical estimation. The Detragiache and Spilimbergo (2001) definition does not provide data on default duration, hence we only construct the default dummy δ. Figure 4.A.1 shows the time trends in the number of defaults and share of countries in default between 1970 and 2010, following the same format as Figure 4.4.1 for the baseline definition, with teal bars showing the number of new defaults, and the solid line – the share of all countries that have newly defaulted or are still negotiating a past default. All four definitions show a wave of defaults in 1980s, continued negotiation and high in-default shares in the early 1990s, and a drop-off in default rates afterwards. The 1980s peak is most pronounced in the Laeven and Valencia (2012) definition, and least pronounced for the definition of Detragiache and Spilimbergo (2001). The trend of the in-default share is similar across different definitions, but its level is higher under the definition of Reinhart and Rogoff (2011b). This is largely because the all-country sample for the other definitions includes more countries, and for the relatively recent data since the 1970s we can be fairly sure that if no default is recorded for these countries in the dataset, this means that no default took place. The Reinhart and Rogoff (2011b) definition goes back to the early 1800s, and their historically consistent sample of countries is somewhat smaller than the universe of all independent countries at any point in time.

Figure 4.A.1. Frequency of sovereign defaults since 1970 under alternative definitions



Notes: Beim & Calomiris defintion data are from Beim and Calomiris (2000), extended by grouping our baseline defaults into one for those time periods and countries not covered by Beim and Calomiris (2000). The share of countries in default is relative to all countries in our sample. Reinhart & Rogoff definition data are from Reinhart and Rogoff (2011b) and Reinhart and Trebesch (2016), and the share of countries in default is relative to the sample in these two papers. The Laeven & Valencia definition, and the number of countries in their sample are sourced from Laeven and Valencia (2012). Detragiache & Spillimbergo definition uses the data from Detragiache and Spilimbergo (2001), and extends it across time and countries using the arrears data in Beers and Nadeau (2015).

4.A.3 Timelines of sovereign defaults and other crisis events

Figure 4.A.1 shows the timeline of the default events under our baseline *Standard & Poor's* definition, and each of the four alternative definitions described in Section 4.A.2. There is substantial overlap across countries but also some heterogeneity, especially when it comes to the arrears-based Detragiache and Spilimbergo (2001) definition, which is conceptually quite different from the other three.

Figure 4.A.2 shows the timeline of sovereign defaults compared to systemic banking, currency and political crises. The banking and currency crisis classification follows Laeven and Valencia (2012). Political crises largely use Polity IV data, and correspond to coups, wars or political transitions. The political distress dummy is set to 1 for the full duration of the crisis, whereas the other three variables only equal 1 in the first year of the crisis event This is mainly done for comparability purposes, since we lack data on the duration of banking and currency crises; also this way the variables correspond directly to ones used in our regression analysis. The solid lines indicate the data coverage.

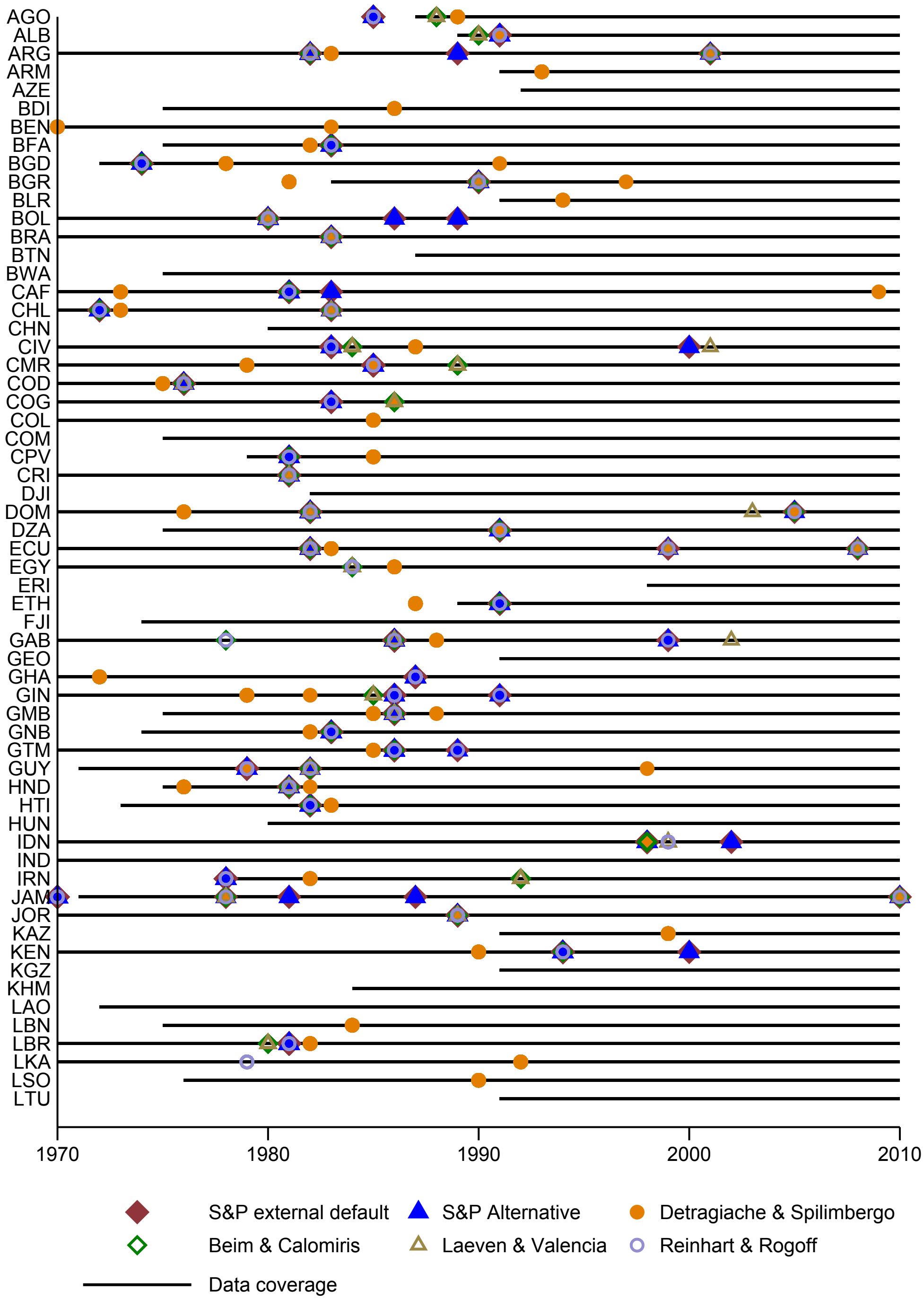
Table 4.A.1 summarises the joint occurrence of the different crisis events throughout our historical sample. We include 92 defaults, 17 of which coincide with systemic banking crises, 35 – with currency crises, 5 – with both banking and currency crises, and between 10 and 20 – with various types of political crises. Taken together, roughly two-thirds of our default observations overlap with other crisis events. This considerable overlap between sovereign defaults and other crisis events motivates our separate analysis of joint crises and standalone defaults in Section 4.6.

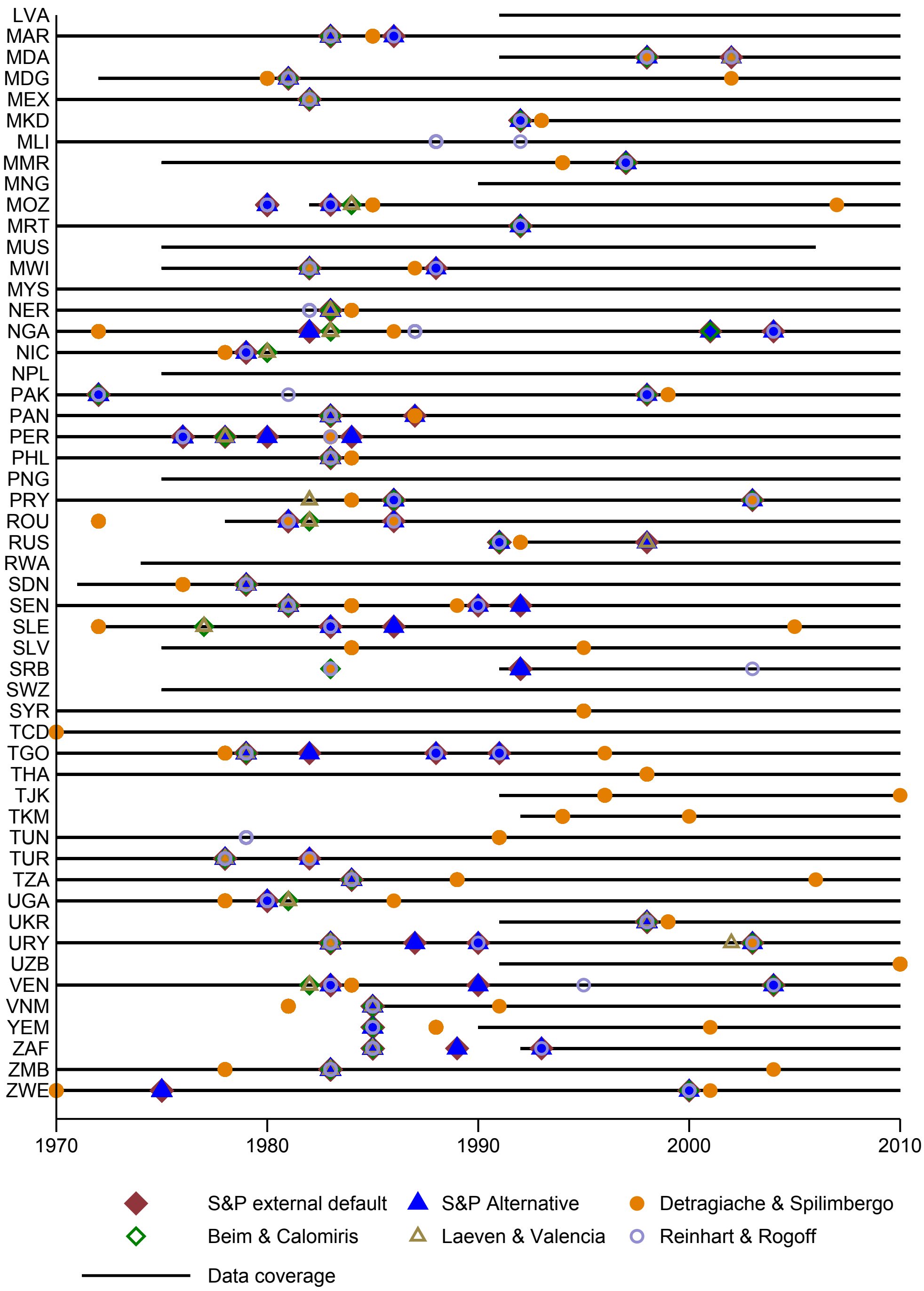
Table 4.A.1. Number of sovereign defaults coinciding with other crisis events

Economic crises:	
Banking crises	17
Currency crises	35
Triple crises (banking + currency + sovereign)	5
Political crises:	
Wars	10
Coups	21
Political transitions	16
All crisis events	59

A joint event is *any* of the above crises occurring concurrently, in the year before or the year after the sovereign default. For more detail on sources for each individual crisis variable, see Table 4.A.1.

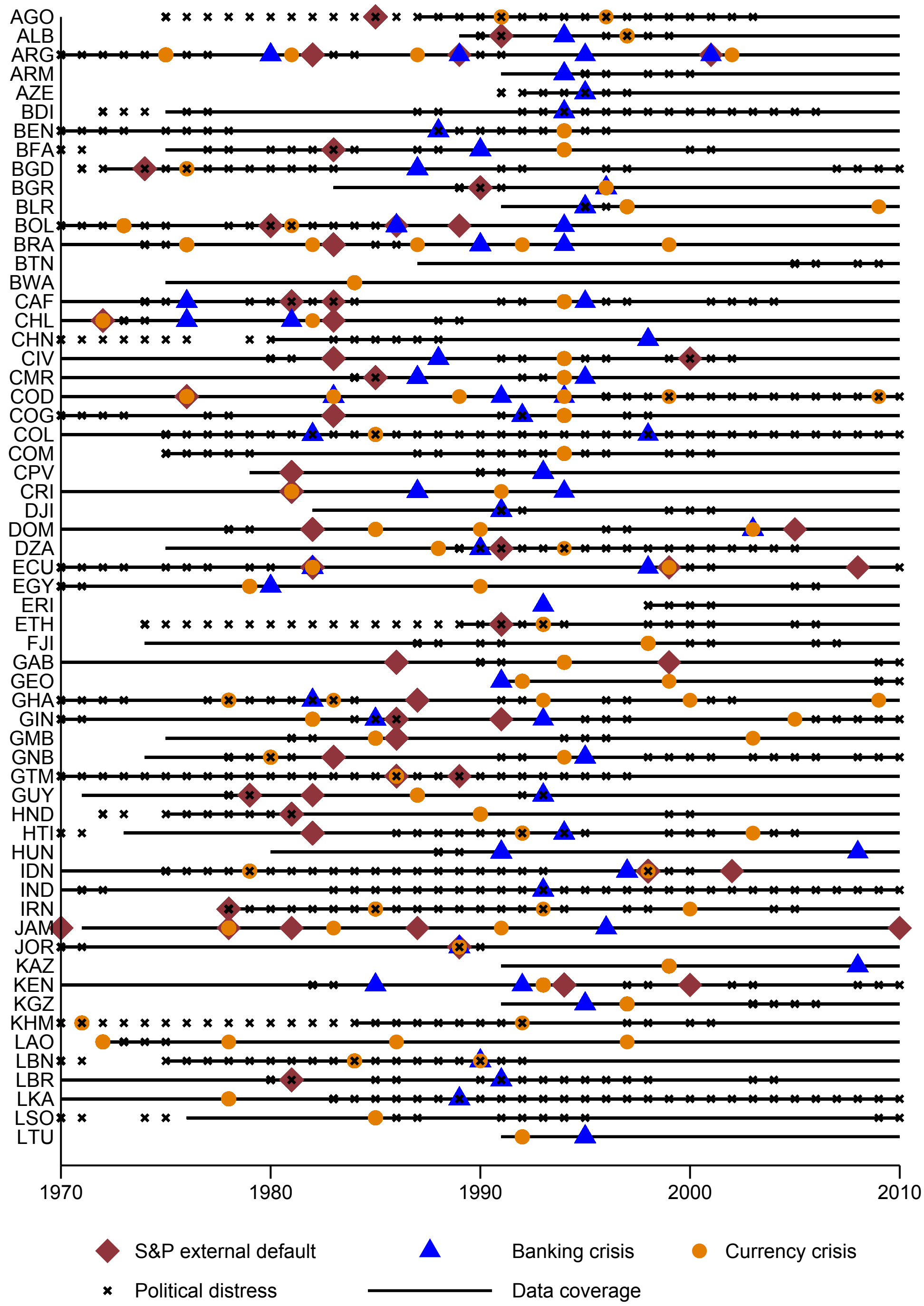
Figure 4.A.1. The timeline of sovereign defaults for individual countries

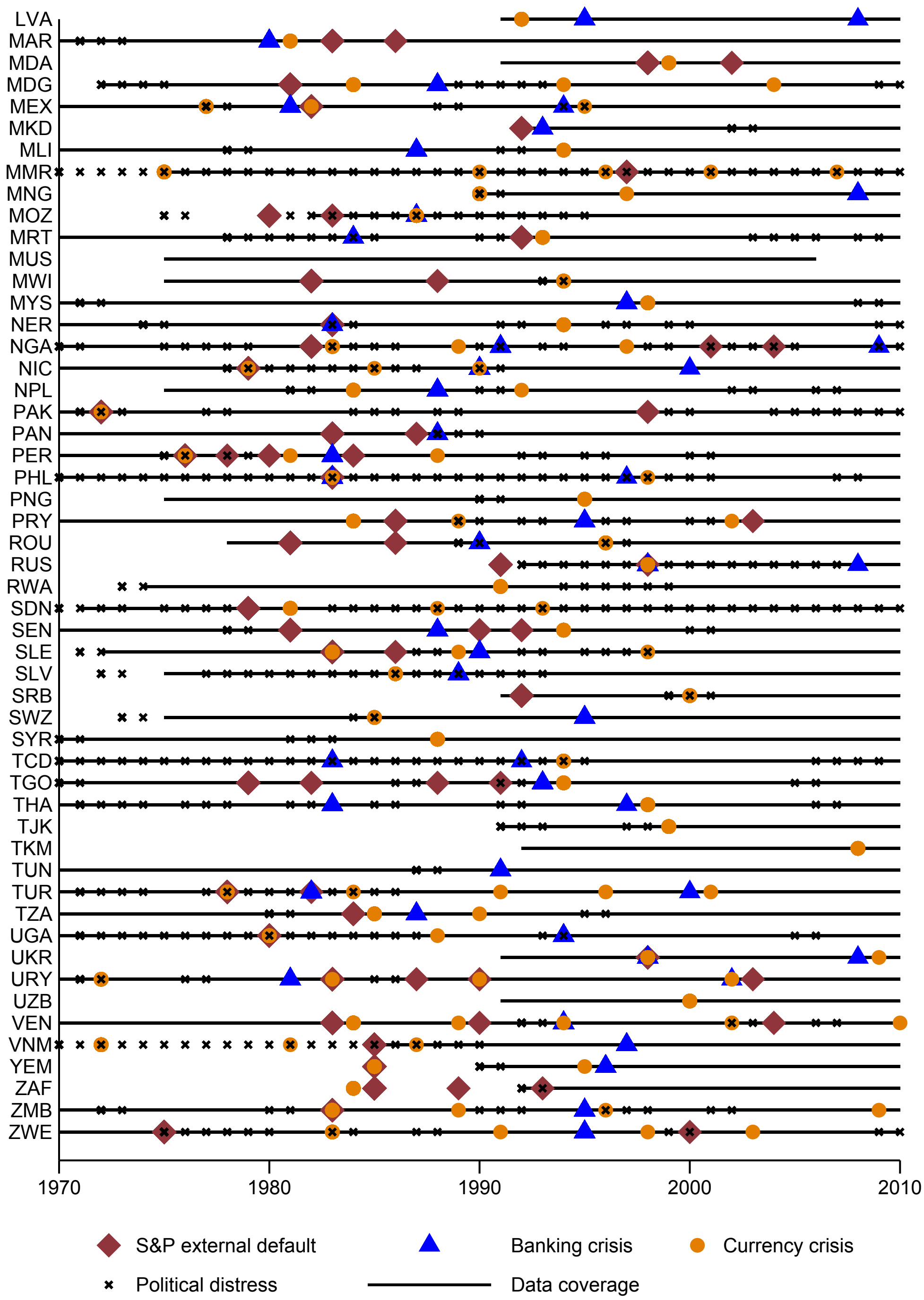




LVA
MAR
MDA
MDG
MEX
MKD
MLI
MMR
MNG
MOZ
MRT
MUS
MWI
MYS
NER
NGA
NIC
NPL
PAK
PAN
PER
PHL
PNG
PRY
ROU
RUS
RWA
SDN
SEN
SLE
SLV
SRB
SWZ
SYR
TCD
TGO
THA
TJK
TKM
TUN
TUR
TZA
UGA
UKR
URY
UZB
VEN
VNM
YEM
ZAF
ZMB
ZWE
1970
1980
1990
2000
2010
S&P external default
S&P Alternative
Detragiache & Spilimbergo
Beim & Calomiris
Laeven & Valencia
Reinhart & Rogoff
Data coverage

Figure 4.A.2. The timeline of other crisis events for individual countries




LVA
MAR
MDA
MDG
MEX
MKD
MLI
MMR
MNG
MOZ
MRT
MUS
MWI
MYS
NER
NGA
NIC
NPL
PAK
PAN
PER
PHL
PNG
PRY
ROU
RUS
RWA
SDN
SEN
SLE
SLV
SRB
SWZ
SYR
TCD
TGO
THA
TJK
TKM
TUN
TUR
TZA
UGA
UKR
URY
UZB
VEN
VNM
YEM
ZAF
ZMB
ZWE
1970
1980
1990
2000
2010
S&P external default
Banking crisis
Currency crisis
Political distress
Data coverage

4.A.4 Joint sovereign-banking crisis events

Table 4.A.1 provides a more detailed description of the events we classify as joint sovereign-banking crises for the purpose of the analysis in Section 4.6.1 Table 4.6.1 and Figure 4.6.1. To focus the analysis on sovereign debt problems and the associated costs, we only include those events where the sovereign default preceded the banking crisis (3 cases), or those where the two events occurred in the same year but either the sovereign default was the main cause of both crises (6 cases), or the two were driven by unrelated events (2 cases). We exclude two events where the sovereign and banking crises happened in the same year – Philippines 1983 and Ecuador 1982 – from the list because in these cases, problems in the banking sector precipitated sovereign default. Our selection of sovereign-to-bank crisis is similar to that obtained in recent work by Balteanu and Erce (2018), who use information in IMF Article IV reports, financial press and country monographs to order the sequence of events within a twin episode.

Table 4.A.1. List of joint sovereign-bankign crisis events, where the sovereign default preceded or was not caused by the banking crisis

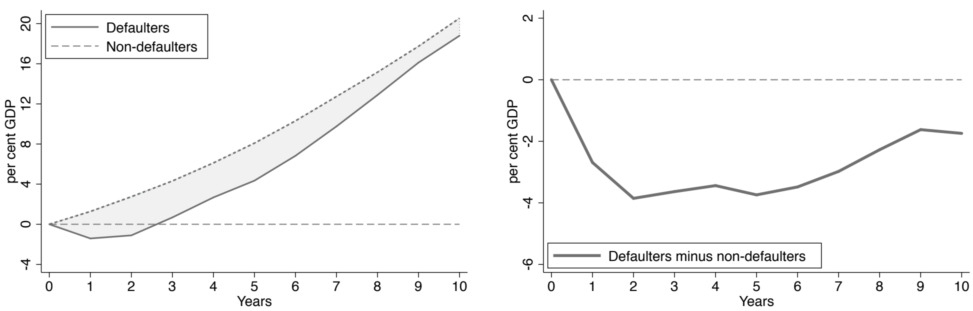
Episode	Narrative of events
Sovereign default and banking crisis in the same year:	
Turkey, 1982	Sovereign finance problems starting in 1970s, default in 1978, coup in 1980, another default in 1982, accompanied by a banking crisis.
Niger, 1983	Boom in uranium prices accompanied by sovereign and private credit boom during 1978–1981; fall in price of uranium triggers sovereign default and a banking crisis.
Bolivia, 1986	1982–1985 economic crisis, sovereign finance problems, hyperinflation due to monetary debt finance. Sovereign default coincided with collapse of state-owned banks.
Jordan, 1989	Accumulated large amount of sovereign and private debt during the oil boom, default on both when oil prices decline.
Argentina, 1989	Government debt sustainability problems throughout the 1980s; currency devaluation, and sovereign default in 1989. Conversion of time deposits into government bonds imposed losses on depostit holders and initiated the banking crisis.
Argentina, 2001	Unsuccessful attempts to eliminate large budget deficit in 1999–2001, political instability in 2001, banking system used to finance deficit needs in early 2001, together with an IMF package. Bank run triggered by the uncertainty about sovereign, deposit freeze in early December. Sovereign default at the end of 2001.
Russia, 1998	Government defaults on domestic and external short-term debt (GKOs), devalues the currency. Sovereign default and devaluation bring down the banking system (banks had large holdings of sovereign bonds, and currency mismatches).
Ukraine, 1998	Structural problems with tax collection and sovereign finances after transition. Banking sector vulnerable as well, with problems in 1997. Sovereign default due to inability to make payments and contagion from Russia, banking crisis soon after.
Banking crisis 1 or 2 years after a sovereign default:	
Cameroon, 1985	Failed coup attempt in 1984 amid climate of political instability, sovereign default in 1985, banking crisis in 1987.
Panama, 1987	Fiscal deficits of 10–15 percent GDP during the 1980s, suspension of external debt payments during 1987-1988, systemic banking crisis in 1988.
Togo, 1991	Political uncertainty, coup and sovereign default in 1991, systemic banking crisis in 1993.

Appendix 4.B Further empirical findings

4.B.1 Interpreting the baseline estimates

To clarify what we mean by the cost of sovereign default, Figure 4.B.1 shows precisely how our estimate is constructed. The left panel displays the expected evolution of cumulative GDP growth for defaulters and non-defaulters – the expected potential outcomes after rebalancing the sample using IPSW and conditioning on the local projection controls in (4.3.2). A representative non-defaulting country is expected to grow close to trend whereas if it defaults, GDP is expected to first fall and then slowly catch up.

Figure 4.B.1. Calculating the average treatment effect of sovereign default



(a) Expected cumulative GDP growth for defaulters and non-defaulters

(b) Average treatment effect of default

The difference between the two expected GDP paths is the average treatment effect – our measure of the default cost. It is the shaded area between the two curves in the left-hand panel, also plotted separately in the right-hand panel. The right-hand-panel figure is the one we present in the tables and graphs in the results section. The idea is similar to comparing the GDP growth performance of defaulters to trend, where the trend is estimated using data for the control group, and both the control and treatment group samples are rebalanced and conditioned on the local-projection controls. Since the cost estimate is computed for a representative country in our broad sample, it captures the “gross” cost of default – that of defaulting compared to doing nothing.

Table 4.B.1. Characteristics of the rebalanced treatment and control groups

	Treatment (defaulters)	Control (non-defaulters)	Difference significant?
GDP growth	-1.10	1.06	Yes(1% level)
External public debt/GDP	47.00	51.12	No
Inflation	22.50	24.63	No
Openness	59.01	62.46	No
Governance quality score (Polity)	-1.99	-1.15	No
Banking crisis probability	0.09	0.08	No
Currency crisis probability	0.12	0.12	No
War intensity (scale 0 – 20)	0.65	1.10	Yes(1% level)
Coup probability	0.11	0.08	No

Notes: All values refer to the year preceding default, and in the case of banking and currency crisis probabilities, to two years before default. Openness is the ratio of gross imports and exports to GDP. Governance quality is scored on a scale from -10 to 10, with a higher score meaning better governance. All ratios are presented as percentage points, all growth rates in percent. The third column tests the equality of the respective means between the treatment and the control group. GDP growth and inflation are winsorized at the 2% level. The sample is rebalanced using the probability weights from the first stage estimation.

4.B.2 IPSWRA estimation: first and second stage

This section separately presents the outcomes of the two stages of IPSWRA. The first stage (equation (4.3.1)) estimates the propensity to default using a logit, including the list of predictors in Table 4.A.1 which are chosen in accordance to the literature on predicting sovereign defaults or debt crises (Manasse, Roubini, and Schimmelpfennig, 2003; Manasse and Roubini, 2009). Table 4.B.2 shows the predictive power of these variables. The decision to default is affected by the country's economic situation – it is negatively correlated with previous year's GDP growth – as well global factors such as commodity prices and interest rates. For the debt variables, liquidity needs and debt service seem to be most important, with the level of debt playing a relatively minor role. When significant, the signs of the coefficients generally follow the economic rationale described in Table 4.A.3, apart from the war index, which, if higher, reduces the probability of default – perhaps reflecting a need to access the debt market to finance expenditures during these times.

Figure 4.B.2 shows the ROC for the logit prediction. The curve compares the true and false positive rates. The fact that the ROC curve is above the 45 degree line (the random prediction) indicates that the logit is informative in predicting defaults. The area under the ROC curve measures the strength of this prediction, and equals 0.84, substantially higher than the naive prediction of 0.5. The value of 0.84 is high considering defaults are rare events, and compares favourably with other estimates in the literature, such as the 0.71 area for predicting systemic banking crises in Schularick and Taylor (2012).

Table 4.B.1 reports the sample characteristics of the treatment and control groups after the sample is rebalanced using the inverse propensity score weights

generated from the logit regression in Table 4.B.2. Compared to the unweighted averages reported in Table 4.3.1, the rebalancing makes the treatment and control groups more similar along a number of dimensions, with differences in inflation, openness, governance quality and various crisis probabilities no longer significant. This suggests that the propensity score weighting procedure brings our data closer to a randomly selected sample. The differences in GDP growth, however, remain significant, even though they shrink somewhat compared to the raw data in Table 4.3.1. This suggests that while the first stage of the IPSWRA helps rebalance the sample, additional regression adjustment through local projections in stage 2 of the IPSWRA is likely necessary in order to control for the remaining control and treatment group differences in observable pre-default characteristics.

Table 4.B.3 presents the second stage of the IPSWRA (equation (4.3.4)) – the local projection estimated on the rebalanced sample. The coefficients on the control variables are generally consistent with the hypothesised signs in Table 4.A.3: for example, GDP growth shows a positive autocorrelation, and is negatively affected by other crises. Higher external debt levels are actually correlated with higher GDP growth, most probably indicating that high-growing countries both want and can borrow more on international markets. The coefficient on the default dummy is equal to the average treatment effect in 4.5.1 (bottom row) and Figure 4.5.1 (solid black line). The control set is able to explain a substantial proportion of the variation in GDP growth, with R^2 statistics of 28% in year 1, rising gradually to 74% in year 10. The high R^2 value at long horizons supports the reliability of our findings on the magnitude of long-run default costs (i.e., the lack thereof).

Table 4.B.2. IPSWRA first stage: Logit regression results

Real GDP per capita growth	-0.132*** (0.036)
Real GDP per capita growth: 1 lag	0.019 (0.020)
Real GDP per capita growth: 2 lags	0.007 (0.022)
GDP deviation from trend	0.876 (0.541)
Real GDP per capital level	-0.000 (0.000)
External public debt to GDP	0.001 (0.015)
External debt to GDP	-0.006 (0.015)
Short-term external debt to GDP	-0.003 (0.018)
Interest payments on external debt to GDP	0.139*** (0.047)
Government share	-0.025 (0.017)
Change in terms of trade	0.078 (0.795)
Change in commodity prices	-3.776*** (1.127)
Change in nominal exchange rate	0.000 (0.000)
Log inflation	-0.232 (0.254)
Openness	-0.002 (0.004)
Current account	0.016 (0.011)
War index	-0.146* (0.080)
Polity index	-0.013 (0.020)
Political transition: continuous measure	-0.373 (0.679)
Political transition dummy	-0.101 (0.893)
Banking crisis dummy	0.576 (0.448)
Currency crisis dummy	-0.017 (0.428)
Coup dummy	0.272 (0.387)
Africa dummy	-0.448 (0.333)
South America dummy	0.491 (0.357)
Asia dummy	-1.096** (0.466)
Number of past defaults	-0.144 (0.122)
Nominal 1-year US T-Bill rate	0.204*** (0.053)
Excess equity return over bills, 17 advanced countries	0.009 (0.009)
Equity dividend yield, 17 advanced countries	-0.027 (0.194)
Observations	3477
Pseudo R-squared	.19

Notes: Regression coefficients on the first-stage predictors (dependent variable: external default one year ahead). Standard errors in parentheses. Regression also includes additional lags of the crisis dummies (political transitions, coups, wars, currency and banking crises), which are insignificant and omitted to save space. Coefficients on these are available from authors upon request.
*, **, ***: Significant at 10%, 5% and 1% levels respectively

Figure 4.B.2. IPSWRA first stage: ROC graph

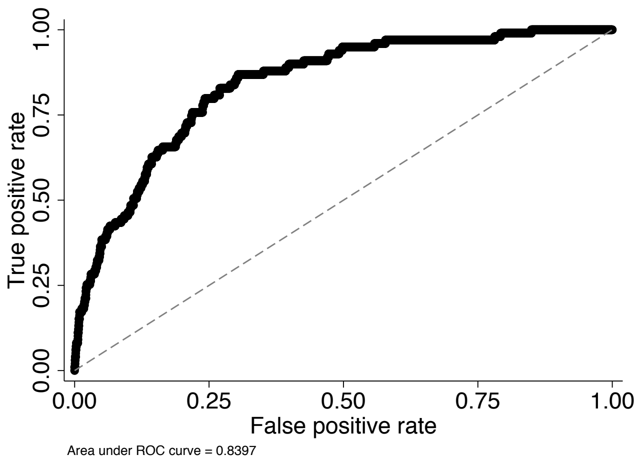


Table 4.B.3. IPSWRA second stage: IPS-weighted regression results

	Year 1	Year 2	Year 3	Year 4	Year 5
F.External default	-2.69*** (0.60)	-3.85*** (1.01)	-3.63*** (1.16)	-3.44** (1.34)	-3.74** (1.54)
Δ real GDP p.c.	0.15* (0.08)	0.06 (0.18)	-0.01 (0.25)	-0.07 (0.31)	-0.17 (0.34)
L.Δ real GDP p.c.	-0.00 (0.03)	-0.02 (0.05)	-0.10 (0.06)	-0.20* (0.11)	-0.16 (0.12)
GDP - HP trend	1.23 (1.46)	3.20 (3.59)	4.11 (4.95)	4.25 (5.73)	4.34 (5.41)
Real GDP p.c. level	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Ext. Public Debt / GDP	-0.01 (0.01)	-0.06** (0.02)	-0.09*** (0.03)	-0.13*** (0.04)	-0.15*** (0.05)
Ext. Debt / GDP	0.02*** (0.01)	0.06*** (0.01)	0.11*** (0.02)	0.15*** (0.02)	0.19*** (0.03)
Govt. share	0.05 (0.06)	0.01 (0.11)	-0.05 (0.14)	-0.14 (0.15)	-0.26 (0.16)
Log inflation	-0.54 (0.39)	-1.08 (0.66)	-1.53* (0.81)	-1.94* (1.01)	-2.62** (1.24)
Openness	0.04*** (0.01)	0.10*** (0.03)	0.13*** (0.04)	0.14*** (0.05)	0.14** (0.06)
Current account	-0.02 (0.04)	-0.01 (0.06)	-0.01 (0.08)	-0.02 (0.10)	0.03 (0.11)
Banking crisis	-2.11* (1.07)	-1.86 (1.31)	-1.97 (1.57)	-2.26 (1.83)	-2.94 (1.79)
L.Banking crisis	0.49 (0.59)	0.12 (0.98)	-0.29 (1.18)	-1.51 (1.34)	-1.68 (1.32)
Currency crisis	0.14 (0.54)	-0.16 (0.90)	-0.31 (1.23)	-0.00 (1.52)	0.07 (1.75)
L.Currency crisis	-1.16** (0.54)	-1.44* (0.80)	-1.36 (0.95)	-1.08 (1.10)	-1.06 (1.17)
Coup	-0.81 (0.73)	-1.94** (0.81)	-1.94* (1.15)	-2.49 (1.50)	-2.78* (1.64)
L.Coup	-0.55 (0.59)	0.29 (1.08)	-0.10 (1.44)	-0.51 (1.45)	-0.74 (1.52)
Polit. transition	-0.49 (0.47)	-0.79 (0.78)	-1.63 (1.15)	-2.90** (1.13)	-3.29** (1.28)
L.Polit. transition	0.10 (0.49)	-0.64 (0.83)	-1.80 (1.21)	-2.86* (1.50)	-3.09* (1.61)
War index	-0.02 (0.16)	-0.10 (0.25)	-0.22 (0.34)	-0.31 (0.44)	-0.30 (0.55)
Polity index	0.05 (0.03)	0.11 (0.06)	0.17* (0.09)	0.22* (0.12)	0.30** (0.14)
ΔCommodity price	3.01* (1.52)	5.95* (3.17)	6.55* (3.92)	7.52* (4.29)	6.60 (4.98)
Constant	1.86 (2.14)	7.41** (3.34)	9.65** (4.44)	12.06** (5.65)	15.87** (6.72)
N	2609	2609	2609	2609	2609
R-squared	.28	.34	.4	.47	.54

Table 4.B.3. IPSWRA second stage: IPS-weighted regression results, continued

	Year 6	Year 7	Year 8	Year 9	Year 10
F.External default	-3.48* (1.77)	-2.98 (1.92)	-2.27 (2.18)	-1.62 (2.51)	-1.74 (2.84)
Δ real GDP p.c.	-0.11 (0.35)	0.12 (0.29)	0.11 (0.24)	-0.07 (0.22)	-0.16 (0.22)
L.Δ real GDP p.c.	-0.11 (0.13)	-0.10 (0.13)	-0.13 (0.11)	-0.10 (0.10)	-0.18 (0.11)
GDP - HP trend	3.39 (4.69)	0.56 (3.56)	0.62 (2.82)	2.98 (2.63)	5.63* (2.91)
Real GDP p.c. level	-0.00*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)	-0.01** (0.00)
Ext. Public Debt / GDP	-0.13** (0.06)	-0.09 (0.07)	-0.02 (0.07)	0.02 (0.08)	0.04 (0.09)
Ext. Debt / GDP	0.20*** (0.03)	0.18*** (0.03)	0.13*** (0.03)	0.10*** (0.04)	0.10** (0.04)
Govt. share	-0.33* (0.18)	-0.30 (0.21)	-0.24 (0.24)	-0.29 (0.27)	-0.39 (0.30)
Log inflation	-3.23** (1.28)	-3.62** (1.46)	-3.84** (1.55)	-3.99** (1.63)	-4.18** (1.73)
Openness	0.11* (0.06)	0.09 (0.07)	0.08 (0.07)	0.07 (0.08)	0.05 (0.08)
Current account	0.05 (0.12)	0.05 (0.13)	0.02 (0.13)	-0.05 (0.14)	-0.13 (0.16)
Banking crisis	-3.22* (1.82)	-4.57** (1.86)	-4.16** (1.87)	-4.05** (1.91)	-2.85 (1.93)
L.Banking crisis	-3.08** (1.38)	-3.08** (1.45)	-2.83* (1.54)	-2.06 (1.67)	-0.31 (1.76)
Currency crisis	0.69 (2.07)	1.18 (2.57)	1.36 (3.20)	1.77 (3.93)	1.88 (4.67)
L.Currency crisis	-1.03 (1.24)	-1.57 (1.28)	-1.42 (1.43)	-0.98 (1.58)	-0.17 (1.67)
Coup	-2.25 (1.87)	-3.12 (1.99)	-4.12* (2.23)	-4.87** (2.21)	-6.21*** (2.04)
L.Coup	-2.16 (1.58)	-2.22 (1.71)	-2.77 (1.70)	-2.92 (1.95)	-2.14 (1.81)
Polit. transition	-4.38*** (1.47)	-4.41*** (1.49)	-4.48*** (1.61)	-4.62*** (1.67)	-3.61** (1.70)
L.Polit. transition	-2.45 (1.81)	-2.61 (1.75)	-2.60 (1.72)	-2.85 (1.92)	-3.76** (1.89)
War index	-0.31 (0.63)	-0.35 (0.69)	-0.36 (0.75)	-0.25 (0.82)	-0.14 (0.87)
Polity index	0.42** (0.17)	0.60*** (0.18)	0.76*** (0.19)	0.89*** (0.21)	1.00*** (0.22)
ΔCommodity price	5.25 (5.48)	5.48 (5.59)	6.88 (5.95)	1.78 (6.24)	1.52 (6.30)
Constant	19.55*** (7.18)	21.40*** (8.12)	22.21** (9.13)	24.49** (9.97)	28.84*** (10.59)
N	2609	2609	2609	2609	2609
R-squared	.59	.64	.67	.7	.74

Notes: This table shows the estimation results for the IPSWRA local projection (dependent variable: cumulative real per capita GDP growth). Clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

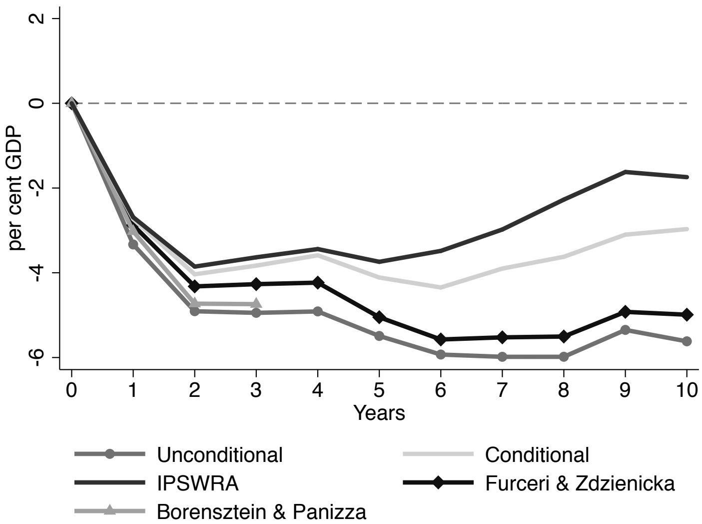
4.B.3 Comparison of baseline results with existing literature

The differences between our baseline sovereign default cost estimate and the existing literature can arise from several potential sources: methodology, control set and sample or default definition. Sections 4.5.2 and 4.B.4.1 discuss the issues around the default definition in more detail. Here we see how different our results would be if we applied the methodology of two existing papers – those of Furceri and Zdzienicka (2012) and Borensztein and Panizza (2008). Why do we choose these two papers? First, they are both based on a broad and comprehensive sample of defaults. Second, Furceri and Zdzienicka (2012) is one of the few studies that consider longer-term default costs, and Borensztein and Panizza (2008) is one of the more widely-cited studies looking at shorter-term default costs.

We generally keep to the exact same specification as the authors of these two papers. For Furceri and Zdzienicka (2012), we use a local projection with country fixed effects, past GDP growth and a country-specific HP-filtered GDP time trend as controls. For Borensztein and Panizza (2008), we use the investment/GDP ratio, government spending/GDP ratio, population growth, civil rights index, change in terms of trade, openness, and a banking crisis dummy as controls. Borensztein and Panizza (2008) further use dummies for groups of countries, and a GDP per capita level in the early 1970s as controls, but no country fixed effects. We instead use country fixed effects, because adding a GDP per capital level in early 1970s would result in missing data and a slightly different sample to our baseline estimates. For the same reason (sample size), we do not include the level of secondary education as a control. This does not, however, have any material bearing on the results. Finally, we use local projections up to a period of 3 years. Borensztein and Panizza (2008) instead use a panel regression with three lags of the sovereign default dummy to achieve a similar end.

Figure 4.B.3 shows that the sovereign default cost estimates obtained using the Borensztein and Panizza (2008) and Furceri and Zdzienicka (2012) specification fall in-between those of our conditional and unconditional estimates. In the long run, the Furceri and Zdzienicka (2012) cost estimate is very close to our unconditional specification. The sizeable differences between our conditional and IPSWRA specification, and those used in the preceding literature show that it is important to control for selection into defaulters using a broad conditioning set, and allowing for non-linearities in selection to arrive at an accurate estimate of the sovereign default cost, particularly in the longer run.

Figure 4.B.3. The cost of default under alternative estimation methods



Notes: Cumulative treatment effect, GDP per capita growth. The Furceri & Zdzienicka specification has past GDP growth and HP-filtered time trend as controls. The Borensztein & Panizza specification has investment/GDP, government spending/GDP, population growth, civil rights index, change in terms of trade, openness, banking crisis dummy as controls, and country fixed effects. Unconditional specification controls for country fixed effects only. Conditional and IP-SWRA specifications control for country fixed effects and the full list of variables in Table 4.A.1.

4.B.4 Alternative treatments for the baseline specification

4.B.4.1 Alternative default definitions

Table 4.B.5 presents the results using the alternative default definitions described in Section 4.A.2. As well as the four alternative definitions in Table 4.A.1, it includes the baseline estimates of Table 4.5.1, and a slightly different manipulation of the *Standard & Poor's* data, which excludes defaults that occurred while a country was still negotiating another default (for example, a country defaulting on its bond obligation while negotiating a default on loans). To ease comparability, all the definitions were extended to use the same sample as baseline. For each default definition, we compare the results under an unconditional local projection, with country fixed effects only, to those using our preferred IPSWRA specification with the full set of controls.

Two broad facts emerge from this comparison. First, sovereign default is costly under all six definitions. Second, controlling for endogeneity using IPSWRA attenuates the size of the cost is substantially, especially at long horizons. This suggests that our main findings discussed in Section 4.5 also hold under these alternative definitions. The size of the short run cost is roughly the same across all six definitions, with the correspnding IPSWRA estimates falling in-between 2.5% and 3.5% of GDP. When it comes to longer horizons, the Beim and Calomiris (2000) and *Standard & Poor's* alternative definitions result in similar costs to baseline, while the cost estimates using the definitions of Reinhart and Rogoff (2011b), Laeven and Valencia (2012) and Detragiache and Spilimbergo (2001) (panels b, d and e) are somewhat higher. This fact is likely to reflect the focus of Laeven and Valencia (2012) and Detragiache and Spilimbergo (2001) on the more severe crisis events, with the cost estimates using these two definitions being, perhaps, more endogenous for this reason.

4.B.4.2 Default magnitude

To calculate a proxy for default magnitude, we make use of the new Bank of Canada CRAG Database (2015) which records total sovereign debt in default for a given country in a given year. Using this, we first record the debt in default to private creditors, or on international financial markets, in proportion to GDP, during the year of sovereign default.[19] We then split our default observations into two groups: those where debt in default was high – "high-magnitude" defaults – and those where debt in default was low. We use two different thresholds to classify defaults as "high-magnitude". The lower threshold of 5% debt-in-default-to-GDP aims to filter out

19. The debt haircut would be a better proxy for magnitude (see, for example Cruces and Trebesch, 2013; Trebesch and Zabel, 2017). However, since default negotiations take some time, information on haircuts is not available at the time of default, and we cannot use it in our local projection or propensity score prediction.

Table 4.B.5. Alternative default definitions

Year	1	2	3	4	5	6	7	8	9	10
(a) Baseline										
Uncond.	-3.33***	-4.91***	-4.95***	-4.91***	-5.49***	-5.93***	-5.98***	-5.98***	-5.35***	-5.62**
	(0.64)	(0.97)	(1.06)	(1.24)	(1.34)	(1.53)	(1.70)	(1.88)	(2.14)	(2.48)
IPSWRA	-2.69***	-3.85***	-3.63***	-3.44***	-3.74***	-3.48**	-2.98	-2.27	-1.62	-1.74
	(0.60)	(1.01)	(1.16)	(1.34)	(1.54)	(1.77)	(1.92)	(2.18)	(2.51)	(2.84)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	92	92	92	92	92	92	92	92	92	92
(b) Reinhart & Rogoff										
Unconditional	-3.77***	-5.31***	-6.12***	-6.17***	-6.64***	-7.14***	-7.41***	-8.23***	-8.19***	-8.50***
	(0.81)	(1.13)	(1.17)	(1.26)	(1.38)	(1.51)	(1.74)	(1.93)	(2.13)	(2.47)
IPSWRA	-2.49***	-3.60***	-4.17***	-4.12***	-4.42***	-4.56***	-4.22***	-4.33**	-4.11*	-4.36*
	(0.64)	(0.96)	(1.14)	(1.21)	(1.41)	(1.55)	(1.78)	(2.10)	(2.33)	(2.57)
Observations	2128	2128	2128	2128	2128	2128	2128	2128	2128	2128
Defaults	81	81	81	81	81	81	81	81	81	81
(c) Beim & Calomiris										
Unconditional	-4.32***	-6.33***	-6.99***	-7.34***	-7.86***	-8.21***	-8.29***	-8.59***	-8.87***	-8.23***
	(0.73)	(1.28)	(1.45)	(1.48)	(1.61)	(1.79)	(2.01)	(2.26)	(2.48)	(2.76)
IPSWRA	-3.38***	-4.90***	-5.25***	-5.12***	-5.43***	-5.25***	-4.72**	-4.12	-3.86	-2.74
	(0.63)	(1.20)	(1.48)	(1.55)	(1.81)	(2.07)	(2.30)	(2.65)	(2.99)	(3.37)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	61	61	61	61	61	61	61	61	61	61
(d) Laeven & Valencia										
Unconditional	-3.89***	-6.46***	-7.21***	-7.51***	-8.51***	-8.93***	-9.01***	-10.09***	-10.29***	-9.29***
	(0.71)	(1.18)	(1.37)	(1.51)	(1.76)	(1.90)	(2.06)	(1.94)	(2.17)	(2.49)
IPSWRA	-3.14***	-5.44***	-6.12***	-6.04***	-7.11***	-7.01***	-6.62***	-6.81***	-6.52***	-5.30**
	(0.66)	(1.12)	(1.35)	(1.45)	(1.77)	(1.91)	(2.03)	(2.04)	(2.17)	(2.42)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	53	53	53	53	53	53	53	53	53	53

Table 4.B.5. Alternative default definitions, continued

(e) Detragiache & Spilimbergo										
Unconditional	-2.40***	-3.76***	-4.98***	-4.78***	-5.80***	-6.81***	-7.07***	-8.21***	-8.75***	-8.96***
	(0.86)	(1.35)	(1.35)	(1.58)	(1.30)	(1.36)	(1.49)	(1.67)	(1.96)	(2.13)
IPSWRA	-1.74***	-2.78***	-3.76***	-3.29***	-4.09***	-4.92***	-4.61***	-5.62***	-6.29***	-6.40***
	(0.72)	(1.13)	(1.16)	(1.28)	(1.18)	(1.31)	(1.42)	(1.59)	(1.71)	(1.99)
Observations	2541	2541	2541	2541	2541	2541	2541	2541	2541	2541
Defaults	92	92	92	92	92	92	92	92	92	92
(f) S & P alternative										
Unconditional	-3.34***	-4.80***	-4.93***	-5.02***	-5.74***	-6.31***	-6.32***	-6.36***	-5.77***	-6.06***
	(0.65)	(0.97)	(1.08)	(1.28)	(1.37)	(1.53)	(1.73)	(1.91)	(2.17)	(2.50)
IPSWRA	-2.69***	-3.72***	-3.59***	-3.50***	-3.96***	-3.83**	-3.25	-2.49	-1.84	-1.82
	(0.61)	(1.02)	(1.19)	(1.39)	(1.58)	(1.81)	(1.99)	(2.27)	(2.59)	(2.95)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	89	89	89	89	89	89	89	89	89	89

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth. Unconditional specification controls for country fixed effects only. IPSWRA specification controls for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. For each definition, we use the longest possible sample; see panel headings for years covered. *, **, ***: Significant at 10%, 5% and 1% levels respectively

those events where debt in default was relatively small, and which may thus have been ignored by both debtors and creditors. The higher threshold of 15% tries to identify the highest-magnitude defaults in our sample and see whether those are exceedingly costly in comparison.

Therefore, we ask two questions: first, is our estimate too low because it includes many low-magnitude defaults that carry almost no cost? And second, do we find that exceedingly large defaults are also exceedingly costly? Our findings rebuff each of the questions. Table 4.B.7 presents the estimation results, with the lower 5% threshold in panel (a) and the 15% threshold in panel (b). The findings in panel (a) correspond to Figure 4.5.3a in the main text. It turns out that the low magnitude defaults are still costly. But defaulting on a larger quantity of debt does increase the cost somewhat, particularly at short horizons, consistent with the findings of Trebesch and Zabel (2017).

4.B.4.3 Defaulting in good and bad times

As Tomz and Wright (2007) have noted, most countries default during bad times, i.e. periods of below-trend GDP growth. Still, our sample contains a substantial number of defaults that occur during good – or normal – times, with GDP growth at or above trend. Comparing the costs of default during good and bad times is interesting for two reasons. First, as previously mentioned, part of our default cost could be endogenous, which simply reflects the poor economic situation of countries that tend to subsequently default, regardless of whether they actually default or not. For this

Table 4.B.7. Large and small defaults

Year	1	2	3	4	5	6	7	8	9	10
(a) Debt defaulted relative to GDP: 5% threshold										
Small	-2.04***	-2.67**	-2.91**	-3.18*	-3.44*	-3.22	-2.47	-1.48	-0.54	-1.17
(no. defaults = 58)	(0.58)	(1.19)	(1.38)	(1.70)	(1.88)	(2.11)	(2.41)	(2.83)	(3.26)	(3.66)
Large	-3.90***	-6.07***	-5.00***	-3.92*	-4.31*	-3.97	-3.93	-3.76	-3.65	-2.82
(no. defaults = 34)	(1.20)	(1.63)	(1.96)	(2.13)	(2.41)	(2.68)	(2.69)	(2.71)	(3.03)	(3.14)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
p-value:										
large = small	0.16	0.09	0.39	0.79	0.77	0.81	0.67	0.53	0.45	0.70
(a) Debt defaulted relative to GDP: 15% threshold										
Small	-2.06***	-2.76***	-2.95***	-3.35**	-3.92**	-3.84**	-3.30	-2.55	-1.89	-2.07
(no. defaults = 70)	(0.52)	(1.01)	(1.19)	(1.50)	(1.69)	(1.91)	(2.14)	(2.46)	(2.84)	(3.19)
Large	-4.94***	-7.77***	-6.08***	-3.77	-3.10	-2.20	-1.82	-1.28	-0.65	-0.58
(no. defaults = 22)	(1.53)	(1.80)	(2.35)	(2.49)	(2.87)	(3.46)	(3.44)	(3.84)	(4.04)	(4.26)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
p-value:										
large = small	0.06	0.01	0.21	0.88	0.79	0.66	0.70	0.77	0.79	0.75

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth. Large defaults are those where the size of debt in default to private creditors, in the year of default, exceeds the chosen threshold. Treatments are based on a simple sample split of our baseline default definition. All figures are IPSWRA estimates controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

to not be the case, we also need default to be costly when economic fundamentals are favourable. Second, as discussed in Section 4.2, a number of theoretical models impose a higher default cost during good times to justify the relative rarity of defaulting when the country is doing well.

We split our sample of defaults into two subsamples – defaults in good and bad times – and compare the results between these two treatments. Good-time defaults are those that occurred when a country's GDP was above trend, and bad-times – below trend, with the trend calculated using a one-sided HP filter with a smoothing parameter of 6.25 (Ravn and Uhlig, 2002).

Table 4.B.8 presents the results. Panel (a) compares deviations from the trend in the year before default, and panel (b) – in the three years preceding default. Panel (b) corresponds to Figure 4.5.3b in the main text. We find that defaulting in good times is costly under both specifications, which suggests that our results are not driven by a subsample of defaulters who simply have poor economic fundamentals. However, defaulting in good times is no more costly than defaulting during

Table 4.B.8. Defaulting in good and bad times

Year	1	2	3	4	5	6	7	8	9	10
(a) 1 year before default										
Bad Times	-3.14***	-4.65***	-4.12***	-3.84***	-4.71***	-6.09***	-5.67***	-4.43*	-3.07	-2.80
(no. defaults = 59)	(0.79)	(1.29)	(1.52)	(1.63)	(1.91)	(2.01)	(2.13)	(2.52)	(2.93)	(3.15)
Good Times	-1.97**	-2.58	-2.85	-2.80	-2.18	0.70	1.35	1.19	0.70	-0.05
(no. defaults = 33)	(0.92)	(1.58)	(2.32)	(2.70)	(2.84)	(3.06)	(3.26)	(3.44)	(3.82)	(4.32)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
p-value:										
good = bad	0.34	0.31	0.68	0.76	0.48	0.05	0.06	0.15	0.39	0.56
(b) 1–3 years before default										
Bad Times	-3.20***	-3.88***	-3.28**	-3.06*	-3.64*	-3.21	-2.74	-1.86	-1.66	-2.61
(no. defaults = 59)	(0.74)	(1.29)	(1.47)	(1.72)	(1.94)	(2.21)	(2.26)	(2.37)	(2.66)	(2.81)
Good Times	-1.82*	-3.81***	-4.24***	-4.08**	-3.91*	-3.95	-3.39	-2.98	-1.55	-0.27
(no. defaults = 33)	(0.98)	(1.50)	(1.63)	(1.88)	(2.12)	(2.51)	(2.94)	(3.45)	(3.88)	(4.45)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
p-value:										
good = bad	0.26	0.97	0.65	0.68	0.92	0.81	0.85	0.76	0.98	0.59

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth, conditional on default occurring during or times. Good times are defined as growth above HP-filtered trend, bad times – growth below trend, either in the year before, or over the three years before default. Treatments are based on a simple sample split of our baseline default definition. All figures are IPSWRA estimates controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses.
*, **, ***: Significant at 10%, 5% and 1% levels respectively.

bad times, which seems to go against the assumptions often made in theoretical literature.

4.B.4.4 Default cost among different groups of countries

Table 4.B.9 computes the cost of default estimate for different groups of countries. Panel (a) compares the cost estimate for "heavily indebted poor countries" (HIPC) with the rest of the sample, using the country grouping provided by the World Bank. These countries constitute a little less than half of all defaults in our sample. It turns out that the cost of default does not differ substantially across country groups: both heavily indebted poor countries, and other defaulting economies experience similar costs across the time horizon. The reasons for why these costs arise may, however be different. The analysis in Section 4.7 suggest that autarky and banking distress are the two main channels responsible for generating the default costs. HIPCs are likely to have undeveloped financial systems, hence the role for banking distress in these defaults is limited. But they are also likely to have higher external dependence, either through borrowing or aid flows, and hence suffer more from autarky. Table

Table 4.B.9. Default cost among different groups of countries

Year	1	2	3	4	5	6	7	8	9	10
(a) Countries grouped by economic development										
Default + not HIPC	-2.91***	-4.98***	-4.62***	-3.80*	-3.93*	-3.50	-2.94	-2.87	-1.91	-0.70
(no. defaults = 52)	(0.82)	(1.55)	(1.77)	(2.05)	(2.21)	(2.46)	(2.73)	(3.19)	(3.68)	(4.34)
Default in a HIPC	-2.43***	-2.56**	-2.50*	-3.03**	-3.52*	-3.47	-3.02	-1.59	-1.29	-2.94
(no. defaults = 40)	(0.85)	(1.17)	(1.30)	(1.50)	(1.86)	(2.14)	(2.24)	(2.38)	(2.61)	(2.36)
(b) Countries grouped by continent										
Africa	-2.15***	-2.67*	-2.68	-2.82	-3.34	-2.94	-2.89	-1.76	-1.40	-2.90
(defaults: 42)	(0.70)	(1.51)	(1.72)	(2.08)	(2.27)	(2.46)	(2.62)	(2.83)	(3.11)	(3.15)
Americas	-2.87***	-4.39***	-3.33***	-3.22***	-3.21*	-3.36	-2.68	-2.44	-2.47	-2.14
(defaults: 35)	(1.18)	(1.48)	(1.31)	(1.29)	(1.69)	(2.36)	(2.34)	(2.72)	(2.99)	(3.52)
Other	-3.90***	-6.25***	-7.00**	-5.65	-5.93	-5.31	-3.81	-3.45	-0.63	2.37
(defaults: 15)	(1.55)	(2.08)	(3.06)	(3.55)	(4.04)	(4.53)	(5.47)	(6.55)	(7.94)	(9.38)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth. IPSWRA estimates using country fixed effects. Clustered standard errors in parentheses. Heavily indebted poor countries are those countries currently eligible for special assistance from the World Bank and the IMF, due to their high levels of poverty and debt. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

4.B.9 panel (b) compares the default cost estimates across different continents and finds that they are, broadly, similar. The cost estimate for African countries loses significance somewhat earlier than that for other continents, but this may be driven by larger measurement error for these countries' data rather than underlying cost differentials.

4.B.4.5 Robustness to alternative regression specifications

To check the stability of our results under different variations of the IPSWRA method, we explore alternative ways of calculating the propensity score and selecting the control group. We also check if common trends across countries matter for our results by adding year fixed effects to our baseline specification. Table 4.B.10 shows the results. The top row shows our baseline specification from Section 4.5 and the second row shows the result with year fixed effects in the second stage. The default cost is somewhat lower with year fixed effects compared to the baseline estimate and the recovery from default happens at a faster pace.

Recall that we truncate the estimated inverse propensity scores at 10 following Imbens (2004). Additionally, the control group in the baseline includes those countries still negotiating a past default. Finally, following Jordà and Taylor (2016), we use a somewhat larger set of predictors than controls.

Table 4.B.10. Alternative propensity scores and control groups

Year	1	2	3	4	5	6	7	8	9	10	Obs.
Baseline	-2.69***	-3.85***	-3.63***	-3.44***	-3.74***	-3.48**	-2.98	-2.27	-1.62	-1.74	2609
(defaults: 92)	(0.60)	(1.01)	(1.16)	(1.34)	(1.54)	(1.77)	(1.92)	(2.18)	(2.51)	(2.84)	
Year fixed effects	-2.34***	-3.05***	-2.52**	-1.98	-1.99	-1.70	-1.24	-0.62	0.38	0.69	2609
(defaults: 92)	(0.62)	(1.04)	(1.22)	(1.43)	(1.58)	(1.79)	(1.96)	(2.28)	(2.66)	(2.90)	
Less truncation	-2.61***	-3.82***	-3.79***	-3.65***	-3.97**	-3.66*	-3.14	-2.33	-1.63	-1.57	2609
(defaults: 92)	(0.61)	(1.07)	(1.28)	(1.49)	(1.71)	(1.96)	(2.11)	(2.40)	(2.77)	(3.17)	
Low weight in	-2.67***	-3.73***	-3.66***	-3.56***	-4.03***	-3.77**	-3.25	-2.57	-1.82	-1.64	2606
default (defaults: 89)	(0.60)	(1.06)	(1.27)	(1.48)	(1.67)	(1.88)	(2.09)	(2.40)	(2.72)	(3.12)	
Clean control	-3.13***	-4.37***	-4.42***	-4.50***	-5.14***	-4.90**	-4.38*	-3.51	-3.01	-2.98	1939
group (defaults: 89)	(0.64)	(1.12)	(1.37)	(1.63)	(1.90)	(2.15)	(2.46)	(2.79)	(3.14)	(3.58)	
Predictors as	-2.48***	-3.21***	-2.59**	-1.98	-2.01	-1.69	-1.16	-0.48	0.39	0.66	2609
controls (defaults: 92)	(0.63)	(1.03)	(1.19)	(1.42)	(1.59)	(1.81)	(1.95)	(2.24)	(2.58)	(2.88)	
Varying the	-2.31***	-3.51***	-2.83***	-2.49*	-2.86*	-2.49	-1.95	-1.57	-1.68	-1.74	
sample size	(0.64)	(1.06)	(1.21)	(1.40)	(1.61)	(1.84)	(1.96)	(2.22)	(2.42)	(2.84)	
Observations	3477	3477	3369	3262	3155	3047	2937	2829	2719	2609	
Defaults	99	99	98	98	97	97	97	96	94	92	

Notes: Average treatment effect of sovereign default on cumulative real GDP per capita growth. Clustered standard errors in parentheses. IPSWRA specifications control for country fixed effects and the full list of variables in Table 4.A.1. Less truncation: inverse propensity score weights truncated at 20 instead of 10. Low weight in default: alternative S & P default definition, IPS-weights equal to one during default. Clean control group: countries negotiating a past default excluded from the control group. Predictors as controls: all predictors from the first stage are also included as controls in the second stage. Varying the sample size: all baseline results are estimated on a consistent sample by imposing horizon restrictions. This specification lossens these restrictions. *, **, ***: Significant at 10%, 5% and 1% levels respectively

The third row shows the results using a larger truncation threshold of 20. This effectively makes the rebalancing stronger but less robust. The estimated effect does not differ substantially compared to baseline. The fourth shows the results using an alternative inverse propensity score weight for those observations that are still negotiating a past default. We set the weight for in-default observations equal to 1 (zero default probability). This weight is smaller than that in our baseline specification and gives these countries a smaller prominence among the control group. We use the alternative S & P default definition from 4.B.4.1 for this exercise (no new default can occur while the country is still negotiating a past default). The results under this specification, however, remain close to baseline.

The results in the fifth row provide further robustness to the control group choice. Rather than treating countries negotiating a past default as "normal" observations or giving them a low weight, we remove them from the control group. Even though this alters the sample substantially, it has relatively little bearing on our results.

The sixth row uses the same controls and predictors in both stages of the IPSWRA, adding variables such as the advanced economies' dividend yield as controls in the LP (these variables are listed in Table 4.A.1 under "predictors used in Stage 1 (logit) only”). The inclusion of these variables is likely to make the estimates somewhat more robust but less precise. As in the baseline specification, sovereign default is costly but the cost goes away in the long run. The medium and long run cost estimate is slightly smaller and less significant. Including the predictors also as control variables strengthens our baseline finding: sovereign default is costly in the short and medium run, but the cost is temporary.

The bottom row of Table 4.B.10 varies the sample size over the LP horizon. Our baseline estimates use a consistent estimate for each horizon h. This means that we include defaults up to 2001 only, since we need data on outcomes up to 2010 for each observation. The year 1 cost estimate in Table 4.B.10 bottom row includes all defaults with one-year ahead GDP data, i.e. all defaults up to 2009, 99 in total. The year 2 estimate includes all defaults up to 2008, and so on. Extending the sample over the LP horizon has no effect on our results.

4.B.5 Amplification of the cost: additional results

4.B.5.1 Sovereign defaults and banking crises

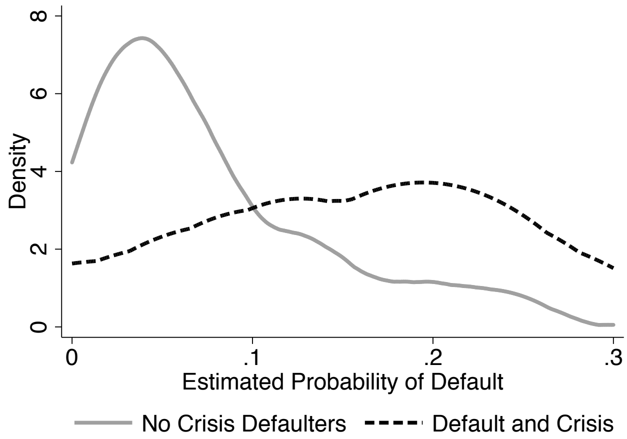
Estimating the cost of joint sovereign-banking crises faces two main challenges. First, our control and predictor variables are primarily selected to forecast sovereign defaults, and may thus do a poor job of forecasting banking crises and joint sovereign-banking crisis events. Second, we want to exclude events where distress in the banking sector, rather than the sovereign, is the main cause of the crisis. To this end, we exclude those events where the banking crisis preceded sovereign default, or the two occurred in the same year and the banking panic was the primary cause of the default. This selection process is, however, by nature imprecise, and our joint event list could include some where the banking, not the sovereign crisis is the main driver of the downturn.

To see if stage 1 of our IPSWRA procedure – complemented by additional predictor variables relating to credit and loan-deposit ratios – does a good job at forecasting these joint events, Figure 4.B.1 compares the propensity scores – i.e. the estimated default probabilities – for those defaults which are followed by a systemic banking crisis (dashed line), with those of standalone default events (solid line). The logit prediction does well at predicting both types of events, with predicted probabilities of 10%–20% substantially above the sample default average of 2%. The prediction is slightly more accurate for joint sovereing-banking crisis events with, on average, higher predicted default probabilities. This suggests that IPSWRA does a good job at controlling for endogeneity of selection into joint sovereign-banking crisis as well as sovereign defaults more generally.

To evaluate the importance of the joint sovereign-banking crisis definition, Figure 4.B.2 and Table 4.B.1 estimate the cost of default for several alternative joint crisis definitions: standalone defaults, events where a systemic banking crisis precedes the sovereign default, those where the two crises occur in the same year – regardless of which type of crisis was the underlying cause – and those where the banking crisis followed the sovereign default. Figure 4.B.2 shows both the unconditional and IPSWRA cost estimates, and Table 4.B.1 provides more detail on the IPSWRA estimates. These estimates should only be regarded as indicative, because of the small number of events in each subgroup. Nevertheless, a clear broad pattern emerges.

All three types of joint sovereign-banking crisis are substantially more costly than the standalone defaults, by a factor of 2 or more depending on the time horizon. The cost of joint events differs somewhat depending on the definition, with the cost of joint events occurring in the same year somewhat less persistent than the other two categories. Under all three definitions, the cost of joint sovereign-banking crises remains substantial, both unconditionally and under IPSWRA (Figure 4.B.2). The 67 standalone defaults in the sample are, on average, around 1 percentage point less costly than our baseline estimate in Table 4.5.1. Taken together, these results sug-

Figure 4.B.1. Predicted default probabilities for "standalone" sovereign defaults, and those followed by a systemic banking crisis



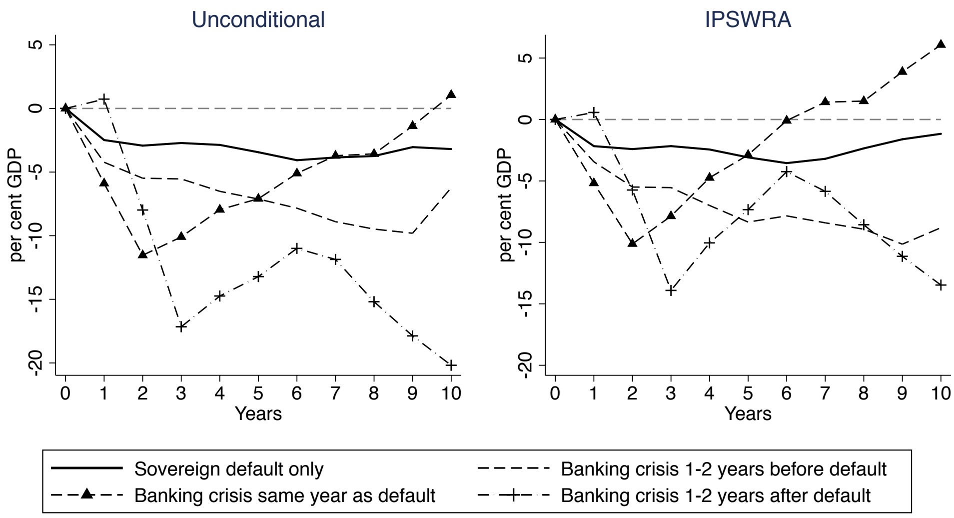
Notes: Kernel density plot of the predicted default probability based on IPSWRA first-stage logit model (as specified in equation (4.3.1), predictors as in Table 4.A.1.). The "default and crisis" observations are those where a sovereign default is followed by a systemic banking crisis within two years. The "no crisis defaulters" are defaults not followed by a systemic banking crisis.

gest that a considerable part of the sovereign default cost arises through the banking channel, and that sovereign-banking interactions are an important factor for amplifying the cost. Finally, if we combine the different joint sovereign-banking crisis events into a single ±1-year or ±2-year window, giving us more crisis observations and hence more precision, the cost of these joint events also remains significantly above that of standalone crises (results available from authors upon request).

4.B.5.2 Currency and political crises

Tables 4.B.2 and 4.B.3 provide the cost estimates for sovereign defaults which coincide, respectively, with currency or political crises. For both of these crisis events, standalone defaults are still costly, and the occurrence of another crisis increases the default cost, but by much less than that of a systemic banking crisis in Table 4.6.1.

Figure 4.B.2. Cost of sovereign default and systemic banking crises: alternative crisis definitions



Notes: Cumulative treatment effect, GDP per capita growth. Unconditional specification controls for country fixed effects only. IPSWRA specification controls for country fixed effects and the full list of variables in Table 4.A.1.

Table 4.B.1. Cost of sovereign default and systemic banking crises: alternative crisis definitions

Year	1	2	3	4	5	6	7	8	9	10
Default + no crisis	-2.17***	-2.41**	-2.17*	-2.44	-3.08*	-3.55*	-3.20	-2.35	-1.61	-1.16
(no. defaults = 67)	(0.70)	(1.10)	(1.21)	(1.50)	(1.65)	(1.93)	(2.17)	(2.52)	(2.95)	(3.50)
Default + crisis 1-2y before	-3.44	-5.49***	-5.54**	-7.00***	-8.35***	-7.84***	-8.41***	-8.93***	-10.15***	-8.81**
(no. defaults = 10)	(2.23)	(2.22)	(2.40)	(2.92)	(2.55)	(2.95)	(3.52)	(3.75)	(4.03)	(4.45)
p-value: crisis = no crisis	0.60	0.24	0.24	0.21	0.11	0.26	0.24	0.19	0.13	0.24
Default + crisis same year	-5.18***	-10.12***	-7.86**	-4.74	-2.88	-0.09	1.41	1.50	3.88	6.07
(no. defaults = 10)	(1.60)	(2.59)	(3.39)	(3.82)	(4.87)	(5.85)	(5.57)	(5.95)	(6.64)	(6.97)
p-value: crisis = no crisis	0.07	0.01	0.12	0.57	0.97	0.57	0.43	0.54	0.44	0.33
Default + crisis 1-2y after	0.57	-5.74**	-13.91***	-10.04***	-7.34*	-4.24	-5.85	-8.57	-11.14	-13.47**
(no. defaults = 3)	(1.26)	(2.48)	(2.51)	(1.95)	(4.29)	(6.79)	(7.59)	(6.76)	(6.83)	(6.32)
p-value: crisis = no crisis	0.12	0.25	0.00	0.01	0.38	0.92	0.74	0.41	0.22	0.11
Observations	2245	2245	2245	2245	2245	2245	2245	2245	2245	2245

Notes: Average treatment effect on cumulative real GDP per capita growth: defaults that are preceded, accompanied or followed by a systemic banking crisis, compared to those that are not. Treatments are based on a simple sample split of our baseline default definition. All figures are IPSWRA estimates controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Table 4.B.2. Sovereign default and currency crises

Year	1	2	3	4	5	6	7	8	9	10
Default + no Crisis	-1.92***	-2.87***	-2.38	-2.07	-2.69	-2.22	-2.21	-1.75	-0.73	-1.15
(no. defaults = 57)	(0.61)	(1.22)	(1.45)	(1.71)	(1.85)	(2.16)	(2.27)	(2.38)	(2.66)	(2.75)
Default + Crisis	-4.06***	-5.62***	-5.89***	-5.89***	-5.62***	-5.75**	-4.36	-3.20	-3.22	-2.81
(no. defaults = 35)	(1.11)	(1.45)	(1.73)	(1.86)	(2.27)	(2.49)	(2.96)	(3.52)	(4.04)	(4.72)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
p-value:										
crisis = no crisis	0.08	0.12	0.11	0.11	0.27	0.24	0.53	0.70	0.55	0.71

Notes: Average treatment effect on cumulative real GDP per capita growth: defaults that coincide, or do not coincide, with a currency crisis, within a one-year window. Treatments are based on a simple sample split of our baseline default definition. All figures are IPSWRA estimates controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Table 4.B.3. Sovereign default and political crises

Year	1	2	3	4	5	6	7	8	9	10
Default + no Crisis	-1.97***	-3.26***	-2.26**	-2.70**	-2.97**	-2.71	-2.35	-1.40	-1.44	-1.76
(no. defaults = 65)	(0.61)	(1.06)	(1.05)	(1.18)	(1.41)	(1.72)	(1.69)	(1.74)	(1.99)	(2.42)
Default + Crisis	-4.12***	-5.05***	-6.38**	-4.92	-5.27	-5.03	-4.24	-4.02	-1.99	-1.72
(no. defaults = 27)	(1.15)	(2.03)	(2.75)	(3.11)	(3.56)	(4.03)	(4.46)	(5.07)	(5.80)	(6.37)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
p-value:										
crisis = no crisis	0.08	0.42	0.16	0.49	0.54	0.60	0.69	0.61	0.92	0.99

Notes: Average treatment effect on cumulative real GDP per capita growth: defaults that coincide, or do not coincide, with a political crisis, within a one-year window. A political crisis is defined as a coup, a political transition or a war intensity of more than 3 (MEPV total conflict variable; scale 0 – 20: sum of interstate and civil conflict, each scaled from 0 to 10). Treatments are based on a simple sample split of our baseline default definition. All figures are IPSWRA estimates controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

4.B.6 Decomposition of the cost: additional details

Table 4.B.1. Share of each component in GDP for defaulters

	Ratio to GDP
Consumption	0.67
Investment	0.18
Government consumption	0.15
Exports	0.21
Imports	0.27

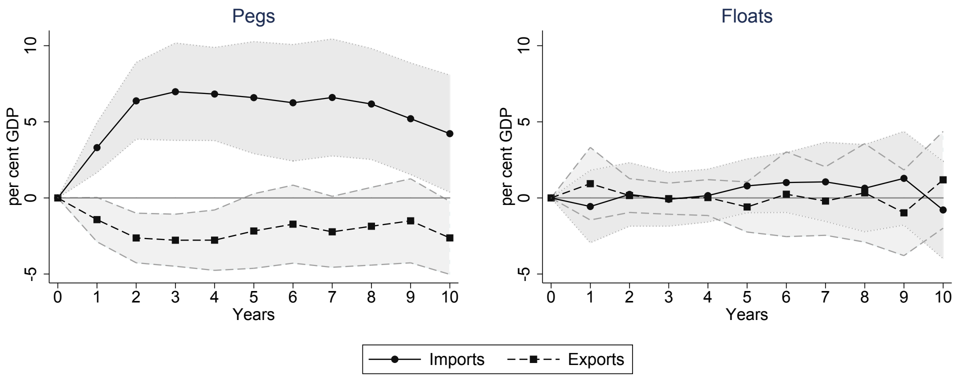
The shares refer to the year before the default episode.

This section provides additional results to help with interpretation of the GDP component cost decomposition in Section 4.7. Table 4.B.1 shows the share of each component in GDP before the default. The larger the share, the higher should be the contribution of this component to the cost in Figure 4.7.1a, all other things being equal. Investment and imports have relatively small GDP shares, which highlights the disproportionally large adjustment in these variables after the default takes place.

Figure 4.B.1 and Table 4.B.2 provide further detail on how the default cost varies according to the exchange rate regime. Figure 4.B.1 decomposes the external adjustment undertaken by pegs and floats into changes in imports and exports. For pegged exchange rates, it is imports that bear the brunt of the adjustment, which is likely to contribute to the high costs experienced by these economies. For floats, neither exports nor imports change much after a default, as little external adjustment needs to be undertaken. Table 4.B.2 decomposes the cost for pegs into different components of GDP. As in the baseline specification (Table 4.7.1), most of the adjustment takes place via investment and imports.

Table 4.B.3 shows the GDP component changes after a default which is followed by a systemic banking crisis. Falls in investment and imports, again, record much larger drops than the other GDP components.

Figure 4.B.1. Pegged and floating country defaults: Trade



Notes: IPSWRA estimates of the change in imports and exports relative to GDP in year 0. Shaded bands indicate 90% confidence intervals. Pegged exchange rates include countries with no separte legal tender, hard pegs, crawling pegs and narrow exchange rate corridors.

Table 4.B.2. Default of countries with pegged exchange rates: impact on components of GDP

Year	1	2	3	4	5	6	7	8	9	10
Investment	-4.06***	-5.91***	-5.22***	-5.01***	-5.37***	-4.68***	-4.26**	-3.24*	-2.28	-1.37
	(1.27)	(1.61)	(1.71)	(1.71)	(1.86)	(1.92)	(1.93)	(1.72)	(1.52)	(1.77)
Consumption	-1.92**	-3.17***	-3.92***	-3.98***	-4.31***	-4.40***	-3.53*	-4.22**	-4.02**	-3.95**
	(0.84)	(1.07)	(1.43)	(1.46)	(1.50)	(1.75)	(2.01)	(1.99)	(2.01)	(1.99)
Government	-0.21	-1.02***	-1.32***	-1.20***	-0.89*	-0.70	-0.44	-0.28	0.15	-0.40
Consumption	(0.22)	(0.42)	(0.44)	(0.45)	(0.46)	(0.43)	(0.46)	(0.48)	(0.49)	(0.41)
Exports	-1.42	-2.63***	-2.78***	-2.77**	-2.17	-1.72	-2.22	-1.86	-1.50	-2.62*
	(0.89)	(0.99)	(1.04)	(1.21)	(1.49)	(1.56)	(1.42)	(1.56)	(1.68)	(1.46)
Imports	3.31***	6.38***	6.98***	6.83***	6.59***	6.25***	6.60***	6.17***	5.21***	4.22*
	(1.00)	(1.53)	(1.95)	(1.86)	(2.24)	(2.33)	(2.34)	(2.22)	(2.23)	(2.34)
Real GDP	-3.60***	-6.00***	-6.19***	-6.16***	-6.11***	-5.39***	-4.21*	-3.86	-3.27	-4.43*
(total)	(0.96)	(1.41)	(1.51)	(1.55)	(1.78)	(2.04)	(2.23)	(2.43)	(2.53)	(2.57)
Observations	2609	2609	2609	2609	2609	2609	2609	2609	2609	2609
Defaults	50	50	50	50	50	50	50	50	50	50

Notes: Pegged exchange rates include countries with no separte legal tender, hard pegs, crawling pegs and narrow exchange rate corridors. The outcome variable is the absolute change in a GDP component between t and t + h, scaled by the GDP level at t. Here t is the year before default, and h is the horizon. IPSWRA specification, controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. Effects do not sum exactly to the treatment effect on GDP; small residual. *, **, ***: Significant at 10%, 5% and 1% levels respectively

Table 4.B.3. Default followed by a systemic banking crisis: impact on components of GDP

Year	1	2	3	4	5	6	7	8	9	10
Investment	-4.33***	-7.08***	-6.72***	-2.15	-1.05	2.04	1.61	2.97	3.81	5.25*
	(1.76)	(2.31)	(2.04)	(1.59)	(2.22)	(2.21)	(2.49)	(2.36)	(2.49)	(2.88)
Consumption	-1.59	-5.23**	-5.56	-5.18*	-4.18	-3.38	-4.58	-5.48	-3.47	-2.43
	(1.57)	(2.56)	(3.54)	(2.89)	(3.37)	(4.23)	(4.98)	(4.40)	(4.68)	(4.86)
Government	0.14	-2.70***	-3.05***	-2.74***	-1.70*	-0.98	-0.67	-0.50	-0.50	-0.35
Consumption	(0.51)	(0.87)	(0.96)	(1.04)	(0.95)	(0.84)	(0.78)	(0.77)	(0.84)	(0.90)
Exports	-1.92***	-2.59*	-3.05	-1.40	4.09	3.36	2.05	0.46	-0.89	-0.93
	(0.78)	(1.34)	(2.29)	(2.48)	(5.51)	(3.91)	(2.83)	(1.93)	(1.97)	(2.18)
Imports	2.45*	7.95***	9.22**	5.97*	-0.45	-0.80	1.68	1.39	2.05	0.49
	(1.36)	(1.79)	(4.05)	(3.36)	(5.72)	(4.77)	(3.81)	(3.44)	(3.81)	(4.24)
Real GDP	-4.39***	-9.42***	-9.51***	-5.91*	-3.69	0.02	0.13	-0.51	1.02	1.99
(total)	(1.54)	(2.28)	(3.03)	(3.11)	(4.11)	(5.06)	(5.23)	(5.48)	(6.16)	(6.62)
Observations	2245	2245	2245	2245	2245	2245	2245	2245	2245	2245
Defaults	11	11	11	11	11	11	11	11	11	11

Notes: Sovereign default followed by a banking crisis within two years: average treatment effect on individual components of GDP. The outcome variable is the absolute change in a GDP component between t and t + h, scaled by the GDP level at t. Here t is the year before default, and h is the horizon. IPSWRA specification, controlling for country fixed effects and the full list of variables in Table 4.A.1. Clustered standard errors in parentheses. Effects do not sum exactly to the treatment effect on GDP; small residual. *, **, ***: Significant at 10%, 5% and 1% levels respectively

References

Acharya, Viral V., Tim Eisert, Christian Eufinger, and Christian W. Hirsch. 2018. "Real Effects of the Sovereign Debt Crisis in Europe: Evidence from Syndicated Loans." *Review of Financial Studies* 31 (8): 2855–2896.

Aguiar, Mark, and Gita Gopinath. 2006. "Defaultable Debt, Interest Rates and the Current Account." *Journal of International Economics* 69 (1): 64–83.

Aguiar, Mark, and Gita Gopinath. 2007. "Emerging Market Business Cycles: The Cycle Is the Trend." *Journal of Political Economy* 115 (1): 69–102.

Andrade, Sandro C, and Vidhi Chhaochharia. 2017. "The Costs of Sovereign Default: Evidence from the Stock Market." Working Paper.

Angrist, Joshua D., Òscar Jordà, and Guido Kuersteiner. 2017. "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited." *Journal of Business & Economic Statistics*, 1–17.

Arellano, Cristina. 2008. "Default Risk and Income Fluctuations in Emerging Economies." *American Economic Review* 98 (3): 690–712.

Asonuma, Tamon, Marcos Chamon, and Akira Sasahara. 2016. "Trade Costs of Sovereign Debt Restructurings: Does a Market-Friendly Approach Improve the Outcome?" IMF Working Paper 16/222.

Asonuma, Tamon, and Christoph Trebesch. 2016. "Sovereign Debt Restrurings: Preemptive or Post-Default." *Journal of the European Economic Association* 14 (1): 175–214.

Auerbach, Alan J, and Yuriy Gorodnichenko. 2012. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy After the Financial Crisis*. Edited by Alberto Alesina and Francesco Giavazzi. Chicago, Ill.: University of Chicago press, 63–98.

Balteanu, Irina, and Aitor Erce. 2018. "Linking Bank Crises and Sovereign Defaults: Evidence from Emerging Markets." *IMF Economic Review* 66 (4): 617–664.

Beck, Thorsten, Aslı Demirgüç-Kunt, and Ross Levine. 2010. "Financial Institutions and Markets across Countries and over Time: The Updated Financial Development and Structure Database." *World Bank Economic Review* 24 (1): 77–92.

Beers, David T., and John Chambers. 2006. "Default Study: Sovereign Defaults at 26-Year Low, to Show Little Change in 2007."

Beers, David T, and Jean-Sébastien Nadeau. 2015. "Database of Sovereign Defaults, 2015." Technical Report 101.

Beim, David O, and Charles W Calomiris. 2000. *Emerging Financial Markets*. New York.

Bocola, Luigi. 2016. "The Pass-Through of Sovereign Risk." *Journal of Political Economy* 124 (4): 879–926.

Borensztein, Eduardo, and Ugo Panizza. 2008. "The Costs of Sovereign Default." IMF Working Paper 08/238.

Borensztein, Eduardo, and Ugo Panizza. 2010. "Do Sovereign Defaults Hurt Exporters?" *Open Economies Review* 21 (3): 393–412.

Cantor, Richard, and Frank Packer. 1996. "Determinants and Impact of Sovereign Credit Ratings." *Economic Policy Review* 2 (2):

Cerra, Valerie, and Sweta Chaman Saxena. 2008. "Growth Dynamics: The Myth of Economic Recovery." *American Economic Review* 98 (1): 439–457.

Cole, Harold L, and Timothy J Kehoe. 1996. "A Self-Fulfilling Model of Mexico's 1994–1995 Debt Crisis." *Journal of International Economics* 41 (3): 309–330.

Cole, Harold L, and Timothy J Kehoe. 2000. “Self-Fulfilling Debt Crises.” *Review of Economic Studies* 67 (1): 91–116.

Cruces, Juan J., and Christoph Trebesch. 2013. “Sovereign Defaults: The Price of Haircuts.” *American Economic Journal: Macroeconomics* 5 (3): 85–117.

De Paoli, Bianca, Glenn Hoggarth, and Victorical Saporta. 2009. “Output Costs of Sovereign Crises: Some Empirical Estimates.” Bank of England Working Paper 362.

Detragiache, Enrica, and Antonio Spilimbergo. 2001. “Crises and Liquidity: Evidence and Interpretation.” IMF Working Paper 01/2.

Driscoll, John C, and Aart C Kraay. 1998. “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data.” *Review of Economics and Statistics* 80 (4): 549–560.

Eaton, Jonathan, and Mark Gersovitz. 1981. “Debt with Potential Repudiation: Theoretical and Empirical Analysis.” *Review of Economic Studies* 48 (2): 289–309.

Furceri, Davide, and Aleksandra Zdzienicka. 2012. “How Costly Are Debt Crises?” *Journal of International Money and Finance* 31 (4): 726–742.

Gennaioli, Nicola, Alberto Martin, and Stefano Rossi. 2014. “Sovereign Default, Domestic Banks, and Financial Institutions.” *Journal of Finance* 69 (2): 819–866.

Gennaioli, Nicola, Alberto Martin, and Stefano Rossi. 2018. “Banks, government Bonds, and Default: What do the data Say?” *Journal of Monetary Economics* 98: 98–113.

Gornemann, Nils. 2014. “Sovereign Default, Private Investment, and Economic Growth.” Working paper.

Hébert, Benjamin, and Jesse Schreger. 2017. “The Costs of Sovereign Default: Evidence from Argentina.” *American Economic Review* 107 (10): 3119–45.

Ilzetzki, Ethan, Carmen M Reinhart, and Kenneth S Rogoff. 2017. “Exchange Arrangements Entering the 21st Century: Which Anchor Will Hold?” NBER Working Paper 23134.

Imbens, Guido W. 2004. “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review.” *Review of Economics and Statistics* 86 (1): 4–29.

Jordà, Òscar. 2005. “Estimation and Inference of Impulse Responses by Local Projections.” *American Economic Review* 95 (1): 161–182.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. “The Rate of Return on Everything, 1870–2015.” *Quarterly Journal of Economics* forthcoming:

Jordà, Òscar, Björn Richter, Moritz Schularick, and Alan M. Taylor. 2017. “Bank Capital Redux: Solvency, Liquidity, and Crisis.” NBER Working Paper 23287.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. “When Credit Bites Back.” *Journal of Money, Credit and Banking* 45 (2): 3–28.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2016. “Sovereigns Versus Banks: Credit, Crises, and Consequences.” *Journal of the European Economic Association* 14 (1): 45–79.

Jordà, Òscar, and Alan M. Taylor. 2016. “The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy.” *Economic Journal* 126 (590): 219–255.

Kaminsky, Graciela L. 2006. “Currency Crises: Are They All the Same?” *Journal of International Money and Finance* 25 (3): 503–527.

Laeven, Luc, and Fabian Valencia. 2008. “Systemic Banking Crises: A New Database.” IMF Working Paper 08/224.

Laeven, Luc, and Fabian Valencia. 2012. “Systemic Banking Crises Database: An Update.” IMF Working Paper 12/163.

Levy-Yeyati, Eduardo, and Ugo Panizza. 2011. “The Elusive Costs of Sovereign Defaults.” *Journal of Development Economics* 94 (1): 95–105.

Manasse, Paolo, and Nouriel Roubini. 2009. ""Rules of Thumb" for Sovereign Debt Crises." *Journal of International Economics* 78 (2): 192–205.

Manasse, Paolo, Nouriel Roubini, and Axel Schimmelpfennig. 2003. "Predicting Sovereign Debt Crises." IMF Working Paper 03/221.

Marshall, Monty. 2014. "Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2013." Center for Systemic Peace.

Marshall, Monty, Ted Gurr, and Keith Jaggers. 2014. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2013." Center for Systemic Peace.

Marshall, Monty, and Donna Marshall. 2014. "Coup Détat Events, 1946–2013." Center for Systemic Peace.

Mendoza, Enrique G, and Vivian Z Yue. 2012. "A General Equilibrium Model of Sovereign Default and Business Cycles." *Quarterly Journal of Economics* 127 (2): 889–946.

Morais, Bernardo, and Mark L. J. Wright. 2008. "International Financial Crisis Facts." Working Paper. Department of Economics, University of California at Los Angeles.

Na, Seunghoon, Stephanie Schmitt-Grohé, Martín Uribe, and Vivian Yue. 2018. "The Twin Ds: Optimal Default and Devaluation." *American Economic Review* 108 (7): 1773–1819.

Owyang, Michael T, Valerie A Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater During Periods of Slack? Evidence from Twentieth-Century Historical Data." *American Economic Review* 103 (3): 129–134.

Perez, Diego. 2015. "Sovereign Debt, Domestic Banks and the Provision of Public Liquidity." SIEPR Discussion paper 15-016.

Ramey, Valerie A, and Sarah Zubairy. 2014. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." NBER Working Paper 20719.

Ravn, Morten O, and Harald Uhlig. 2002. "on Adjusting the Hodrick-Prescott Filter for the Frequency of Observations." *Review of Economics and Statistics* 84 (2): 371–376.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011a. "The Forgotten History of Domestic Debt." *Economic Journal* 121 (551): 319–350.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011b. "From Financial Crash to Debt Crisis." *American Economic Review* 101 (5): 1676–1706.

Reinhart, Carmen M., and Christoph Trebesch. 2016. "The International Monetary Fund: 70 Years of Reinvention." *Journal of Economic Perspectives* 30 (1): 3–28.

Reinhart, Carmen M, and Kenneth S Rogoff. 2011. "From Financial Crash to Debt Crisis." *American Economic Review* 101 (5): 1676–1706.

Rose, Andrew K. 2005. "One Reason Countries Pay Their Debts: Renegotiation and International Trade." *Journal of Development Economics* 77 (1): 189–206.

Schularick, Moritz, and Alan M. Taylor. 2012. "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review* 102 (2): 1029.

Sosa-Padilla, Cesar. 2018. "Sovereign Defaults and Banking Crises." *Journal of Monetary Economics*,

Tomz, Michael, and Mark L. J. Wright. 2007. "Do Countries Default in "Bad Times"?" *Journal of the European Economic Association* 5 (2-3): 352–360.

Trebesch, Christoph, and Michael Zabel. 2017. "The Output Costs of Hard and Soft Sovereign Default." *European Economic Review* 92: 416–432.

Yue, Vivian Z. 2010. "Sovereign Default and Debt Renegotiation." *Journal of International Economics* 80 (2): 176–187.

Chapter 5

Deleveraging, Deflation and Depreciation in the Euro Area*

Joint with Gernot J. Müller, and Martin Wolf

5.1 Introduction

Following the onset of the global financial crisis, the euro area has experienced a protracted economic slump, with aggregate output in 2015 still below the 2008 level. The dynamics of the slump among individual countries, however, have been highly heterogenous. Whilst all countries underwent a deep recession in 2009, some rebounded quickly whereas others have experienced a persistent and protracted further decline. These most adversely affected economies are also undergoing a severe deleveraging process (Martin and Philippon, 2014; Reinhart and Rogoff, 2014).[1] Meanwhile, inflation in the euro area has been subdued, both during the initial downturn and the subsequent years. Moreover this subdued inflation has taken hold throughout the union, both in countries experiencing a deep slump and those doing relatively well. As a consequence, intra-euro-area real exchange rates have hardly moved during the post-crisis period.

In this paper we ask why—despite the heterogenous deleveraging and output performance across the euro area—there has been no significant adjustment of intra-euro-area real exchange rates. To provide an answer, we put forward a stylised two-country model of a currency union which accounts for the key features of the

* Kuvshinov: University of Bonn, Email: dmitry.kuvshinov@uni-bonn.de. Müller: University of Tübingen and CEPR, London, Email: gernot.mueller@uni-tuebingen.de, Wolf: University of Vienna, Email: martin.wolf@uni-bonn.de. We thank our discussant Tommaso Monacelli and participants in the conference on “The Post-Crisis Slump” at the EU commission in October 2015 for very useful comments. The usual disclaimer applies.

1. For an early account of how deleveraging in the global banking sector helped the financial crisis in the United States to morph into a global crisis, see Kollmann, Enders, and Müller (2011).

crisis in the euro area. The two countries represent the two heterogenous groups in the euro area, namely the "stressed" economies undergoing a severe slump and the "non-stressed" economies which have performed relatively well. Each country specialises in production of differentiated goods. The real exchange rate fluctuates with the relative price of these goods because goods markets are imperfectly integrated. Whilst goods prices are flexible in both economies, wages are downwardly rigid as in Schmitt-Grohé and Uribe (2016). Monetary policy aims to stabilise union-wide inflation, but may be constrained by the zero lower bound on nominal interest rates.

Drawing on Eggertsson and Krugman (2012), we assume that a group of households in the stressed economy are forced to reduce their debt because their borrowing limit tightens exogenously. We study how the repercussions of this deleveraging shock transmit through the entire currency union. The model is able to account for key features of the data. First, the output performance is heterogenous. Output collapses in the stressed economy, but is hardly affected in the rest of the union. Second, there is—at the same time—union-wide deflationary pressure such that, third, the real exchange rate hardly moves. Our analysis reveals that size matters for these results: both the size of the country which is subject to the deleveraging shock and the size of the shock itself.

To establish this analytically, we first consider the case when the stressed economy is generically small. In this case, deleveraging does not impact the rest of the union at all and depreciates the stressed economy's real exchange rate. In line with conventional wisdom, we find that the extent of depreciation is limited by the extent of downward wage rigidity in the stressed economy (Friedman, 1953).[2] By the same token, the recession turns out to be less severe, the more flexible wages are.

Once we turn to the other polar case and assume that the stressed economy is large, the effects of the shock turn out to be fundamentally different. This holds in particular if the shock, in addition to the economy, is large as well. In this case, monetary policy finds itself pushed to the zero lower bound, and hence unable to contain the deflationary effects of the shock—not only in the stressed economy, but in the entire union. Whilst relative wage rigidities in the two countries still play a role in the adjustment process, deflationary spillovers from the stressed to the non-stressed economy will generally dampen the extent of real depreciation. Under specific conditions the real exchange rate may even appreciate, echoing earlier findings of a "perverse" response of real exchange rates at the zero lower bound when exchange rates are flexible (Cook and Devereux, 2013).

Furthermore in this scenario—somewhat paradoxically—the real exchange rate may depreciate less, the more flexible wages in the stressed economy are. The reason for this can be traced back to the root cause of the crisis itself—debt, or more

2. Kollmann (2001) and Monacelli (2004) perform analyses of how nominal rigidities impact real exchange rate volatility under flexible and fixed exchange rates in small open economies. Broda (2004) provides evidence for developing countries in support of the received wisdom.

precisely debt deflation à la Fisher (1933). As prices in the domestic economy decline, the real value of debt increases, thwarting the initial efforts to reduce the debt. Increased wage flexibility gives rise to more debt deflation, amplifying the recession and dampening the real exchange rate response further—an instance of the "paradox of flexibility", as established by Eggertsson and Krugman (2012) for the closed economy. We offer an important qualification to this paradox in the currency union setting, however: it only applies if both the shock and the size of the domestic economy are large. In particular, if the domestic economy is small, the drop in domestic prices depreciates the real exchange rate, stabilising the economy. Moreover in this case, long-run purchasing power parity implies that the decline in domestic prices is met by future inflation, which is also stabilising (Corsetti, Kuester, and Müller, 2013).

What remains to be determined is whether the size of stressed economies in the euro area, and the magnitude of deleveraging are sufficiently large to push the union to the zero lower bound and generate enough deflationary spillovers for the real exchange rate response to be muted. In our quantitative analysis, we find this to be the case for plausible parametrisations of the model.[3] As monetary policy becomes constrained by the zero lower bound, domestic output and prices decline, as do prices in the entire monetary union. Importantly, the decline in prices in the foreign economy implies that foreign output remains close to full employment. The model thus generates a heterogenous output response across the two countries. At the same time, deflationary spillovers ensure that the real exchange rate remains basically flat. Counterfactual simulations suggest that greater wage flexibility in the stressed economies is unlikely to be stabilising, and may instead deepen the recession.

The effects of debt deleveraging in an open economy context have been analysed in a number of other studies. Benigno and Romei (2014) examine the implications of deleveraging by one country within the world economy and study how monetary policy should be optimally set at a global level. However they do not consider the case of a monetary union. Fornaro (2015) studies the implications of deleveraging within a part of a monetary union under the zero lower bound constraint, as in our paper. But he abstracts from internal debt and debt deflation within countries, and does not focus his attention on relative price movements across stressed and non-stressed economies, which is the main focus of our paper. Gilchrist, Schoenle, Sim, and Zakrajsek (2015) also study the lack of real exchange rate adjustment across the euro area, but focus on financial constraints of firms rather than households. In their account, adverse financing conditions induce firms in stressed economies to keep prices high relative to what would be optimal under benign conditions. Firms in non-stressed economies, in turn, find it optimal to reduce prices in order to cannibalise the market share of stressed firms.

3. Stressed economies make up 37% of the union and deleveraging is 34% of annual GDP per borrower, based on eurozone data.

The remainder of the paper is organised as follows. The next section provides a number of basic facts regarding the post-2008 dynamics in the euro area. Section 5.3 introduces the model. We discuss analytical results for the limiting cases in Section 5.4. Section 5 presents results obtained from model simulations. A final section concludes.

5.2 Some facts

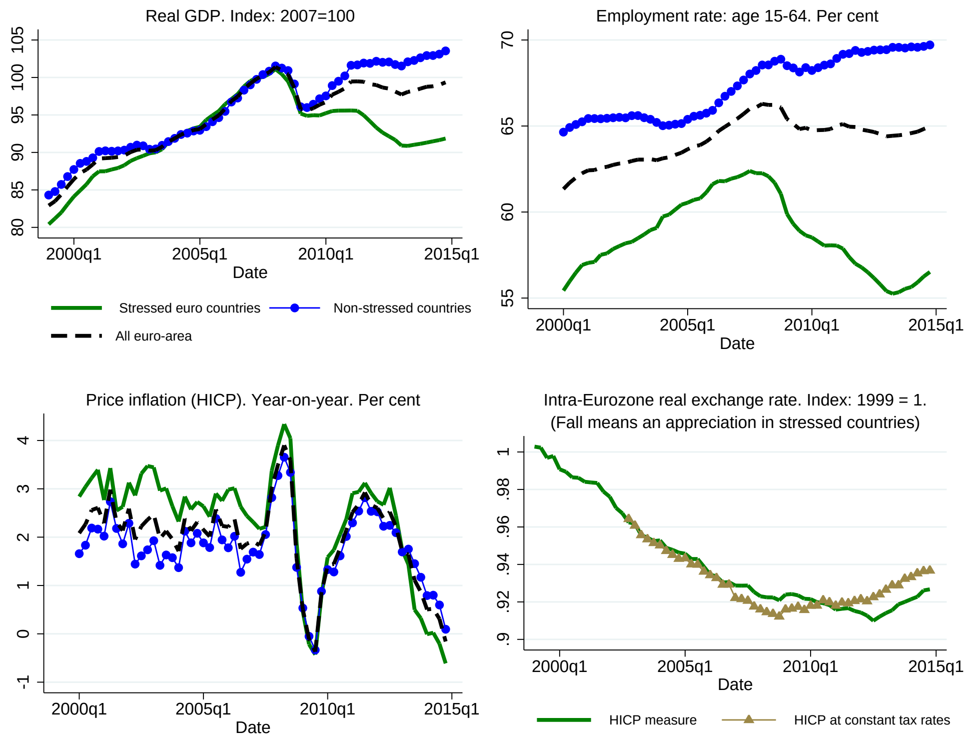
In this section we present time series evidence that highlights important aspects of the post-crisis slump in the euro area and provides some background for our model-based analysis. Our focus is on two regions, each consisting of a group of countries. The "stressed" region comprises countries where the crisis was particularly severe: Greece, Italy, Ireland, Portugal and Spain. The "non-stressed" region consists of Austria, Belgium, Finland, France, Germany and the Netherlands. For each group we aggregate time series using 2007 GDP weights. The stressed economies make up roughly 37% of GDP of all the countries in our sample.[4] Our sample, in turn, covers about 97% of euro-area GDP.

Figure 5.2.1 displays quarterly time series data for macroeconomic aggregates, covering the period 1999Q1–2014Q4. Dashed lines correspond to data for the euro area as a whole, solid lines correspond to the stressed economies and dotted lines—to the non-stressed economies. The upper-left panel shows real GDP normalised to 100 in the pre-crisis year 2007. In the run-up to the crisis output growth was quite synchronised across the two regions. Even during the early stages of the crisis, the GDP collapse in stressed and non-stressed countries was roughly the same. But since 2010 the growth performance has been quite distinct. The decline in GDP in the stressed economies seemed to bottom out in 2011–2012, only to decline further afterwards. While there is a small recovery at the end of our sample, stressed-economy GDP is still almost 10 percent below its pre-crisis level. In contrast, GDP recovered relatively quickly in the non-stressed countries, surpassing the pre-crisis level in early 2011. A similar picture emerges for employment data (top-right panel).

The intra-euro-area real exchange rate is displayed in the bottom-right panel of Figure 5.2.1. The solid line indicates the exchange rate measure based on the harmonised index of consumer prices. The line with markers (triangles) corresponds to a series which controls for tax changes. In both instances, we normalise the exchange rate to unity in 1999 and define it such that a decline corresponds to an appreciation for the stressed economies. In the years prior to the crisis the real exchange rate appreciated by about 8 percent, but has moved little after 2008. The HICP-based measure indicates that there was a further, if very mild, appreciation in

4. The 37% figure corresponds to pre-crisis levels. During the crisis, the share falls to around 34%.

Figure 5.2.1. Development of macroeconomic aggregates

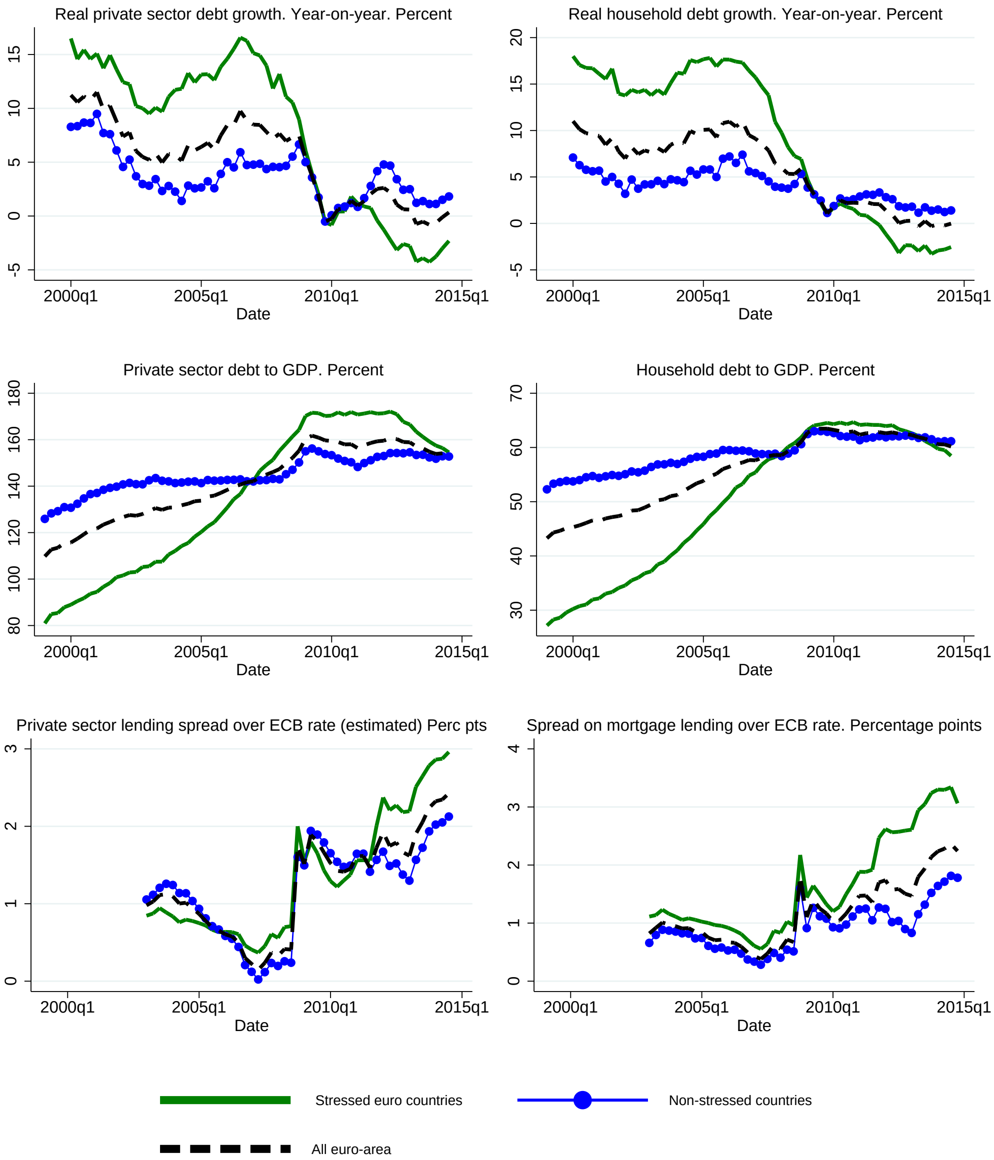


Notes. Source: Eurostat. Averages are weighted by 2007 nominal GDP. Irish and Finnish tax-adjusted exchange rates series start in 2013 and 2005 respectively; unadjusted data used beforehand.

the early stage of the crisis and an equally mild depreciation after 2013. The relative price adjustment is somewhat more pronounced but still muted for the series which is purged of the effect of tax changes. Within the euro area, changes in the real exchange rate are the result of differential inflation developments across the two regions, which are shown in the lower left panel. Prior to the crisis inflation in the stressed economies always exceeded inflation in the non-stressed region. Since the start of the crisis, however, inflation dynamics across the two regions have been markedly similar: inflation in both regions briefly turned negative in 2009, recovered afterwards but declined again after 2012.

In the model-based analysis that follows we explore the adjustment of the real exchange rate to a deleveraging process which takes place in one region of the currency union. That such a process contributed to the post-crisis slump has been suggested by many observers (see, for instance, Martin and Philippon, 2014; Reinhart and Rogoff, 2014) and is hardly controversial in light of the facts. The upper panels of Figure 5.2.2 display real credit growth (top row) and credit volumes relative to GDP (middle row) in the euro area, again distinguishing between area-wide devel-

Figure 5.2.2. Development of debt and spreads



Notes. Sources: Eurostat, ECB and BIS. Averages are weighted by 2007 nominal GDP. Irish household debt data estimated before 2001Q4.

opments and those in the stressed and non-stressed economies. In the left and right panels we show private and household debt respectively.

We observe that prior to the crisis the stressed economies experienced a particularly rapid expansion of credit. In fact their real credit growth averaged close to 13% per year, roughly three times the corresponding rate in the non-stressed economies. The advent of the crisis in 2008–09 coincided with a collapse in lending growth. Credit growth in both regions was close to zero in 2009 and then turned negative in the stressed economies whilst recovering somewhat in the non-stressed economies. In the former, it is still negative at the end of our sample. Overall, credit volumes thus declined considerably during the crisis—as far as the stressed economies are concerned. Expressed relative to GDP, overall private credit (household debt) peaked at some 170 (65) percent of GDP in the stressed economies in 2009. Since then the decline in credit relative to GDP has been muted by the sizeable decline in GDP.

In the bottom panels of Figure 5.2.2 we show the dynamics of interest rates. Specifically, we compute the difference of private-sector lending (left panel) and mortgage (right panel) rates relative to the ECB main refinancing rate. The spread on mortgages is higher in the stressed economies throughout the sample period. There is, however, a marked widening in the gap between stressed and non-stressed economies after 2010, reaching some 2 percentage points. The picture is less clear-cut for the spread on private sector lending rates. However, private sector spreads also tightened by considerably more in the stressed economies after 2009.

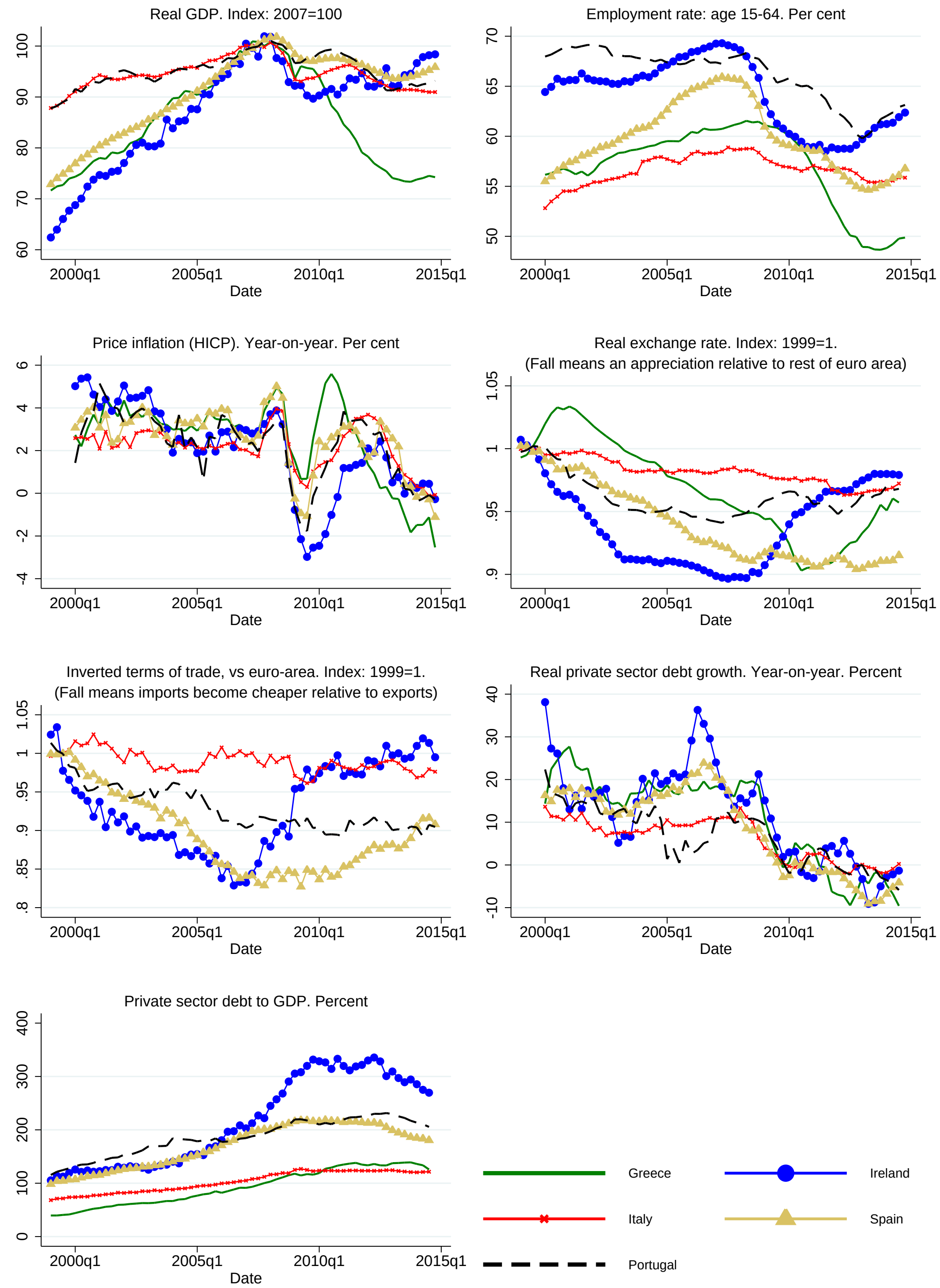
Figure 5.2.3 zooms in on the developments of individual countries of the stressed-economies aggregate. As we might expect, the disaggregated picture is more nuanced, with somewhat different dynamics from country to country. Ireland stands out with the largest lending boom, and also the most significant real exchange rate adjustment. Italy, on the contrary, did not experience much of a boom in lending or a significant real exchange rate appreciation before the crisis. Greece, in turn, shows the most dismal performance in terms of post-crisis output and employment.

Despite these differences, we find sufficiently strong similarities across the stressed euro area countries for our stressed/non-stressed country classification to be justified. First, the top panels of Figure 5.2.3 (real GDP and employment) show that all countries are experiencing a prolonged slump and are yet to fully recover. Second, inflation dynamics are quite synchronised (second row, left panel). Third, in all countries real exchange rates are still appreciated relative to the beginning of the sample (second row, right panel). For our sample, we also observe that the movements in real exchange rates reflect to a large extent movements of the terms of trade (displayed in the left panel of the third row).[5] This observation squares well with formal results of earlier studies which decompose real exchange rate movements (e.g., Engel, 1999).[6] Fourth and finally, note that all countries experienced a lending surge before the crisis and a lending slowdown after 2009, albeit to different degrees.

5. For ease of comparison, the graph depicts "inverted" terms of trade: the price of imports relative to exports. That way it is directly comparable to our real exchange rate series. No observations are available for Greece.

6. However, in a recent study Berka, Devereux, and Engel (2015) find that for a pre-crisis sample of euro area countries, cross-country productivity differentials in the traded and non-traded good sector matter a great deal for real exchange rate movements, even in the short run. According to the authors this result is likely due to the fact that the countries in their sample maintain a common currency.

Figure 5.2.3. Developments in individual countries of the stressed region



Notes. Sources: Eurostat, BIS and OECD Economic Outlook.

In sum, a simple inspection of the facts supports the view that the stressed economies of the euro area are experiencing a fully-fledged balance sheet recession. A sizeable build-up of debt was followed by a lengthy period of deleveraging, with very adverse consequences for economic activity. On the contrary, there is a recovery in the non-stressed region. Against this background it may be surprising that inflation is subdued not only in the stressed economies, but in the non-stressed region, too. Equivalently, the lack of real deprecation in the stressed economies relative to the non-stressed economies may appear puzzling. We investigate this issue further by means of a model-based analysis.

5.3 The model

Our analysis is based on a simple two-country model of a currency union. Countries specialise in the production of specific goods which are traded across countries. Good market integration is incomplete, however, as countries' consumption is biased towards domestically produced goods. The real exchange rate may therefore deviate from unity. Given that the real exchange rate and the terms of trade co-move strongly in our sample (see Section 5.2), we abstract from the production of non-traded goods. Countries may differ in size and we assume "Home" makes up a mass $[0, n)$ of the total union population, where $n \in [0, 1]$. Each country is populated by a unit mass of agents, who supply labour inelastically to domestic firms. Within Home we distinguish between households with high and low discount factors as in Eggertsson and Krugman (2012). We refer to these households as "savers" and "borrowers" respectively. The rest of the union ("Foreign") is populated by savers only. Savers in one country can trade a nominally non-contingent bond with the savers in the other country. Savers in Home can, in addition, lend funds to domestic borrowers. Prices are fully flexible, but downward adjustment of nominal wages is restricted as in Schmitt-Grohé and Uribe (2016).

5.3.1 Households

In Home, borrowers account for a fraction $\chi \in (0, 1)$ of the population. They are less patient than savers, which account for the rest. A typical saver in Home maximises

$$\max_{\{C_t^s\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} (\beta^s)^t \ln(C_t^s)$$

subject to

$$P_t C_t^s + R_t^{-1} B_t^s + R_t^{*-1} D_t \ = \ W_t L_t + B_{t-1}^s + D_{t-1} \tag{5.3.1}$$

and a constraint which rules out Ponzi games. Here $\beta^s < 1$ is the discount factor which exceeds the discount factor of the borrower: $\beta^s > \beta^b$. C_t^s denotes savers' per capita consumption and P_t is the consumer price level in Home. Savers earn rate R_t

by lending to borrowers at home (B_t^s), and rate R_t^* by saving abroad (D_t). All debt and interest rates are denominated in nominal terms. W_t is the nominal wage rate in Home, and L_t are hours worked. Optimality requires the following Euler equation to hold

$$\left(C_t^s\right)^{-1} = \beta^s R_t \left(C_{t+1}^s\right)^{-1} \frac{P_t}{P_{t+1}}, \tag{5.3.2}$$

as well as a transversality condition. The absence of arbitrage possibilities between domestic and foreign assets requires that

$$R_t = R_t^*. \tag{5.3.3}$$

In turn, a typical borrower maximises

$$\max_{\{C_t^b\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} (\beta^b)^t \ln(C_t^b)$$

subject to

$$P_t C_t^b + B_{t-1}^b = R_t^{-1} B_t^b + W_t L_t \tag{5.3.4}$$

$$B_t^b \leq \bar{B}_t. \tag{5.3.5}$$

Here, B_t^b denotes nominal debt vis-à-vis the savers which may not exceed an exogenous, potentially time-varying, debt limit $\bar{B}_t$. First order conditions imply that (5.3.5) holds with equality at all times.[7] Aggregate consumption in Home is given by

$$C_t = (1-\chi) C_t^s + \chi C_t^b. \tag{5.3.6}$$

In the rest of the union, all households are savers and we suppress the superscript s for simplicity. The objective is

$$\max_{\{C_t^*\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} (\beta^s)^t \ln(C_t^*)$$

subject to

$$P_t^* C_t^* + R_t^{*-1} D_t^* = W_t^* L_t^* + D_{t-1}^*, \tag{5.3.7}$$

where stars denote variables in the rest of the union. First order conditions imply

$$\left(C_t^*\right)^{-1} = \beta^s R_t^* \left(C_{t+1}^*\right)^{-1} \frac{P_t^*}{P_{t+1}^*}. \tag{5.3.8}$$

7. More precisely, optimality requires that $\left(C_t^b\right)^{-1} \geq \beta^b R_t \left(C_{t+1}^b\right)^{-1} \frac{P_t}{P_{t+1}}$, holding with equality whenever $B_t^b < \bar{B}_t$, and requiring $B_t^b = \bar{B}_t$ whenever holding with strict inequality. In the steady state of the model as well as during the deleveraging phase, the latter case always obtains, such that we omit this case distinction from the main text.

We allow for home bias in consumption in both countries, and the elasticity of substitution between domestic and imported goods is unity. Specifically, aggregate consumption is a composite of two goods in both countries:

$$C_t = \frac{C_{H,t}^{\lambda} C_{F,t}^{1-\lambda}}{\lambda^{\lambda}(1-\lambda)^{1-\lambda}}, \qquad C_t^* = \frac{C_{H,t}^{\lambda^*} C_{F,t}^{1-\lambda^*}}{\lambda^{*\lambda^*}(1-\lambda^*)^{1-\lambda^*}},$$

where $\lambda = 1-(1-n)\omega$ and $\lambda^* = n\omega$. Here, $C_{H,t}$ is the locally produced good, $C_{F,t}$ is the good produced in the rest of the union, and $\omega \in [0,1]$ determines the degree of home bias in consumption, which we assume is symmetric across countries. If $\omega = 1$ there is no home bias. This formulation of relative consumption weights follows Sutherland (2005) and De Paoli (2009). Note that both country size n and home bias $1-\omega$ affect the consumption shares of Home- and Foreign-produced goods; however they do so in different ways. As the size of the domestic country grows (as n increases), both domestic and foreign households consume a bigger share of the good produced in Home. By contrast, an increase in $1-\omega$ implies that both domestic and foreign households consume a bigger share of goods produced in their own country (Home and Foreign respectively)—the classic notion of "home bias".[8]

The local good sells at price $P_{H,t}$, while the foreign good sells at price $P_{F,t}$. Expenditure minimisation implies

$$P_t = P_{H,t}^{\lambda} P_{F,t}^{1-\lambda}, \qquad P_t^* = P_{H,t}^{\lambda^*} P_{F,t}^{1-\lambda^*}, \tag{5.3.9}$$

that is, the consumer price indices domestically and abroad are a weighted average of the producer prices of the two goods.

Furthermore, we define the real exchange rate Q_t as the price of foreign consumption in terms of domestic consumption,

$$Q_t = \frac{P_t^*}{P_t}, \tag{5.3.10}$$

such that an increase in Q_t indicates a depreciation of Home's real exchange rate.

5.3.2 Firms

Firms operate in competitive goods and labour markets. They maximise profits $P_{H,t}Y_t - W_t L_t$ in Home, $P_{F,t}Y_t^* - W_t^* L_t^*$ in Foreign, subject to

$$Y_t = L_t, \qquad Y_t^* = L_t^* \tag{5.3.11}$$

8. This in turn implies that the real exchange rate is independent of the distribution of wealth *only if* home bias is zero ($\omega = 1$), regardless of the value of n. A formal analysis of this issue is available on request. Note that furthermore, the size of the economy n impinges on the supply side of the model, as more goods are being produced in a country that is larger (formally, this is implied from equation (5.3.21)).

respectively, and their first order conditions imply

$$P_{H,t} = W_t, \qquad P_{F,t} = W_t^*. \tag{5.3.12}$$

As in Schmitt-Grohé and Uribe (2016), the labour market is characterised by downward nominal wage rigidity. In each period, a maximum of $\bar{L}$ hours can be sold to firms

$$L_t \leq \bar{L}, \qquad L_t^* \leq \bar{L} \tag{5.3.13}$$

while wages may fall by at most $(1-\gamma)$ in Home, $(1-\gamma^*)$ in Foreign, in proportion to their previous level

$$W_t \geq \gamma W_{t-1}, \qquad W_t^* \geq \gamma^* W_{t-1}^*. \tag{5.3.14}$$

We require that $1 \geq \gamma > 0$ and $1 \geq \gamma^* > 0$, where $\gamma, \gamma^* \to 0$ characterises flexible wages, and $\gamma, \gamma^* = 1$—full rigidity. The labour markets are closed by complementary slackness conditions of the form

$$(L_t - \bar{L})(W_t - \gamma W_{t-1}) = 0, \qquad (L_t^* - \bar{L})(W_t^* - \gamma^* W_{t-1}^*) = 0, \tag{5.3.15}$$

which imply that, as long as wages are free to adjust, the economy must operate at potential. Conversely, involuntary unemployment is possible as (5.3.14) becomes a binding constraint.

5.3.3 Monetary policy

We assume that monetary policy is characterised by a strict inflation targeting rule, adjusting the nominal interest rate such that area-wide inflation is zero, subject to a zero lower bound constraint. It targets

$$\Pi_t^u = 1 \text{ subject to } R_t \geq 1, \tag{5.3.16}$$

where $\Pi_t^u = (P_t)^n (P_t^*)^{1-n} / (P_{t-1})^n (P_{t-1}^*)^{1-n}$ is area-wide inflation, and sets

$$R_t = 1 \tag{5.3.17}$$

if due to deflationary pressure, the inflation target cannot be reached.

5.3.4 Market clearing

Goods market clearing requires that the supply of domestically produced goods equals domestic as well as export demand

$$Y_t = \left(\frac{P_{H,t}}{P_t}\right)^{-1} \left(\lambda C_t + \frac{\lambda^*(1-n)}{n} Q_t C_t^*\right). \tag{5.3.18}$$

Equivalently, we require for the Foreign-produced good[9]

$$Y_t^* = \left(\frac{P_{F,t}}{P_t^*}\right)^{-1}\left(\frac{(1-\lambda)n}{1-n}Q_t^{-1}C_t + (1-\lambda^*)C_t^*\right). \tag{5.3.19}$$

Moreover, asset market clearing requires

$$(1-\chi)B_t^s = \chi B_t^b \tag{5.3.20}$$

within Home and

$$(1-\chi)nD_t + (1-n)D_t^* = 0 \tag{5.3.21}$$

across the two countries.

An equilibrium is a sequence of endogenous variables $\{Y_t, Y_t^*, L_t, L_t^*, C_t, C_t^*, C_t^s, C_t^b, B_t^s, B_t^b, \ldots \quad D_t, D_t^*, R_t, R_t^*, P_t, P_t^*, P_{H,t}, P_{F,t}, W_t, W_t^*, Q_t, \Pi_t^u\}$ solving equations (5.3.1)—(5.3.21), for given parameters and initial conditions, and exogenous $\{\bar{B}_t\}$.

5.3.5 Steady state

We assume that initially the economy is in a symmetric steady state: the real exchange rate, consumer and producer price indices are equal to unity, $P_H = P = P^* = P_F = 1$, which from (5.3.12) implies that $W = W^* = 1$. Moreover, we let $Y = Y^* = \bar{L}$ and $C^* = Y^*$. This implies $C = Y$ from equations (5.3.18) and (5.3.19). We obtain $R = R^* = 1/\beta^s$ from equations (5.3.2) and (5.3.8). Borrowers are up against the borrowing constraint, hence $C^b = Y - (1-\beta^s)\bar{B}$. Savers in Home consume $C^s = Y + (1-\beta^s)(B^s + D)$, and savers in Foreign $C^* = Y^* + (1-\beta^s)D^*$. This combined with the fact that $C^* = Y^*$ yields $D^* = 0$, and from (5.3.21) $D = 0$. That is, net foreign assets must equal zero in the initial steady state. Note that the economy is characterised by non-stationary dynamics, that is, it will generally not revert back to its initial steady state, once it departs from it.[10]

5.4 Relative prices in a crisis

We now investigate how the economy adjusts to a deleveraging shock. More specifically, we consider a one-off tightening of the debt limit in Home from $\bar{B}_t = \bar{B}^H$ to a

9. Note that for the limiting cases we discuss in Section 5.4, $\frac{\lambda^*(1-n)}{n} \to \omega$ as $n \to 0$ and $\frac{(1-\lambda)n}{1-n} \to \omega$ as $n \to 1$.

10. The economy is non-stationary for two reasons. First, international financial markets are incomplete, and second, households are heterogenous. The distribution of wealth, both across agents and across countries is a state of the economy which induces unit-root behaviour in some variables.

permanently lower level $\bar{B}^L$ in time period t. Our setup mimics Eggertsson and Krugman (2012), except that we consider a two-country model and restrict the tightening of the debt limit to take place in one country only. The adjustment will then depend on the size of this country. We illustrate this by first focusing on two limiting cases: $n \to 0$ and $n \to 1$. For these cases we obtain closed-form results, and develop an intuition of the underlying mechanisms. We discuss numerical results for intermediate n in Section 5.5. Throughout, our main interest is how the real exchange rate in period t responds to the shock.

5.4.1 Deleveraging in a small union member

If $n \to 0$, Home is effectively a small open economy (see Galí and Monacelli, 2005; De Paoli, 2009). The Home-good consumption weights are $\lambda \to 1-\omega$ and $\lambda^* \to 0$ in Home and Foreign, respectively. In this case, $P_t^* = P_{F,t}$ from equation (5.3.9), $Y_t^* = C_t^*$ from (5.3.19) and $D_t^* = 0$ from (5.3.21). In other words, the rest of the union resembles a closed economy and is not affected by the deleveraging shock in Home. It therefore remains in the initial steady state during the whole deleveraging process. This has important implications for monetary policy: from (5.3.16) we obtain $\Pi_t^u = P_t^*/P_{t-1}^* = 1$, such that average inflation in the union is zero, and as a result the nominal interest rate remains unchanged as well: $R_t^{-1} = \beta^s$. We summarise our main result in what follows.

Proposition 1. *Consider the economy defined in Section 5.3 and let $n \to 0$. Suppose that in period t, the debt limit in Home is unexpectedly and permanently reduced from $\bar{B}^H$ to $\bar{B}^L$. The real exchange rate at time t is then given by*

$$Q_t = P_t^*/P_t = \min\left(\gamma^{\omega-1}, [1-\eta(\bar{B}^H - \bar{B}^L)/Y]^{\omega-1}\right) \geq 1, \tag{5.4.1}$$

where $\eta = (\beta^s(1-\omega)\chi)/(1-(1-\omega)\chi) > 0$. Therefore, there is no depreciation ($Q_t = 1$) if wages are completely rigid ($\gamma = 1$), and a greater depreciation, the more flexible wages are, where the upper threshold $[1-\eta(\bar{B}^H - \bar{B}^L)/Y]^{\omega-1} > 1$ is reached once (5.3.14) ceases to bind.

Proof. See Appendix. □

Intuitively, the deleveraging shock forces borrowers to cut consumption in order to repay their debts. This reduces aggregate demand and puts downward pressure on prices, resulting in a real exchange rate depreciation, as long as wages are allowed to adjust sufficiently. We now establish a second result.

Proposition 2. *Consider again the economy defined in Section 5.3 with $n \to 0$. In the period of deleveraging, output, saver consumption, borrower consumption and real wage income all decline strictly less if wages are more flexible (that is, if γ is reduced) up until (5.3.14) ceases to bind, point beyond which they do not vary further. The recession is deepest if wages are completely downwardly rigid ($\gamma = 1$).*

Proof. See Appendix. □

Propositions 1 and 2 establish that wage flexibility and the associated movements in the real exchange rate dampen the response to country-specific shocks, as the received wisdom—going back to at least Friedman (1953)—suggests. Still, it is interesting to analyse this case for the following reasons. First, it will serve as a useful benchmark once we consider $n > 0$. Second, the fact that a real depreciation plays a stabilising role is actually not obvious during a deleveraging recession. The fall in Home prices required to bring about the depreciation increases the real value of debt, giving rise to debt deflation à la Fisher (1933).

To see this, consider how the borrowers respond to the deleveraging shock. When the shock hits, nominal debt of $\bar{B}^H - \beta^s \bar{B}^L$ has to be repaid to satisfy the new, lower, borrowing limit. By rearranging the budget constraint (5.3.4) as follows

$$C_t^b \;=\; -\frac{\bar{B}^H - \beta^s \bar{B}^L}{P_t} + \frac{W_t L_t}{P_t}, \tag{5.4.2}$$

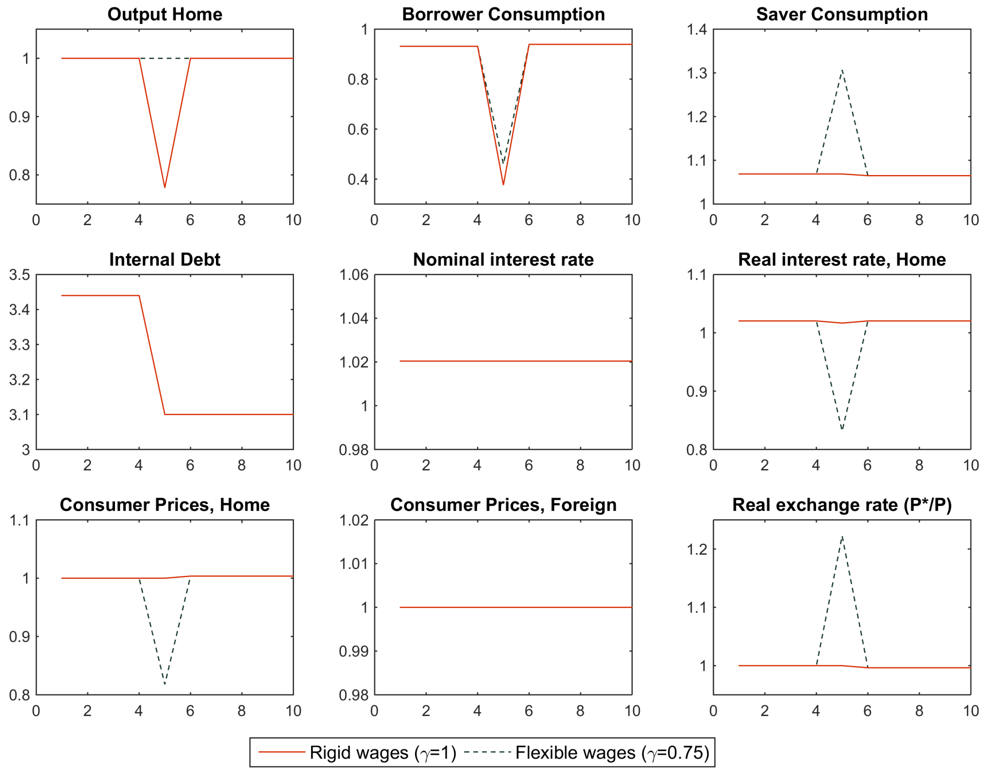
we see that borrowers' consumption depends on debt repayment as well as on real wage incomes $W_t L_t / P_t$. For a given real income, a lower price level increases real debt and reduces consumption. Still, recall that borrower consumption declines less in general equilibrium with greater wage flexibility and real depreciation (Proposition 2). The reason is that in this case, overall economic activity is higher, which helps sustain the real wage income of the borrowers. First, a weaker real exchange rate crowds in foreign demand for the domestic good. Second, since long-run prices are pinned down by purchasing power parity, a temporary drop in the price level generates expected inflation, which reduces the real interest rate and increases spending by savers.[11] It turns out that these two mitigating factors necessarily outweigh the adverse impact of debt deflation. In other words, while rigidity in wages may rule out debt deflation altogether, the resulting drop in real wage income (which operates via a drop in working hours L_t) depresses borrower consumption all the same—in fact, depresses it by more, the more rigid the wages.

In sum, in the case of a small open economy, a lack of relative price adjustment reflects the presence of nominal rigidities, in our case downwardly sticky wages. We show the adjustment dynamics in Figure 5.4.1, which contrasts results for the cases of fully rigid ($\gamma = 1$, solid lines) and flexible ($\gamma = 0.75$, dashed lines) wages. We discuss the parameter choices which underlie the model simulations in Section 5.5 below. Importantly, we assume that a deleveraging of 17% GDP is undertaken over one year, in period 5.[12] Because the share of borrowers is 0.5, this amounts to a

11. Thus, implicit in fixed exchange rate regimes is an element of price level targeting, which has been emphasised in previous work (Corsetti, Kuester, and Müller, 2013). A more detailed discussion is provided in Section 5.4.2.

12. Here and in the rest of the paper, we refer to deleveraging in terms of nominal GDP before the crisis, which equals 1 in our parametrisation.

Figure 5.4.1. Deleveraging in a small member of a currency union ($n \longrightarrow 0$)



Notes. Parameter values: $\bar{L} = 1, \bar{B}^H = 3.44, \bar{B}^L = 3.1, \chi = 0.5, \omega = 0.2, \beta^s = 0.98, \beta^b = 0.97$.

deleveraging of 34% GDP per borrower. Figure 5.4.1 echoes our discussion above: if wages are sticky, the exchange rate response is flat, while output collapses. If wages are fully flexible, the exchange rate depreciates strongly, while output remains constant.[13]

The response of consumption also differs across the two scenarios: both saver and borrower consumption are higher under flexible wages, relative to the rigid-wage scenario. This implies, in the context of our model, that welfare is higher under more flexible wages. Galí and Monacelli (2013), in contrast, find that higher wage flexibility may reduce welfare whenever monetary policy seeks to stabilise the exchange rate. This is because Galí and Monacelli (2013) assume a monopolistically competitive labour market and staggered wage setting. As a result, higher wage inflation induces wage dispersion which is detrimental to welfare. We do not consider this possibility. Moreover, recall that in our set-up labour effort has no direct bearing on household utility.

13. In our calibration, $\gamma = 0.75$ provides sufficient flexibility for (5.3.14) not to bind in the period of deleveraging. A further increase in wage flexibility would then leave results unaltered.

Martin and Philippon (2014) also study a deleveraging shock within a small member country of a currency union. They find their model to perform well in accounting for the dynamics of nominal GDP, employment and net exports in a number of euro area countries during the period 2000–2012. Through counterfactual simulations they explore the role of macro-prudential and fiscal policy in influencing macroeconomic outcomes. Their analysis does not consider movements in relative prices, and in particular real exchange rates, which is the main focus of our paper. Furthermore, focussing on shocks to a small union member—as we do in this section—implies that a deleveraging shock does not generate spillovers on the rest of the union. As we show next, allowing for such spillovers has important implications for the area-wide response to the shock and, hence, for the adjustment of the real exchange rate.

5.4.2 Deleveraging in a (very) large union member

If $n \to 1$, the Home-good consumption weights are $\lambda \to 1$ and $\lambda^* \to \omega$ in Home and Foreign respectively. We obtain $Y_t = C_t$ from equation (5.3.18), $P_t = P_{H,t}$ from (5.3.9) and $D_t = 0$ from (5.3.21). Thus, Home accounts for almost the entire currency union and behaves like a closed economy. Foreign, in turn, effectively becomes a small open economy. Again, this has important implications for monetary policy: it is entirely geared towards developments in Home, as Foreign has a negligible effect on average union-wide inflation: $\Pi_t^u = P_t/P_{t-1}$ from (5.3.16).

As we now show, in this case the real exchange rate response hinges critically on the size of the shock, and the related monetary policy response. We establish that for a large shock, monetary policy becomes constrained by the zero lower bound (5.3.17) and, as a result, the real exchange rate response becomes muted, provided wages in Home are not much more flexible than in Foreign. Moreover, under certain conditions the real exchange rate may in fact appreciate rather than depreciate, reversing the usual dynamics.

Proposition 3. *Consider the economy defined in Section 5.3 and let $n \to 1$. Suppose that in period t, the debt limit in Home is unexpectedly and permanently reduced from $\bar{B}^H$ to $\bar{B}^L$. The real exchange rate response at time t depends on whether this shock is large enough to push the union to the zero lower bound.*

(a) If the deleveraging shock is small, $\beta^s \bar{B}^H - \bar{B}^L < \underline{\zeta}$, monetary policy is unconstrained by the zero lower bound and the real exchange rate depreciates. Formally, we have

$$Q_t \; (:= \; Q_t^{NoZLB}) \;=\; \left[(1-\omega)\left(1-\beta^s\left(1-\frac{(1-\chi)Y+\chi\bar{B}^H}{(1-\chi)Y+\chi\bar{B}^L}\right)\right)+\omega\right]^{1-\omega} \;>\; 1.$$

(b) If the deleveraging shock is large, $\beta^s \bar{B}^H - \bar{B}^L > \underline{\zeta}$, the zero lower bound binds in the period of deleveraging. If wages in Home are not much more flexible than in

Foreign, $\gamma^*/\gamma < 1+\kappa$, where $\kappa > 0$, then the following inequality holds

$$Q_t\ (:=\ Q_t^{ZLB})\ =\ \max\left(\left[\frac{\gamma^*}{\gamma}\right]^{1-\omega}, \left[1-\omega+\left(1-(1-\omega)\beta^s\right)\frac{Y_t}{Y}\right]^{1-\omega}\right)\ <\ Q_t^{NoZLB}.$$

That is, the zero lower bound generally dampens the real depreciation. Moreover, for a sufficiently large shock, $\beta^s\bar{B}^H - \bar{B}^L > \bar{\zeta} \geq \underline{\zeta}$, the second part in $\max(\cdot,\cdot)$ above is below one such that the real exchange appreciates ($Q_t^{ZLB} < 1$), provided wages in Home are less flexible than in Foreign, $\gamma^*/\gamma < 1$.

Proof. See Appendix. □

The critical values $\underline{\zeta}$, $\bar{\zeta}$ and κ are provided in the appendix along with the proof of Proposition 3. The solution for Y_t is given in equation (5.4.4) below; note that $Y_t < Y$, the full-employment output level. The first part of the proposition establishes that country size *per se* does not alter the sign of the real exchange rate response. Intuitively, if monetary policy is not constrained in pursuing the inflation target, it reduces the nominal interest rate sufficiently in response to the deleveraging shock. Average inflation in the currency union remains at zero, because Home inflation is zero ($n \to 1$). Lower interest rates, in turn, raise consumption in Foreign. In the presence of home bias, this pushes up the price level in Foreign. Thus the Home real exchange rate depreciates.

The second part of the proposition shows that if Home is large, the real exchange rate response is generally hampered whenever monetary policy becomes constrained by the zero lower bound. In this case, monetary policy is unable to stabilise Home prices, as this would require pushing nominal interest rates into negative territory. As Home demand collapses, nominal wages (and as a result: prices from (5.3.12)) decline up to the floor set by downward rigidity, parametrised by γ. This has implications for Foreign, too. As before, Foreign consumption will tend to increase to the extent that monetary policy reduces interest rates. However, there is now a second effect: the demand for Foreign-produced goods falls with Home consumption. This exerts downward pressure on the Foreign price level: there are deflationary spillovers which dampen the real exchange rate depreciation. In fact, if the shock is large enough ($\beta^s\bar{B}^H - \bar{B}^L > \bar{\zeta}$) and if Foreign wages are more flexible than Home wages, the Foreign price level may decline more strongly than the Home price level: the Home real exchange rate appreciates.[14]

This finding qualifies a result obtained by Cook and Devereux (2014). They use a two-country model to contrast the effect of a negative demand shock under flexible exchange rates with that under a common currency. In case the zero lower

14. Thus a sufficient condition for the real exchange rate depreciation to be dampened for an intermediate-sized shock, and even to be reversed for a larger shock, is given by $\gamma^* < \gamma$—that is, Foreign wages are more flexible than Home wages.

bound binds, they find the Home real exchange rate to appreciate, but only if the nominal exchange rate is flexible. Under a common currency, instead, there is no such "perverse adjustment" of the real exchange rate. As we allow for differential degrees of stickiness in the two countries, we find that, for a large enough shock, the real exchange rate response may in fact also reverse its usual pattern under a common currency.

Having established that real exchange rate movements are dampened at the zero lower bound, we now turn to the role of wage rigidity in the adjustment process. It turns out that at the zero lower bound, increasing wage flexibility is actually destabilising, in line with the findings of Eggertsson and Krugman (2012) for the closed economy. We summarise this result in the following Proposition.

Proposition 4. *Paradox of flexibility. Consider the economy defined in Section 5.3 and let* $n \to 1$*. Assume the zero lower bound is binding in the period of deleveraging* ($\beta^s \bar{B}^H - \bar{B}^L > \underline{\zeta}$)*. In this case, if wages become more flexible domestically (that is, as* γ *is reduced)*

(a) Output, borrower consumption and real wage income all decline strictly more.

(b) The real exchange rate depreciates strictly less (or, if $\beta^s \bar{B}^H - \bar{B}^L > \bar{\zeta} \geq \underline{\zeta}$*, appreciates strictly more), provided the wage rigidity condition (5.3.14) is not binding in Foreign.*

Proof. See Appendix. □

The reason why more flexibility is harmful is simple: debt deflation. To see this, we again rearrange the borrower budget constraint (5.3.4) to obtain

$$C_t^b \ = \ -\frac{\bar{B}^H - \bar{B}^L}{P_t} + \frac{W_t L_t}{P_t}, \tag{5.4.3}$$

where we have used that $R_t = 1$ in the period of deleveraging. As before, the shock exerts downward pressure on domestic prices, making real debts harder to repay. And again, the general equilibrium response of real wage income is crucial for how the borrower responds to the shock.

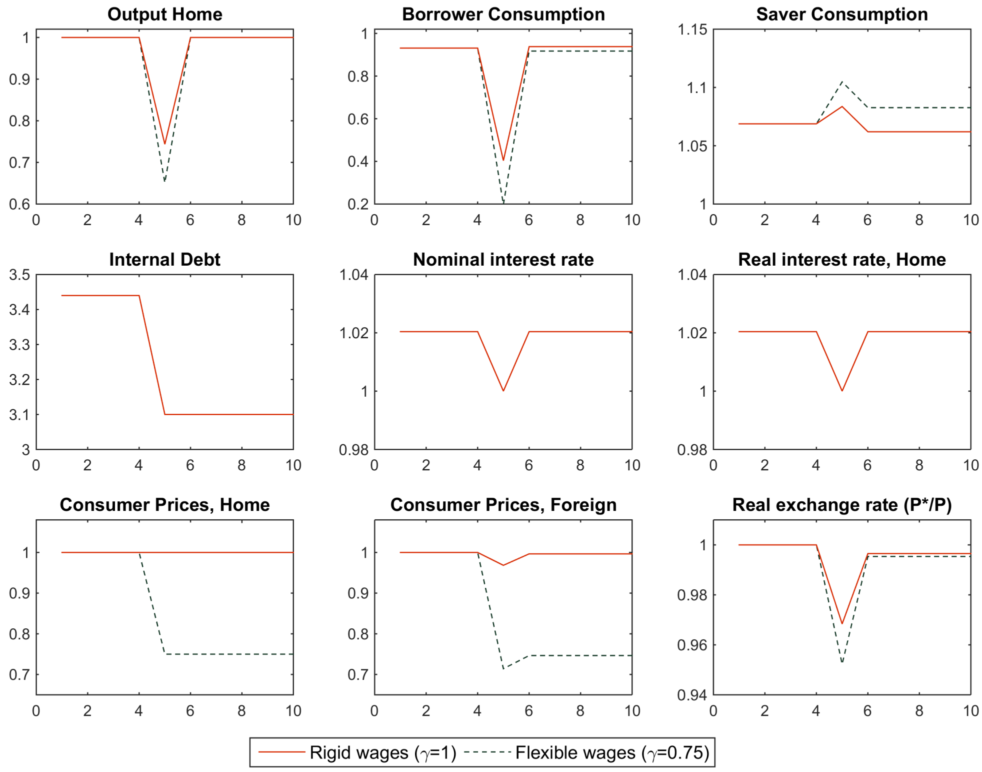
As $n \to 1$, real wage income corresponds to domestic output, $W_t L_t / P_t = Y_t$, which follows from combining (5.3.11) with (5.3.12), and using that $P_t = P_{H,t}$ (see above). In turn, domestic output in the period of deleveraging solves[15]

$$Y_t \ = \ (\beta^s)^{-1} \left[Y - \frac{\chi}{1-\chi} \left(\frac{\beta^s \bar{B}^H - \bar{B}^L}{\gamma} \right) \right] \ < \ Y. \tag{5.4.4}$$

The second part of this expression is negative, and more so, the more flexible the domestic wages (the lower the γ). That is, in contrast to the case of $n \to 0$ analysed

15. See the appendix, the proof of Proposition 4.

Figure 5.4.2. Deleveraging in a large union member (n ⟶ 1)



Notes. Parameter values: $\bar{L} = 1, \bar{B}^H = 3.44, \bar{B}^L = 3.1, \chi = 0.5, \omega = 0.2, \beta^s = 0.98, \beta^b = 0.97, \gamma^* = 0.7$.

earlier, economic activity (and therefore real wage incomes) now decline with more flexible wages.

The intuition here is that while lower prices still increase the real value of debts, they no longer stimulate economic activity by crowding in foreign demand or by lowering the real interest rate. For foreign demand, at the zero lower bound deflationary spillovers imply that the real exchange rate response is dampened (Proposition 3).[16] Real interest rates could fall because of either a cut in nominal rates, or expected inflation. But the zero lower bound implies that nominal interest rates are stuck at zero, and under inflation targeting a drop in prices does not generate expected inflation.

As pointed out in Proposition 3, for a large enough shock the fall in Home consumption creates large enough spillovers such that Foreign prices fall by more than Home prices, to the extent that the former are more flexible than the latter. Part a) of Proposition 4, in turn, states that these spillovers can only *increase* as domestic rigidi-

16. Incidentally, in the case of $n = 1$, foreign demand would not be crowded in even if the real exchange rate depreciated substantially, given Foreign's negligible size. However, as we turn to $n < 1$ next, we believe the deflationary spillovers are the more relevant intuition.

ties are reduced. Thus, somewhat paradoxically, more flexible wages in Home may lead to an even larger appreciation of its real exchange rate (part b) of Proposition 4).[17] Figure 5.4.2 shows the full adjustment dynamics for precisely such a case, by comparing fully rigid ($\gamma = 1$, solid lines) and more flexible wages ($\gamma = 0.75$, dashed lines).

From a conceptual point of view, we also note that there is a critical level of wage flexibility beyond which the model has no solution—or equilibrium—because the real value of debt repayments exceeds the level of output under full employment, which would imply $Y_t < 0$ in equation (5.4.4). Moreover, there is no solution even as wages become fully flexible. Intuitively, for a large deleveraging shock, the flexible-wage equilibrium requires a real interest rate below unity to induce savers to consume all of the repayment made by the borrowers. At the zero lower bound, this requires expected inflation—which is inconsistent with the central bank's inflation targeting mandate.[18] A flexible-wage equilibrium would exist, however, under an alternative monetary policy rule which generates the required expected inflation, such as a price level targeting rule. Moreover, such an alternative monetary policy rule could maintain full employment even under partially rigid wages, avoiding both the paradox of flexibility and the deleveraging recession itself.

The above insight relates to a point made by Cochrane (2015), who maintains that in New Keynesian models equilibria involving deep recessions and paradoxes of flexibility at the ZLB are the result of monetary policy rules containing an element of equilibrium selection. In our model, we account for alternative monetary policy rules as we vary the country size n. When $n \to 0$, monetary policy in Home corresponds to a simple exchange rate peg. As discussed in Corsetti, Kuester, and Müller (2013), such a peg provides an implicit price level commitment—in our case, to prices in Foreign—which is stabilising because it generates expected inflation. As n increases, a Foreign country with more rigid wages can still provide such a commitment mechanism, albeit to a lesser extent. Instead, for $n \to 1$ the price level commitment disappears and we are left with a simple inflation targeting rule.

In the next section we turn to analyse the numerical solution of our model for an empirically relevant, intermediate value of n. Generally speaking though, the insights gained from looking at the limiting cases $n \to 0$ and $n \to 1$ carry over to our quantitative analysis. Most importantly, as long as the deleveraging shock is large enough to push the union to the zero lower bound, deflationary spillovers will generally mute the adjustment of relative prices, and may sometimes even reverse them. This lack of relative price adjustment, and the lack of movement in the real

17. Note that $\gamma^* < \gamma$—the condition for the real exchange to appreciate if the shock is large (part b) of Proposition 3)—is implied by the wage rigidity condition (5.3.14) not binding in Foreign, as assumed in part b) of Proposition 4. See the appendix, the proofs of the two propositions for further details.

18. We show this formally in note which is available on request.

interest rate, imply that Home could enter a deep recession even if wages are quite flexible.

5.5 Quantitative analysis

The analytical results established in the previous section show that country size matters for the adjustment dynamics to a region-specific deleveraging process. In 2009, credit growth stalled in most countries of the euro area. However only the stressed economies of the euro area experienced a full-fledged deleveraging process in the years thereafter (see Section 5.2). In the following we perform a quantitative assessment of our model and explore to what extent it can account for key features of the post-crisis slump in the euro area. Precisely, we examine the non-linear impulse response to a deleveraging shock in two settings. First, we consider the baseline model introduced in Section 5.3, where deleveraging takes place over one period. Second, we use a modified version of the model, where borrowers are allowed to deleverage gradually and choose the optimal path of deleveraging over a number of periods.

5.5.1 Baseline model

We assign parameter values in order to solve the model numerically. The specific values are summarised in Table 5.5.1. The parameters which govern the size of the stressed economy and the size of the shock are determined by the following observations. First, we set $n = 0.37$ to match the share of stressed countries in euro-area GDP on the eve of the crisis. Second, we observe that the level of total private sector debt in the stressed economy declines from 172% of GDP to 155% of GDP, that is, by 17 percentage points of GDP. The debt limit in our model, $\bar{B}_t$, is debt *per borrower*, which means that we have to scale up the economy-wide debt-to-GDP values by a factor $1/\chi$, in our case 2. This gives us deleveraging from 344% GDP per borrower ($\bar{B}^H = 3.44$) to 310% GDP per borrower ($\bar{B}^L = 3.1$), a total of 34% GDP.[19] We use the borrower share parameter χ from Martin and Philippon (2014), which captures the share of liquidity-constrained households in the data from Eurosystem Household Finance and Consumption Survey (HFCS).

The other parameters are standard. Assuming a discount factor β^s of 0.98 for the patient households implies an annual real interest rate of 2% in steady state. We set the home-bias parameter $\omega = 0.2$. Given $n = 0.37$ this implies an import share of 0.12, the average GDP weight of imports from the rest of the euro area in the stressed economies.[20] Finally, we vary the parameter γ which captures downward wage rigidities. Specifically, we consider a range of values between 0.93 and 0.99,

19. The upper limit is the peak value, observed in 2012Q2, whilst the trough value is the latest observation in our sample, 2014Q3. As before, in our model nominal GDP equals 1 in the pre-crisis steady state, and a deleveraging shock of 34% is equivalent to 34% of pre-crisis nominal GDP.

20. Source: OECD, Monthly Foreign Trade Statistics, period: 1999–2006.

Table 5.5.1. Parameters for quantitative analysis: baseline model

Parameter:	n	$\bar{B}^H$	$\bar{B}^L$	β^s	ω	χ	γ
Value:	0.37	3.44	3.1	0.98	0.2	0.5	0.93–0.99

which is equivalent to maximum wage deflation of between 7% and 1% per year respectively. The high downward wage rigidity implied by $\gamma = 0.99$ is broadly in line with estimates for the euro area (Schmitt-Grohé and Uribe, 2016).

Figure 5.5.1 shows the response of this economy to the deleveraging shock, contrasting three scenarios. Solid lines represent our baseline, characterised by high wage rigidities ($\gamma = 0.99$) in Home. In addition we consider two counterfactuals. The lines with crosses correspond to a scenario where Home rigidities are reduced, but remain above Foreign ($\gamma = 0.96$ vs $\gamma^* = 0.95$). The dashed lines correspond to a case where rigidity in Home is significantly lower, to the extent that Home becomes more flexible than Foreign ($\gamma = 0.93$ vs $\gamma^* = 0.95$).

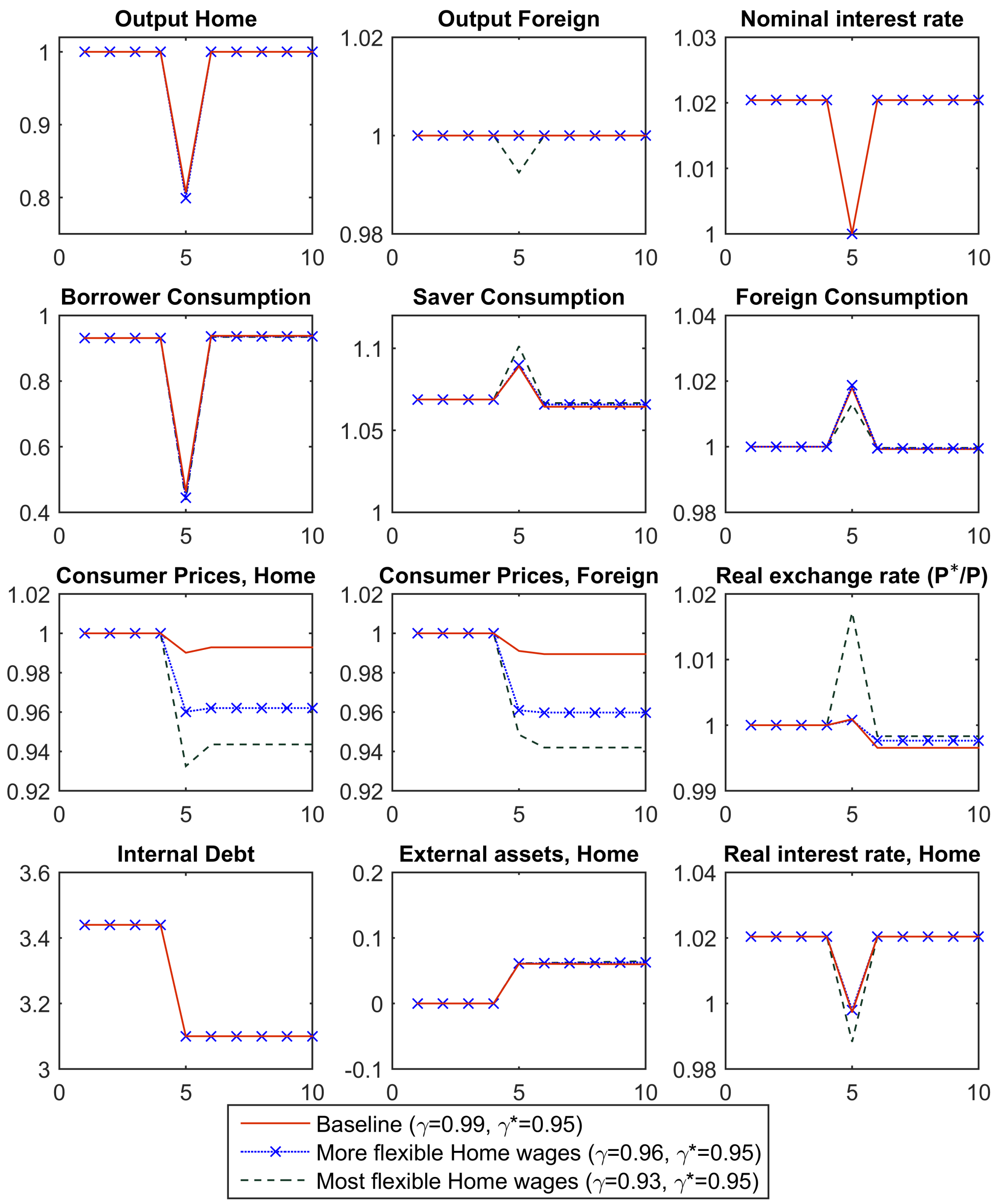
The deleveraging shock is displayed in the bottom-left panel. Under our baseline scenario, it pushes the economy into a deep but asymmetric recession: while Home output collapses, Foreign output remains unaffected, even though monetary policy becomes constrained by the zero lower bound (first row of Figure 5.5.1). The dynamics of consumption also differ. While borrowers reduce consumption to repay their debts, Home savers' and Foreign households' consumption increases (second row of Figure 5.5.1), reflecting a reduced real interest rate (bottom-right panel for the case of Home).

The third row shows the responses of variables of particular interest. For the baseline case we find a mild decline of the price level in Home, reflecting the presence of downward wage rigidities. At the same time, Foreign prices decline—if only by little—due to deflationary spillovers. As a result, the real exchange rate remains broadly unchanged in the baseline scenario.

Turning to the counterfactuals, consider first the case of somewhat increased wage flexibility ($\gamma = 0.95$, crossed lines) in Home. Despite the sharper drop in prices in Home, the real exchange rate movement is virtually identical to that under the baseline scenario of higher wage rigidity (third row). This is because more flexible wages increase debt deflation and deepen the recession in Home—an instance of the paradox of flexibility (see Proposition 4 for the $n \to 1$ case). Deflationary spillovers increase, such that prices in Foreign decline by more to stabilise output in the face of an export collapse.

Under our second counterfactual, wage rigidities in Home are relaxed yet further, to the point where wages are more downwardly flexible than in the rest of the union ($\gamma = 0.93$, $\gamma^* = 0.95$, dashed lines). Following the shock, both countries find themselves against the wage rigidity constraint in (5.3.15), but because Home

Figure 5.5.1. Deleveraging in a medium-sized union member, $n = 0.37$



Notes. Baseline model. Parameter values in Table 5.5.1. One period is a year.

wages are more flexible, prices in Home fall by more (third row). The real exchange rate depreciates, more so than in the baseline scenario. Turning to output (top row), Foreign now enters a mild recession because its prices cannot fall by enough to fully offset the negative demand spillover from Home. Even though aggregate demand and Home output are boosted by the real depreciation and a lower real interest rate, the benefits of these are roughly offset by higher debt deflation, and Home output falls by roughly the same amount as in the baseline scenario.

On reflection, the results in this section resemble the $n \to 1$ case of Section 5.4.2 more than they do the $n \to 0$ case of Section 5.4.1. The deleveraging is associated with a large output drop in Home and deflationary spillovers to Foreign, which increase in size with higher wage flexibility. Unlike Home output and prices, the real exchange rate and Foreign output change little in all three of our simulated scenarios.

In the next section, we examine a fuller model which allows us to relate our quantitative results to the dynamics of the post-crisis slump in the euro area. To do this, we allow the deleveraging to take place gradually, and over a longer time period.

5.5.2 Dynamic Deleveraging

Euro-area deleveraging has been taking place for the best part of six years, and the pace of the deleveraging has been quite gradual. To account for this, we modify our model in one key way. Borrowers no longer face a fixed borrowing limit $\bar{B}_t$. Instead, when debts in the economy exceed some level perceived as "safe"—denoted $\bar{B}_{S,t}$—the savers (through financial intermediaries) start charging higher interest rates on borrowing. A deleveraging shock is then a reduction in this safe level of debt, which increases interest spreads and incentivises borrowers to reduce their indebtedness. In this regard we mimic the set-up of Benigno, Eggertsson, and Romei (2014) who consider a closed economy.[21]

Formally, borrowers now maximise

$$\max_{\{C_t^b\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} (\beta^b)^t \ln(C_t^b)$$

subject to

$$P_t C_t^b + B_{t-1}^b = (R_t^b)^{-1} B_t + W_t L_t, \tag{5.5.1}$$

where R_t^b is the gross borrowing rate determined according to

$$R_t^b = R_t \left(B_t^{\mathrm{Av}} / \bar{B}_{S,t}\right)^{\phi} \tag{5.5.2}$$

$$R_t^b \geq R_t. \tag{5.5.3}$$

Note that the spread is determined by the average debt per borrower throughout the economy, rather than individual debt, even though in equilibrium $B_t^{\mathrm{Av}} = B_t^b$. Here, ϕ

21. Our set-up differs only slightly from Benigno, Eggertsson, and Romei (2014), in the following ways. First, we examine two open economies. Second, to improve tractability, the spread is a function of nominal, not real, debt. Last, we provide a full non-linear solution of the model under perfect foresight.

is the elasticity of the gross borrowing rate with respect to excessive debt. As argued in Benigno, Eggertsson, and Romei (2014), one can interpret this set-up as capturing financial intermediation in a very stylised way: banks lend to borrowers at rate R_t^b and pay the savers R_t. The premium for excessive borrowing can be interpreted as a charge for default risk in the presence of asymmetric information, or as compensating for fraud. And the profits of these transactions are distributed to savers, who own the banks.[22]

Optimality requires that borrowers satisfy an Euler equation

$$\left(C_t^b\right)^{-1} \;=\; \beta^b R_t^b \left(C_{t+1}^b\right)^{-1} \frac{P_t}{P_{t+1}}. \tag{5.5.4}$$

Furthermore, as in Benigno, Eggertsson, and Romei (2014), we let $\beta^b \to \beta^s = \beta$, such that in steady state $B^b = \bar{B}_S$ from combining (5.5.2) and (5.5.4), and so banks make zero profits.

Lastly, for our numerical analysis of the gradual deleveraging model, we assume that the central bank implements its inflation target via a Taylor-type rule of the form

$$R_t \;=\; \left(\beta^s\right)^{-1} (\Pi_t^u)^{\varphi^\pi} \;\; \text{subject to } R_t \;\geq\; 1, \tag{5.5.5}$$

where we assume $\varphi^\pi > 1$.

We measure periods in quarters, and adjust the model parameters accordingly. All parameter values are listed in Table 5.5.2. Values for n, ω and χ are unchanged from before. The quarterly discount factor $\beta = 0.995$ is equivalent to a 2% annualised real interest rate in steady state. Debt limits of 344% and 310% annual GDP per borrower translate to 1376% and 1240% of quarterly GDP respectively. We set the Taylor rule parameter φ^π to the conventional value of 1.5. Wage rigidity occupies a range of values, allowing for maximum wage falls of between 4% and 8% per year.[23] Finally, we set the spread elasticity to 0.38 in order to target a zero-lower-bound episode which lasts 6 quarters.

We now turn to our quantitative analysis. We assume that at time t, the safe debt limit unexpectedly and permanently tightens from $\bar{B}^H$ to $\bar{B}^L$. From (5.5.2), this opens up a spread between borrowing and lending rates, since the level of debt becomes judged as "excessive": $\bar{B}^H > \bar{B}^L = \bar{B}_{S,t}$, triggering a deleveraging from equation (5.5.4), until the new safe debt level $\bar{B}^L$ is reached in the long run. The dynamic

22. The savers' budget constraint is now given by $P_t C_t^s + R_t^{-1} B_t^s + R_t^{*-1} D_t = W_t L_t + B_{t-1}^s + D_{t-1} + \frac{\chi}{1-\chi}\left(R_t^b - R_t\right) B_t^b$, where the last term are banking profits distributed to savers in a lump-sum manner.

23. Note that we examine a slightly narrower range than in Section 5.5.1, to preserve stability of the numerical solution. Because of dynamic feedback effects, large changes in wage flexibility can have dramatic effects on model outcomes: they trigger a deflationary spiral, such that a stable model solution does not exist.

Table 5.5.2. Parameters for quantitative analysis: dynamic deleveraging model

Parameter:	n	$\bar{B}^H$	$\bar{B}^L$	β	ω	χ	γ	φ	φ^π
Value:	0.37	13.76	12.4	0.995	0.2	0.5	0.98–0.99	0.38	1.5

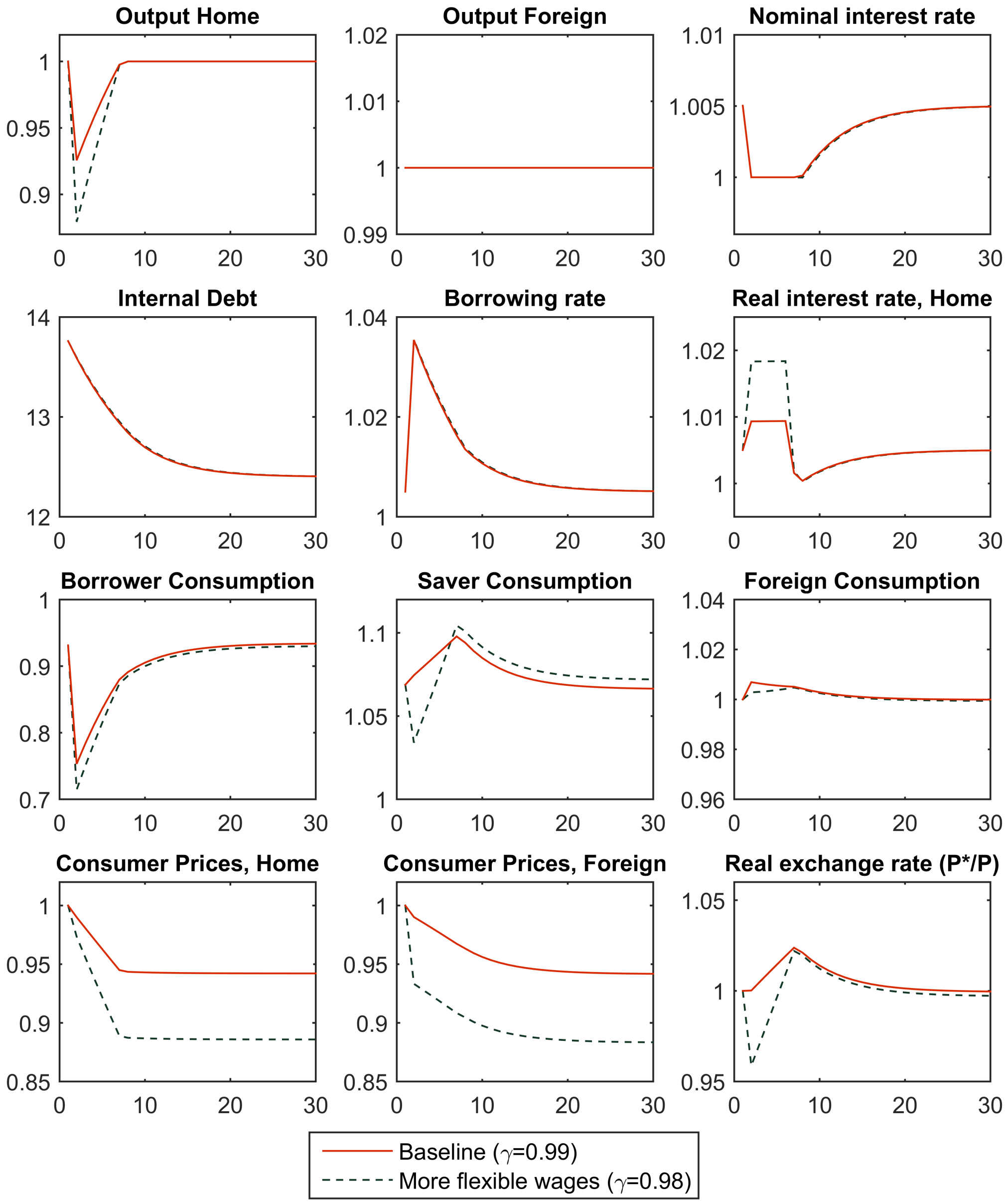
response to the shock is depicted in Figure 5.5.2. We compare the response in a baseline scenario of rigid wages in stressed countries (red solid lines) to a counterfactual scenario of more flexible wages (black dashed lines). Throughout this experiment, we assume that wages in Foreign are flexible enough to support full employment.

From the top row, we can see that the deleveraging shock generates a recession and pushes the union to the zero lower bound. Looking at the mechanism in more detail, the second row shows the associated increase in spreads which triggers a slow and gradual debt reduction towards the new, lower, safe level. The debt repayment acts as a long-lasting drag on borrower consumption (third row, left), which lowers aggregate demand and generates deflationary pressures. In the early stages of the adjustment, the central bank is unable to cut interest rates sufficiently in order to offset these pressures, and the economy enters a recession.

Turning to the focus of our experiment, the bottom row depicts the movements in relative prices. In every period of the recession, Home wages fall by the maximum amount permitted by the rigidity constraint, which leads to a fall in Home prices. Under the baseline scenario, Foreign prices fall slightly due to the deflationary spillovers, dampening the movement in the real exchange rate. Prices in both countries fall by around 1% in the first quarter such that there is no real depreciation. Thereafter prices in Home fall by around 5.5% and in Foreign—by 3%, so just over half of the price fall in Home is dampened by the deflationary spillover to Foreign, and the resulting real depreciation is rather modest in size.

The absolute price movements under the counterfactual scenario of more flexible wages (bottom row, dashed lines) are much more pronounced. However this does not mean that *relative* prices adjust by more—in fact, they adjust by less and even move in the “wrong” direction. This is because the sharp fall in Home prices triggers a destabilising deflationary spiral and increases the negative demand spillover to Foreign, and with it the pressure on Foreign prices to drop *relative* to Home prices. Debt deflation in Home forces borrowers to reduce consumption further (third row, left). At the same time, expected deflation raises real interest rates (second row, right), such that even savers cut back on consumption (third row, middle). This in turn reduces real incomes throughout the Home economy and deepens the recession. The resulting fall in demand for Foreign exports pushes down Foreign prices to the extent that the real exchange rate appreciates sharply at first, further depressing real incomes and exacerbating the deleveraging in Home. The paradox of flexibility

Figure 5.5.2. Dynamic deleveraging in a medium-sized union member, $n = 0.37$



Notes. Parameter values in Table 5.5.2. One period is a quarter. Interest rates shown on a quarterly basis.

is in full force, and the recession is deeper than under the baseline scenario of more rigid wages.

The discussion above supports the conclusions of our analytical model in Section 5.4 and the one-period deleveraging experiment in Section 5.5. Further, the impact of higher wage flexibility—and the associated paradoxes—seem to be much more pertinent in the dynamic setting. More precisely, the differences in output and real exchange rate movements between the baseline and flexible-wage counterfactuals are much larger. Whilst under one-period deleveraging higher wage flexibility was not helpful but largely harmless, under dynamic deleveraging it can be highly destabilising. The reason for this is twofold. First, because deleveraging is more gradual, the adverse effects of debt deflation can persist for longer. Second, because some of the deflation is anticipated, it raises real interest rates and reduces consumption and demand further.

Even though the model in this section remains fairly stylised, the impulse responses for our baseline scenario match a number of facts which characterise the post-crisis slump in the euro area (Section 5.2). Just like the stressed euro-area economies, there is a deep recession in the Home country, triggered by a long and gradual deleveraging process. The recession persists for several periods and pushes the union to the zero lower bound. Prices in Home and Foreign move in a synchronised manner, which means that the real depreciation is small, around 2.4% at its peak. This is still somewhat larger than the "raw" 0.6% real depreciation we observe for the stressed euro countries in the data, but broadly in line with the 2% depreciation calculated using tax-adjusted price data.[24] The fall in the Foreign price level helps the rest of the union remain at full employment, so whilst prices move together, the output performance of the two country groups diverges. Our counterfactual simulation suggests that had wages in stressed economies been more flexible, we would have still seen little relative price adjustment within the euro area. Indeed, the real exchange rate may have even appreciated, deepening the recession.

5.6 Conclusion

Why—given the heterogenous economic performance across the euro area after 2009—has there been no significant adjustment of intra-euro-area real exchange rates? In this paper, we address the question by setting up a model of a currency union in which one of the two member countries experiences a large deleveraging shock. A key feature of the model is that the size of countries may differ. This allows us to study polar cases for which we are able to solve the model in closed form. In doing so we clarify aspects of the transmission mechanism, which also turn out to be relevant in the quantitative assessment of the model.

In case the domestic economy is small, there are no spillovers from the deleveraging shock into the rest of the union. Also, monetary policy remains unchanged. In this case, in line with much of the received wisdom, the extent of real deprecation is inversely related to the degree of wage rigidities. In the context of our model, this is simply the result of prices fluctuating one-for-one with wages. More flexibility in wages not only makes the real exchange rate more flexible, it also makes the

24. Figures quoted are for 2008Q4—2014Q4.

economy more resilient in the face of adverse country-specific shocks. This result is notable because more wage flexibility also leads to more debt deflation in response to the deleveraging shock. Still, this effect is not strong enough to offset the benefits from increased flexibility for a small member country of a currency union.

A different picture emerges once we consider a large currency union member. Monetary policy lowers interest rates in response to the shock, and reaches the zero lower bound if the shock is large enough. The output response in the domestic economy and in the rest of the union will generally be different, in particular if there is sufficient wage flexibility in the rest of the union: while domestic output collapses, it may stay at its initial level in the other countries. At the same time there is deflationary pressure across the entire union, and the response of the real exchange rate to the deleveraging shock is dampened if the zero lower bound binds. In fact, the real exchange rate may even appreciate if the shock is large and/or wages are relatively flexible in the rest of the union. More wage flexibility in the domestic economy can also be destabilising and induce an even stronger real appreciation—an instance of the paradox of flexibility, established by Eggertsson and Krugman (2012) in a closed-economy context.

In our quantitative assessment we assume that the domestic economy accounts for 37 percent of the currency union, corresponding to the GDP weight of the stressed economies prior to the crisis. In response to a deleveraging shock of 34 percent GDP, we find the model able to account for some of the basic facts of the post-crisis slump in the euro area: there is large divergence in terms of output performance, there is deflationary pressure in the entire union, and the real exchange rate response is muted.

We also find that the paradox of flexibility reigns in our calibrated model. In a counterfactual scenario where wages are more flexible, the real exchange rate appreciates and the output loss is even larger than in the baseline case. Hence, we think that our analysis has some relevance for the ongoing policy debate in the euro area. Namely, similarly to Eggertsson, Ferrero, and Raffo (2014), we find that measures geared towards raising the flexibility of the economies hit hardest by the crisis are likely to be ineffective—unless they are accompanied by expansionary monetary policy measures which may make up for the zero-lower-bound constraint on interest rates.

Appendix 5.A Deleveraging, Deflation and Depreciation: Appendix

This Appendix presents proofs of Propositions 1—4.

5.A.1 Proposition 1 proof

For $n \to 0$, we can reduce the system of equations (5.3.1)—(5.3.21) to the following 7 equations:

$$
\begin{aligned}
P_t C_t^s &= P_{t+1} C_{t+1}^s && (5.A.1)\\
P_t C_t^b &= \beta^s \bar{B}_t - \bar{B}_{t-1} + P_{H,t} Y_t && (5.A.2)\\
C_t &= (1-\chi) C_t^s + \chi C_t^b && (5.A.3)\\
(1-\chi)(\beta^s D_t - D_{t-1}) &= P_{H,t} Y_t - P_t C_t && (5.A.4)\\
P_t &= P_{H,t}^{1-\omega} && (5.A.5)\\
P_{H,t} Y_t &= (1-\omega) P_t C_t + \omega Y && (5.A.6)\\
0 &= (Y_t - Y)(P_{H,t} - \gamma P_{H,t-1}), && (5.A.7)
\end{aligned}
$$

as well as two inequalities:

$$
\begin{aligned}
Y_t &\leq Y && (5.A.8)\\
P_{H,t} &\geq \gamma P_{H,t-1} && (5.A.9)
\end{aligned}
$$

The first equation is derived from the savers' Euler equation (5.3.2), the second—from the borrowers' budget constraint (5.3.4) (combined with (5.3.11), (5.3.12)), and the third—from aggregate consumption (5.3.6). Equation (5.A.4) is the country's budget constraint, obtained by combining the saver's and borrower's budget constraints (5.3.1) and (5.3.4), and substituting for nominal incomes from (5.3.11) and (5.3.12). The remaining equations are the price index (from 5.3.9), Home market clearing (from 5.3.18), and the complementary slackness condition (from 5.3.15). The inequality in (5.A.9) is obtained by combining (5.3.14) and (5.3.12). Additionally, we have used that $R_t = (\beta^s)^{-1}$, $P_t^* = P_{F,t} = 1$, and $C_t^* = Y_t^* = Y$.

The perfect foresight solution is a sequence of endogenous variables $\{C_t^s, C_t^b, C_t, Y_t, P_t, P_{H,t}, D_t\}$ that solves equations (5.A.1)—(5.A.9), given the initial conditions (stated in Section 5.3.5), and an exogenous path for $\{\bar{B}_t\}$, known in the initial period.

We can solve for real and nominal variables separately. The system in nominal variables can be reduced further to three equations

$$
\begin{aligned}
P_t C_t^s &= P_{t+1} C_{t+1}^s && (5.A.10)\\
(1-\chi)(\beta^s D_t - D_{t-1}) &= (1-\chi) P_{H,t} Y_t - (1-\chi) P_t C_t^s - \chi(\beta^s \bar{B}_t - \bar{B}_{t-1}) && (5.A.11)\\
(1-(1-\omega)\chi) P_{H,t} Y_t &= (1-\omega)(1-\chi) P_t C_t^s + (1-\omega)\chi(\beta^s \bar{B}_t - \bar{B}_{t-1}) + \omega Y && (5.A.12)
\end{aligned}
$$

in three unknowns $(P_t C_t^s)$, $(P_{H,t} Y_t)$ and D_t.

We solve this system forward analytically. The full solution is presented in the proof of Proposition 2. Here, we only present the solution for Home nominal income $(P_{H,t} Y_t)$:

$$
P_{H,t} Y_t = Y + (1-\beta^s)\frac{(1-\chi)(1-\omega)}{\omega} D_{t-1} - \eta(\bar{B}_{t-1} - \bar{B}_t),
$$

where $\eta = (\beta^s(1-\omega)\chi)/(1-(1-\omega)\chi) > 0$.

In the period of deleveraging, $D_{t-1} = 0$, $\bar{B}_{t-1} = \bar{B}^H$ and $\bar{B}_t = \bar{B}^L$, which yields

$$
P_{H,t} Y_t = Y - \eta(\bar{B}^H - \bar{B}^L), \tag{5.A.13}
$$

Under full employment, $Y_t = Y$, and hence

$$
P_{H,t} = 1 - \eta(\bar{B}^H - \bar{B}^L)/Y. \tag{5.A.14}
$$

If there is unemployment, $Y_t < Y$ and from (5.A.7)

$$
P_{H,t} = \gamma P_{H,t-1} = \gamma
$$

From (5.A.9),

$$
P_{H,t} = \max(\gamma, 1 - \eta(\bar{B}^H - \bar{B}^L)/Y \leq 1. \tag{5.A.15}
$$

The inequality holds because $\gamma \leq 1$.

From (5.3.10) and (5.A.5), $Q_t = P_{H,t}^{\omega-1}$, where $\omega - 1 < 0$. Combining this with (5.A.15) gives us the real exchange rate formula in Propostion 1:

$$
Q_t = P_t^*/P_t = \min\left(\gamma^{\omega-1}, [1 - \eta(\bar{B}^H - \bar{B}^L)/Y]^{\omega-1}\right) \geq 1.
$$

5.A.2 Proposition 2 proof

We first present the full solution to the system of equations (5.A.10)—(5.A.12):

$$
\begin{aligned}
P_t C_t^s &= Y + (1-\beta^s)\left(\frac{\chi}{1-\chi}\bar{B}_{t-1} + \frac{1-(1-\omega)\chi}{\omega}D_{t-1}\right) \\
P_t C_t^b &= Y - (1-\beta^s)\left(\bar{B}_{t-1} - \frac{(1-\chi)(1-\omega)}{\omega}D_{t-1}\right) - \frac{\beta^s}{1-(1-\omega)\chi}(\bar{B}_{t-1}-\bar{B}_t) \\
P_{H,t}Y_t &= Y + (1-\beta^s)\frac{(1-\chi)(1-\omega)}{\omega}D_{t-1} - \frac{\beta^s(1-\omega)\chi}{1-(1-\omega)\chi}(\bar{B}_{t-1}-\bar{B}_t) \\
D_t &= D_{t-1} + \frac{\omega\chi}{(1-(1-\omega)\chi)(1-\chi)}(\bar{B}_{t-1}-\bar{B}_t).
\end{aligned}
$$

In the period of deleveraging, $D_{t-1}=0$, $\bar{B}_{t-1}=\bar{B}^H$ and $\bar{B}_t=\bar{B}^L$. Substituting for this gives the following expressions for nominal spending and incomes:

$$
\begin{aligned}
P_t C_t^s &= Y + (1-\beta^s)\frac{\chi}{1-\chi}\bar{B}^H \\
P_t C_t^b &= Y - (1-\beta^s)\bar{B}^H - \frac{\eta}{(1-\omega)\chi}(\bar{B}^H-\bar{B}^L) \\
P_{H,t}Y_t &= Y - \eta(\bar{B}^H-\bar{B}^L),
\end{aligned}
$$

where $\eta=(\beta^s(1-\omega)\chi)/(1-(1-\omega)\chi)>0$, as above.

We can immediately see that nominal spending and incomes are independent of wage flexibility γ. Savers' nominal spending is independent of the deleveraging shock, whilst borrowers' nominal spending and incomes fall proportionately with the deleveraging shock $\bar{B}^H-\bar{B}^L$. Because of this, the smaller the nominal adjustment in prices, the larger the adjustment of real spending and incomes will be, which gives us some intuition for the benefits of higher wage flexibility. We establish this more formally below.

Suppose first that the wage rigidity condition (5.A.9) is binding. Then $P_{H,t}=\gamma$ and $P_t=\gamma^{1-\omega}$ from (5.A.5). Real spending and output, as well as real incomes W_tL_t/P_t, are given below

$$C_t^s = \gamma^{\omega-1}\left[Y + (1-\beta^s)\frac{\chi}{1-\chi}\bar{B}^H\right] \tag{5.A.1}$$

$$C_t^b = \gamma^{\omega-1}\left[Y - (1-\beta^s)\bar{B}^H - \frac{\eta}{(1-\omega)\chi}(\bar{B}^H-\bar{B}^L)\right] \tag{5.A.2}$$

$$Y_t = \gamma^{-1}\left[Y-\eta(\bar{B}^H-\bar{B}^L)\right] \tag{5.A.3}$$

$$W_tL_t/P_t = P_{H,t}Y_t/P_t = \gamma^{\omega}Y_t = \gamma^{\omega-1}\left[Y-\eta(\bar{B}^H-\bar{B}^L)\right], \tag{5.A.4}$$

where we have used (5.3.11) and (5.3.12) in the last equation. Note that all increase with lower γ, or higher wage flexibility (which follows from $\omega-1<0$).

When γ is low enough such that the wage rigidity condition in (5.A.9) is not binding, from (5.A.15), prices are at their lowest level given by (5.A.14), output is at full employment level, $Y_t = Y$, and consumption and real incomes are at their maximum level.[25] If γ is lowered further beyond this point, the real variables no longer change. This concludes the proof of Proposition 2.

5.A.3 Proposition 3 proof

For $n \to 1$, we can reduce the system of equations (5.3.1)—(5.3.21) to the following 11 equations:

$$\begin{aligned}
\beta^s R_t P_t C_t^s &= P_{t+1} C_{t+1}^s && (5.A.1)\\
P_t C_t^b &= R_t^{-1} \bar{B}_t - \bar{B}_{t-1} + P_t Y_t && (5.A.2)\\
C_t &= Y_t && (5.A.3)\\
C_t &= (1-\chi) C_t^s + \chi C_t^b && (5.A.4)\\
\beta^s R_t P_t^* C_t^* &= P_{t+1}^* C_{t+1}^* && (5.A.5)\\
R_t^{-1} D_t^* - D_{t-1}^* &= \omega (P_t C_t - P_t^* C_t^*) && (5.A.6)\\
P_{F,t} Y_t^* &= (1-\omega) P_t^* C_t^* + \omega P_t C_t && (5.A.7)\\
P_t^* &= P_{F,t}^{1-\omega} P_t^{\omega} && (5.A.8)\\
0 &= (R_t - 1)(P_t - P_{t-1}) && (5.A.9)\\
0 &= (Y_t - Y)(P_t - \gamma P_{t-1}), && (5.A.10)\\
0 &= (Y_t^* - Y)(P_{F,t} - \gamma^* P_{F,t-1}), && (5.A.11)
\end{aligned}$$

as well as these inequalities:

$$\begin{aligned}
Y_t &\le Y, \qquad Y_t^* \le Y && (5.A.12)\\
P_t &\ge \gamma P_{t-1}, \qquad P_{F,t} \ge \gamma^* P_{F,t-1} && (5.A.13)\\
R_t &\ge 1 && (5.A.14)
\end{aligned}$$

The first five equations are the Home savers' Euler equation (5.3.2), the borrowers' budget constraint (5.3.4), Home goods market clearing (5.3.18), Home aggregate consumption (5.3.6), and Foreign saver's Euler equation (5.3.8) respectively. Equation (5.A.6) is the Foreign country's budget constraint, which we obtain by combining the Foreign saver's budget constraint (5.3.7) with the Foreign goods market clearing condition (5.3.19), and using the expressions in (5.3.10), (5.3.11) and (5.3.12). The remaining equations are, in order, the Foreign goods market clearing (5.3.19), Foreign price index (5.3.9), monetary policy (equations 5.3.16 and 5.3.17),

25. That is, the maximum when varying wage flexibility and taking all other parameters as given.

and the complementary slackness conditions in Home and Foreign, in (5.3.15). Inequalities refer to the full employment constraint on output, downward wage rigidity and the zero lower bound. Throughout, we have used $P_t^u = P_t = P_{H,t} = W_t$, and $D_t = 0$ due to the large size of Home.[26]

Additionally, we write down a simplified expression for the savers' budget constraint in Home, which is not necessary to compute the solution, but provides a useful stepping stone in parts of the proof:

$$P_t C_t^s \;=\; \frac{\chi}{1-\chi}(\bar{B}_{t-1} - R_t^{-1}\bar{B}_t) + P_t Y_t. \tag{5.A.15}$$

The perfect foresight solution is a sequence of endogenous variables $\{C_t^s, C_t^b, C_t, Y_t, P_t, R_t, P_t^*, C_t^*, D_t^*, Y_t^*, P_{F,t}\}$ that solves equations (5.A.1)—(5.A.14), given the initial conditions (stated in Section 5.3.5), and an exogenous path for $\{\bar{B}_t\}$, known in the initial period. The solution will depend on whether the union is at the zero lower bound or not. We consider each of these two cases in turn.

(a) Outside of the zero lower bound

For the remainder of the proof, we adopt the same time notation as in Proposition 3: rather than using the general time subscript t, we denote t as the period of deleveraging, and $t-1$ as the initial steady state. Equation (5.A.9) then yields $P_t = P_{t-1} = 1$, and hence $Y_t = Y$ from equation (5.A.10): we have zero inflation and full employment in Home during the deleveraging period. At this point, it is helpful to write down the savers' budget constraint (5.A.15) at t and $t+1$:

$$\begin{aligned} C_t^s \;&=\; Y + \frac{\chi}{1-\chi}\left(\bar{B}^H - R_t^{-1}\bar{B}^L\right) \\ C_{t+1}^s \;&=\; Y + \frac{\chi}{1-\chi}(1-\beta^s)\bar{B}^L \end{aligned}$$

where we have used that $Y_t = Y_{t+1} = Y$, $\bar{B}_{t-1} = \bar{B}^H$ and $\bar{B}_t = \bar{B}^L$.

Plugging the above expressions into the savers' Euler equation (5.A.1), we obtain an expression for the nominal interest rate

$$R_t \;=\; (\beta^s)^{-1}\frac{(1-\chi)Y + \chi\bar{B}^L}{(1-\chi)Y + \chi\bar{B}^H} \tag{5.A.16}$$

Intuitively, the central bank cuts the nominal interest rate sufficiently such that the savers consume all of the debt repayments they receive from the borrowers to maintain full employment.

26. $P_t^u = (P_t)^n (P_t^*)^{1-n}$ is the union-wide price level.

We now turn to the developments in Foreign. In the period before deleveraging $D^*_{t-1} = 0$, and thereafter $D^*_{t+1} = D^*_t$ and $R_{t+1} = (\beta^s)^{-1}$. Also, there is no downward pressure on Foreign prices, and it remains at full employment, $Y^* = Y$.[27] We can then write down a 4x4 equation system which allows us to solve for the Foreign price level, and hence, the real exchange rate.

$$\begin{aligned} R_t P^*_t C^*_t &= (\beta^s)^{-1} P^*_{t+1} C^*_{t+1} && (5.A.17)\\ R_t^{-1} D^*_t &= \omega(P_t C_t - P^*_t C^*_t) && (5.A.18)\\ (1-\beta^s) D^*_t &= \omega(P^*_{t+1} C^*_{t+1} - P_{t+1} C_{t+1}) && (5.A.19)\\ P_{F,t} Y &= (1-\omega) P^*_t C^*_t + \omega P_t C_t && (5.A.20) \end{aligned}$$

The first equation is Foreign saver's Euler in (5.A.5), second and third—the Foreign country budget constraint (5.A.6) at t and $t+1$ respectively, and the last equation is the Foreign goods market clearing in (5.A.7). R_t is exogenous to Foreign and given by (5.A.16). Using that $P_t C_t = Y$, we first combine (5.A.17)—(5.A.19) to solve for $P^*_t C^*_t$

$$P^*_t C^*_t = Y(1 - \beta^s + R_t^{-1}), \tag{5.A.21}$$

and then substitute this expression into (5.A.20) to yield

$$Q_t = P^*_t / P_t = P_{F,t}^{1-\omega} = \left[(1-\omega)(1-\beta^s + R_t^{-1}) + \omega\right]^{1-\omega}, \tag{5.A.22}$$

which, combined with the interest rate expression in (5.A.16), gives us the real exchange rate formula in Proposition 3(a):

$$Q_t \ (:= Q_t^{\text{NoZLB}}) = \left[(1-\omega)\left(1 - \beta^s\left(1 - \frac{(1-\chi)Y + \chi \bar{B}^H}{(1-\chi)Y + \chi \bar{B}^L}\right)\right) + \omega\right]^{1-\omega}. \tag{5.A.23}$$

Because $\bar{B}^H > \bar{B}^L$, $Q_t^{\text{NoZLB}} > 1$, hence the real exchange rate depreciates.

(b) At the zero lower bound

Suppose that stabilising union-wide inflation (and therefore economic activity from (5.A.10)) would require $R_t < 1$. From equation (5.A.16), this requires a shock large enough, such that

$$\beta^s \bar{B}^H - \bar{B}^L > \frac{(1-\chi)}{\chi}(1-\beta)Y =: \underline{\zeta}.$$

In this case, we know that union-wide inflation cannot be stabilised (for if it were, output would be at potential as argued above, and thus the implied R_t would be

27. One can verify this ex post by checking the real exchange rate formula in (5.A.16). Since $Q_t > 1$, $P_{F,t} > 1$ and thus $P_{F,t} > \gamma^* P_{F,t-1}$ since $\gamma^* \leq 1$ and $P_{F,t-1} = 1$. From (5.A.11), this implies $Y^*_t = Y$.

negative). We shall demonstrate in Section 5.A.4 that in equilibrium, the price level will fall to its lower bound provided by downward wage rigidity, from (5.A.10), $P_t = \gamma P_{t-1} = \gamma$, and that $Y_t < Y$ so that output strictly drops below potential. Further note that in period $t+1$, the economy is in steady state, hence $R_{t+1} = (\beta^s)^{-1}$, $P_{t+1} = P_t = \gamma$, and $Y_{t+1} = Y$.

Before doing so, however, we establish the response of the real exchange rate by taking for granted the equilibrium at the zero lower bound described above, and to be established in the following section. To do so, first focus on the developments in Foreign. Suppose first that Foreign wages are fully flexible. This means Foreign remains at full employment from (5.A.11), and $Y_t^* = Y$. Furthermore, we have $R_t = 1$ from (5.A.9), $P_t C_t = \gamma Y_t$ and $P_{t+1} C_{t+1} = \gamma Y$ (where we use $P_t = P_{t+1} = \gamma$ in (5.A.3), and that $Y_{t+1} = Y$). Plugging these values into the 4x4 system of equations in (5.A.17)—(5.A.20) yields an expression for Foreign nominal consumption

$$P_t^* C_t^* \ = \ \gamma Y (1 + (1 - \beta^s) \frac{Y_t}{Y}).$$

Combining this with the market clearing condition in (5.A.7) gives us an expression for the real exchange rate under flexible Foreign prices:

$$Q_t \ = \ \left[\frac{P_{F,t}}{P_t}\right]^{1-\omega} \ = \ \left[1 - \omega + \left(1 - (1-\omega)\beta^s\right)\frac{Y_t}{Y}\right]^{1-\omega} \qquad (5.A.24)$$

We refer the reader to equation (5.4.4) and the proof of Proposition 4 for the precise expression for Y_t.

Suppose now that Foreign wages are downwardly rigid, and furthermore, the rigidity is sufficiently high such that (5.A.11) binds and hence $P_{F,t} = \gamma^*$. Then, trivially,

$$Q_t \ = \ \left[\frac{\gamma^*}{\gamma}\right]^{1-\omega}.$$

We now combine the cases of rigid and flexible wages in Foreign. From (5.A.13), we know that Foreign price level falls to either the flex-price level, or the maximum amount permitted by downward wage rigidity:

$$P_{F,t} \ = \ \max\left(\gamma^*, 1 - \omega + \left(1 - (1-\omega)\beta^s\right)\frac{Y_t}{Y}\right),$$

Putting this expression into the real exchange rate formula in (5.3.10) yields

$$Q_t \ (:= \ Q_t^{\mathrm{ZLB}}) \ = \ \max\left(\left[\frac{\gamma^*}{\gamma}\right]^{1-\omega}, \left[1 - \omega + \left(1 - (1-\omega)\beta^s\right)\frac{Y_t}{Y}\right]^{1-\omega}\right), \quad (5.A.25)$$

which is the formula in Proposition 3(b).

To establish the inequality $Q_t^{\text{ZLB}} < Q_t^{\text{NoZLB}}$, we require that both parts of the max() operator above are smaller in magnitude than Q_t^{NoZLB}. This first of all requires that Foreign wages are not excessively rigid relative to Home:

$$\left[\frac{\gamma^*}{\gamma}\right]^{1-\omega} < Q_t^{\text{NoZLB}},$$

which from (5.A.23) requires

$$\begin{aligned}\gamma^*/\gamma &< 1+\kappa, \text{ where}\\ \kappa &= (1-\omega)\beta^s\left(\frac{(1-\chi)Y+\chi\bar{B}^H}{(1-\chi)Y+\chi\bar{B}^L}-1\right) > 0\end{aligned}$$

Turning to the second part of the max() operator, since $Y_t < Y$ and $1-(1-\omega)\beta^s > 0$, this is bounded above by

$$\left[1-\omega+\left(1-(1-\omega)\beta^s\right)\frac{Y_t}{Y}\right]^{1-\omega} < [2-\omega-(1-\omega)\beta^s]^{1-\omega}.$$

Because $\beta^s\bar{B}^H-\bar{B}^L > \underline{\zeta}$, we know that the counterfactual $R_t < 1$ in (5.A.16), and hence $R_t^{-1} > 1$. Applying this inequality to equation (5.A.22), we get

$$Q_t^{\text{NoZLB}} > [2-\omega-(1-\omega)\beta^s]^{1-\omega},$$

Therefore, the inequality is satisfied for both parts of the max() operator, and $Q_t^{\text{ZLB}} < Q_t^{\text{NoZLB}}$.

To get a real appreciation, we need both parts of the max() operator to be less than 1. For the first part, we simply require $\gamma^*/\gamma < 1$. For the second part, we require

$$\beta^s\bar{B}^H-\bar{B}^L > \frac{(1-\chi)}{\chi}\left(\frac{1-\beta}{1-(1-\omega)\beta}\right)\gamma Y =: \widetilde{\zeta}$$

which is obtained by substituting for Y_t in (5.A.24) (using the formula in (5.4.4), derived in the Proposition 4 proof), and setting the resulting expression to be less than 1. Finally, a real appreciation requires that the deleveraging shock is both large enough to push the union to the zero lower bound ($\beta^s\bar{B}^H-\bar{B}^L > \underline{\zeta}$), and large enough to trigger a real appreciation once at the zero lower bound ($\beta^s\bar{B}^H-\bar{B}^L > \widetilde{\zeta}$). Thus it is required that

$$\beta^s\bar{B}^H-\bar{B}^L > \max(\widetilde{\zeta}, \underline{\zeta}) =: \bar{\zeta}.$$

5.A.4 Proposition 4 proof

Recall that from our definition of monetary policy, Section 5.3.3, monetary policy cuts the nominal interest rate all the way to $R_t = 1$ if due to deflationary pressure, the inflation target $\Pi_t^u = 1$ cannot be reached. Here we show in a first step that for a large enough shock ($\beta^s \bar{B}^H - \bar{B}^L > \underline{\zeta}$), $P_t = P_{t+1} = \gamma$ along with $Y_t < Y$ is such an equilibrium, and in a second step, that it is the only one possible at the ZLB. Finally, we provide a proof of Proposition 4.

First note that as $n = 1$, real wage income and economic activity are directly related (as shown in the main text)

$$\frac{W_t L_t}{P_t} = Y_t.$$

Substituting this and the fact that $P_t = P_{t+1} = \gamma$ into the borrowers' budget constraint in (5.A.2), and the savers' Euler equation in (5.A.1) gives

$$C_t^b = -\frac{\bar{B}^H - \bar{B}^L}{\gamma} + Y_t \tag{5.A.1}$$

$$C_t^s = (\beta^s)^{-1} C_{t+1}^s \tag{5.A.2}$$

Turning to savers' budget constraint (equation 5.A.15) at $t + 1$, and knowing that $\bar{B}_t = \bar{B}^L$ in the new steady state, we get

$$C_{t+1}^s = Y + \frac{\chi}{1-\chi}\left(\frac{(1-\beta^s)\bar{B}^L}{\gamma}\right) \tag{5.A.3}$$

Combining (5.A.3) in (5.A.4) yields

$$Y_t = \chi C_t^s + (1-\chi) C_t^b,$$

and substituting for saver and borrower consumption using equations (5.A.1)—(5.A.3) gives us the expression for output in (5.4.4):

$$Y_t = (\beta^s)^{-1}\left[Y - \frac{\chi}{1-\chi}\left(\frac{\beta \bar{B}^H - \bar{B}^L}{\gamma}\right)\right]. \tag{5.A.4}$$

From this equation, it follows that $Y_t < Y$ whenever

$$\beta^s \bar{B}^H - \bar{B}^L > \gamma \frac{(1-\chi)}{\chi}(1-\beta) Y = \gamma \underline{\zeta},$$

which must always hold, because $\gamma \leq 1$ and $\beta^s \bar{B}^H - \bar{B}^L > \underline{\zeta}$ at the ZLB. Thus we have established that $P_t = P_{t+1} = \gamma$ along with $Y_t < Y$ is an equilibrium.

To see that no other deflationary equilibrium at the ZLB exists, assume that prices fall to some level $1 = P_{t-1} > P_t = P_{t+1} (=: \tilde{P}) > \gamma$. In this case, from slackness condition (5.A.10), output must be at potential in the period of deleveraging, $Y_t = Y$. However as prices fall to $\tilde{P}$ equation (5.4.4), by setting $Y_t = Y$, can be written as

$$\beta^s \bar{B}^H - \bar{B}^L \ = \ \tilde{P}\underline{\zeta}.$$

The fact that at the ZLB, $\beta^s \bar{B}^H - \bar{B}^L > \underline{\zeta}$, then leads to a contradiction because under deflationary pressure, $\tilde{P} < 1$ as mentioned before. Thus, there can be no other equilibrium at the ZLB where due to deflationary pressure, the central bank cuts its interest rate to $R_t = 1$.[28]

We can see that from equation (5.4.4), output Y_t falls with lower γ. This also means that real incomes fall, and from (5.A.1), that borrower consumption falls—since real incomes Y_t are lower and real debt repayments $(\bar{B}^H - \bar{B}^L)/\gamma$, which enter negatively, are higher. This completes the proof of Proposition 4(a). As a side note, from (5.A.2) and (5.A.3) we can see that saver consumption increases slightly in the period of deleveraging with more flexibility (and thus lower prices), because saver consumption becomes higher in the new steady state. This is because the value of saver assets (equal to borrower debt) rises in real terms as prices decline by more. However, this increase in saver consumption comes at the expense of the borrowers (both during the deleveraging period and in the new steady state), and is not enough to offset the fall in borrower consumption during the deleveraging period (because output falls, from equation 5.4.4).

We now turn to part (b) of Proposition 4. If the Foreign wage rigidity constraint is not binding, the real exchange rate is given by expression (5.A.24). Then γ only enters this expression via Y_t. We can see that higher wage flexibility lowers output Y_t, and, since $1 - (1 - \omega)\beta^s > 0$, lowers the real exchange rate Q_t. If $\bar{\zeta} > \beta^s \bar{B}^H - \bar{B}^L > \underline{\zeta}$, $Q_t > 1$ and the real exchange rate depreciates by less as γ declines. If $\beta^s \bar{B}^H - \bar{B}^L > \bar{\bar{\zeta}}$, $Q_t < 1$ and the real exchange rate appreciates by more as γ declines. This completes the proof of the Proposition.

28. Note that an inflationary equilibrium at the ZLB exists, $\tilde{P} > 1$, such that $Y_t = Y$ and $R_t = 1$ in the period of deleveraging. Thus we rule out this equilibrium by maintaining that the central bank would only cut its interest rate to $R_t = 1$ in the case of deflationary pressure in the period of deleveraging. This would be strictly implied if—as we do in our numerical implementation of the model—the central bank *implemented* its strict inflation target via a Taylor-type feedback rule.

References

Benigno, Pierpaolo, Gauti B. Eggertsson, and Federica Romei. 2014. "Dynamic Debt Deleveraging and Optimal Monetary Policy." NBER Working Paper 20556.

Benigno, Pierpaolo, and Federica Romei. 2014. "Debt deleveraging and the exchange rate." *Journal of International Economics* 93 (1): 1–16.

Berka, Martin, Michael B Devereux, and Charles Engel. 2015. "Real Exchange Rates and Sectoral Productivity in the Eurozone." Working Papers 26970. Department of Economics, The University of Auckland.

Broda, Christian. 2004. "Terms of trade and exchange rate regimes in developing countries." *Journal of International Economics* 63 (1): 31–58.

Cochrane, John H. 2015. "The New-Keynesian Liquidity Trap." Mimeo.

Cook, David, and Michael B. Devereux. 2013. "Sharing the Burden: Monetary and Fiscal Responses to a World Liquidity Trap." *American Economic Journal: Macroeconomics* 5 (3): 190–228.

Cook, David, and Michael B. Devereux. 2014. "Exchange rate flexibility under the zero lower bound." Globalization and Monetary Policy Institute Working Paper 198. Federal Reserve Bank of Dallas.

Corsetti, Giancarlo, Keith Kuester, and Gernot J. Müller. 2013. "Floats, pegs and the transmission of fiscal policy." In *Fiscal Policy and Macroeconomic Performance*. Edited by Luis F. Céspedes and Jord Galí. Vol. 17, Central Banking, Analysis, and Economic Policies. Central Bank of Chile. Chapter 7, 235–281.

De Paoli, Bianca. 2009. "Monetary policy and welfare in a small open economy." *Journal of International Economics* 77 (1): 11–22.

Eggertsson, Gauti B., and Paul Krugman. 2012. "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *Quarterly Journal of Economics* 127 (3): 1469–1513.

Eggertsson, Gauti, Andrea Ferrero, and Andrea Raffo. 2014. "Can structural reforms help Europe?" *Journal of Monetary Economics* 61 (C): 2–22.

Engel, Charles. 1999. "Accounting for U.S. Real Exchange Rate Changes." *Journal of Political Economy* 107 (3): 507–538.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica*,

Fornaro, Luca. 2015. "International Debt Deleveraging." CEPR Discussion Paper 10469.

Friedman, Milton. 1953. "Essays in Positive Economics." In. Chapter The case for flexible exchange rates.

Galí, Jordi, and Tommaso Monacelli. 2005. "Monetary Policy and Exchange Rate Volatility in a Small Open Economy." *Review of Economic Studies* 72: 707–734.

Galí, Jordi, and Tommaso Monacelli. 2013. "Understanding the Gains from Wage Flexibility: The Exchange Rate Connection." mimeo.

Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajsek. 2015. "Financial Heterogeneity and Monetary Union." mimeo.

Kollmann, Robert. 2001. "The exchange rate in a dynamic-optimizing business cycle model with nominal rigidities: a quantitative investigation." *Journal of International Economics* 55 (2): 243–262.

Kollmann, Robert, Zeno Enders, and Gernot J. Müller. 2011. "Global banking and international business cycles." *European Economic Review* 55 (3): 407–426.

Martin, Philippe, and Thomas Philippon. 2014. "Inspecting the Mechanism: Leverage and the Great Recession in the Eurozone." NBER Working Paper 20572.

Monacelli, Tommaso. 2004. “Into the Mussa puzzle: monetary policy regimes and the real exchange rate in a small open economy.” *Journal of International Economics* 62 (1): 191–217.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2014. “Recovery from Financial Crises: Evidence from 100 Episodes.” *American Economic Review* 104 (5): 50–55.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2016. “Downward Nominal Wage Rigidity, Currency Pegs, and Involuntary Unemployment.” *Journal of Political Economy* 124 (5): 1466–1514.

Sutherland, Alan. 2005. “Incomplete pass-through and the welfare effects of exchange rate variability.” *Journal of International Economics* 65 (2): 375–399.

Lebenslauf

Dmitry Kuvshinov was born on 31 March 1985 in Novosibirsk, Russia. He studied in gymnasium no. 130 in Novosibirsk between 1992 and 1999, and graduated from the Netherhall School, Cambridge, UK, in September 2001. He studied for his A-levels at the Hills Road Sixth Form College in Cambridge, UK, during 2002–2003, and completed a BA Hons in Economics and Management at Balliol College, Oxford University, in September 2007. He worked for the Bank of England between 2007 and 2013, taking a year's study leave at Warwick University to complete the MSc in Economics and International Financial Economics in September 2011. He started his PhD at the Bonn Graduate School of Economics in September 2013.