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

Niklas Kaspar Zimmermann

aus Bergisch Gladbach, Deutschland

Bonn, 2021

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Moritz Schularick
Zweitreferent:	Prof. Dr. Farzad Saidi
Tag der mündlichen Prüfung:	11. Juni 2021

Acknowledgements

This dissertation would not have been possible without the support I received from many people. Special thanks go to my supervisor, Moritz Schularick, for his guidance, feedback and support. I am deeply grateful to him for providing the environment, freedom, and resources that I needed to conduct my research. I would also like to thank Farzad Saidi, my second reviewer, and the other members of the Bonn faculty – in particular Christian Bayer, Tobias Berg, Benjamin Born, Narly Dwarkasing, Thomas Hintermaier, Keith Kuester, Moritz Kuhn and Isabel Schnabel – for their advice and feedback at various stages of my dissertation.

I consider myself very fortunate to have collaborated with two excellent coauthors. I am grateful to Dmitry Kuvshinov and Björn Richter for countless discussions that shaped this dissertation. Dmitrys enthusiasm and energy and Björns sharp analysis and advice made this journey truly enjoyable. Apart from my coauthors and advisors, I am also thankful to my fellow graduate students and colleagues, who supported me in writing this dissertation. This includes Thilo Albers, Branislav Albert, Francisco Amaral, Luis Bauluz, Alina Bartscher, Yao Chen, Martin Dohmen, Ricardo Duque Gabriel, Katharina Knoll, Sebastian Kohl, Filip Novokhmet, Andrea Papadia, Lucas Ter-Steege, Yevhenii Usenko, Felix Ward and Dominik Wehr among others.

This dissertation originated within the framework of the Bonn Graduate School of Economics and the Institute for Macroeconomics and Econometrics. I wish to thank Heide Baumung and Beate Roters for helping me deal with countless administrative matters. I benefited from the infrastructure and financial support of the Bonn Graduate School of Economics, the Federal Ministry of Education and Research (BMBF) and the Deutsche Forschungsgemeinschaft (DFG). My dissertation has also greatly profited from research visits at the Paris School of Economics through the European Doctoral Programme and Yale University.

Finally, I would like to thank my friends Zvonimir Bašić, Jonas Radbruch, Felix Schran and Louis Strang and many other PhD colleagues at the BGSE, who made the PhD experience a pleasant one. A special thanks goes to Katinka, who has been a wonderful sparring partner for new ideas and a source of support and encouragement.

Contents

List of Figures

List of Tables

Introduction

This thesis provides a comparative perspective on current macro-financial phenomena. Understanding the causes and consequences of rare macroeconomic events and secular trends is inherently difficult. Competing theories and explanations are difficult to evaluate in a macroeconomic context as the potential for experimentation is limited and truly exogenous variation is hard to come by. In addition, macro-financial disasters are rare events, occurring on average less than once in a generation. This means that the time horizons to study them have to be long. Starting with the well-known work *This Time Is Different* by Reinhart and Rogoff (2009), long-run perspectives have come to play a central role in the macro-financial literature. A comparative perspective is also a natural starting point for the analysis of macro-financial trends. It tells us how unusual an observation is from a historical standpoint and helps to tell different explanations for the observed patterns apart.

For this purpose, the research projects included in this thesis introduce new long-run cross-country data to study some of the most pertinent questions in macro-financial research. The new data allow us to put the recent growth of listed equity markets into a long-run perspective (Chapter 1), study the joint evolution of safe and risky rates over the last 150 years (Chapter 2) and learn more about the causes and consequences of systemic banking crises and government bankrutpcies (Chapter 3 and 4).

Chapter 1, titled "The Big Bang: Stock Market Capitalization in the Long Run", written jointly with Dmitry Kuvshinov studies trends in stock market capitalization and their drivers. In the US, the past few decades have seen an increase in stock market capitalization and corporate profits despite slowing economic activity (Fernald, 2015; Greenwald, Lettau, and Ludvigson, 2019; Barkai, 2020). In this chapter we put these recent US trends into a broader cross-country perspective. For this purpose, we introduce new annual data on stock market capitalization in 17 advanced economies between 1870 and 2016. We use these data to trace out the long run evolution of stock market size and disentangle its drivers. After all, without long-run cross-country data one cannot tell whether these recent developments are simply a country-specific short-run deviation, or are part of a much broader shift in the relationship between the stock market and the economy.

We find that during the first century of our data, the ratio of market capitalization to GDP was constant at roughly one-third. But after the 1980s, a sharp structural break took place, with the market cap to GDP ratio tripling to 100% and remaining at this high level thereafter. Exploring the reasons behind this long-run disconnect, we find that the structural increase in market size reflects a sharp and persistent stock price boom. We show that these rising equity prices are driven by a profit shift away from other sectors of the economy towards listed firms, supported by historically low equity discount rates. The share of listed firms' profits in both GDP and capital income has more than doubled since 1990, reaching its highest levels in the last 145 years. Our findings show that the recent corporate profit boom (De Loecker, Eeckhout, and Unger, 2020) is a historically unparalleled global phenomenon.

Chapter 2 – "The Expected Return on Risky Assets: International Long-run Evidence", joint work with Dmitry Kuvshinov – investigates the long-run evolution of the expected return on risky assets and its relationship with the safe rate. We estimate expected returns on two major risky asset classes across 17 countries and 145 years and find that the expected risky return has been declining steadily throughout the last 145 years. But this decline is largely unrelated to movements in the real safe rate and, as a consequence, the risk premium exhibits large secular variation and is strongly negatively correlated with the safe rate.

This risky-safe rate disconnect carries important implications for asset pricing theory. Standard theory puts forward two key drivers of expected returns: growth and risk. These entail opposing predictions for the relationship between risky and safe rates. While the growth channel pushes risky and safe rates in the same direction by affecting the general willingness to save, the risk channel pushes safe rates and risk premia in opposite directions with ambiguous overall effect on the risky rate. Our finding that risk premia and safe rates are strongly negatively correlated suggests that risk, rather than growth, is the key driver of expected risky and safe returns over the long run. Consistent with this view, we show that secular movements in the risk premium can be explained by changes in macroeconomic risk in line with theory (Lettau, Ludvigson, and Wachter, 2008).

Chapter 3 – "The Profit-Credit Cycle", joint work with Björn Richter – studies the underlying drivers of boom-bust cycles in credit markets. Credit supply based explanations of *credit booms gone bust* are increasingly gaining traction in macro-financial research. Empirical evidence points to a general pattern where buoyant conditions in credit markets increase the quantity of credit (Schularick and Taylor, 2012; Jordà, Schularick, and Taylor, 2013; Mian, Sufi, and Verner, 2017) at low costs (Greenwood and Hanson, 2013; Krishnamurthy and Muir, 2017) but are followed by hightened crises risks and severe economic downturns. Yet, the ultimate sources of these credit supply expansions and how they turn into a crisis is still limited (Mian and Sufi, 2018).

In this chapter we revisit the origins and turning points of the credit cycle. We show that what we observe in the data is in fact a "profit-credit cycle". An increase in profitability of the banking sector predicts rising credit-to-GDP ratios and the start of a credit boom in a long-run panel of advanced economies. This relationship remains robust when we include additional controls, time effects and analyze subsamples. It holds for alternative measures of profitability or credit growth, during and outside of financial crises and on a country-by-country level. To study the channels behind this "profit-credit cycle" in more detail we decompose bank profitability into its sources and uses. We find that decreasing loan losses are associated with expanding credit, while lower costs or increasing revenues are not. Our analysis supports a supply side view of the credit cycle with an important role for time-varying beliefs. In line with behavioral credit cycle models, we show that increases in bank profitability are associated with a higher incidence of banking crises, in particular panics, in the medium term.

The final chapter 4 – "Sovereigns Going Bust: Estimating the Cost of Default", joint work with Dmitry Kuvshinov published in the *European Economic Review* – studies the macroeconomic costs of sovereign bankruptcy. 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 with existing empirical studies placing the 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). Our analysis contributes to the literature by providing a conclusive best-practice estimate of the macroeconomic cost of default which relies on up-to-date comprehensive methods and data. We combine local projections and propensity score weighting as in Jordà and Taylor (2016) and study default costs in a large dataset spanning 1970-2010, 112 countries and 92 external defaults.

We find that default generates a long-lasting output cost of 2.7% of GDP on impact and 3.7% at peak after five years – but in the longer term, economic activity recovers. The downturn is characterised by a collapse in investment and gross trade. Our second key contribution is to assess which factors are empirically more important in amplifying the cost of default. We show that the cost rises dramatically if the default is followed by a systemic banking crisis — peaking at some 9.5% of GDP -— but is attenuated for economies with floating exchange rates. Our findings suggest that financial autarky, trade frictions and sovereign-banking spillovers play a key role in generating the cost of default.

References

Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75 (5): 2421–2463.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135 (2): 561–644.

De Paoli, Bianca, Glenn Hoggarth, and Victorical Saporta. 2009. "Output Costs of Sovereign Crises: Some Empirical Estimates." Bank of England Working Paper 362.

Fernald, John G. 2015. "Productivity and Potential Output before, during, and after the Great Recession." *NBER Macroeconomics Annual* 29: 1–51.

Furceri, Davide, and Aleksandra Zdzienicka. 2012. "How Costly Are Debt Crises?" *Journal of International Money and Finance* 31 (4): 726–742.

Greenwald, Daniel L., Martin Lettau, and Sydney C. Ludvigson. 2019. "How the Wealth Was Won: Factors Shares as Market Fundamentals." NBER Working Paper 25769.

Greenwood, Robin, and Samuel G. Hanson. 2013. "Issuer Quality and Corporate Bond Returns." *Review of Financial Studies* 26 (6): 1483–1525.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. "When Credit Bites Back." *Journal of Money, Credit and Banking* 45 (s2): 3–28.

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.

Krishnamurthy, Arvind, and Tyler Muir. 2017. "How Credit Cycles Across a Financial Crisis." NBER Working Paper 23850.

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.

Levy-Yeyati, Eduardo, and Ugo Panizza. 2011. "The Elusive Costs of Sovereign Defaults." *Journal of Development Economics* 94 (1): 95–105.

Mian, Atif, and Amir Sufi. 2018. "Finance and Business Cycles: the Credit-Driven Household Demand Channel." *Journal of Economic Perspectives* 32 (3): 31–58.

Mian, Atif R., Amir Sufi, and Emil Verner. 2017. "Household Debt and Business Cycles Worldwide." *Quarterly Journal of Economics* 132 (4): 1755–1817.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly.* Princeton University Press.

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): 1029–1061.

Chapter 1

The Big Bang: Stock Market Capitalization in the Long Run*

Joint with Dmitry Kuvshinov

1.1 Introduction

Short-run stock market movements are often disconnected from economic fundamentals (Shiller, 1981). As a consequence, much research in finance has studied the nature of these short-run deviations and their underlying drivers (Cochrane, 2011). Yet the focus on the short run has left much of the bigger long-run picture underexplored. In the US, the past few decades have seen an increase in stock market capitalization and corporate profits despite slowing economic activity (Greenwald, Lettau, and Ludvigson, 2019; Barkai, 2020). Without long-run cross-country data, however, one cannot tell whether these recent developments are simply a country-specific short-run deviation, or are part of a broader shift in the relationship between the stock market and the economy. After all, Rajan and Zingales (2003) argued that for a broad cross-section of economies, post-1980 increases in stock market capitalization were merely a reversal to previously high levels observed in the early 20th century. And when it comes to profits, there is some evidence that the recent increases may be specific to the US (Gutiérrez and Piton, 2020) and may simply represent reversals to a long-run mean following profit declines in the 1960s and 70s (Nordhaus, 1974; Feldstein and Summers, 1977; Barkai and Benzell, 2018).

* 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 Michael Bordo, Jan Eeckhout, Leslie Hannah, Yueran Ma, Björn Richter, Kristian Rydqvist, Moritz Schularick, Nikolaus Wolf, the participants of seminars in Bonn, Berlin School of Economics, CREI, UPF, the 7th CEPR economic history symposium, Paris EHES congress, Oxford FRESH-UPIER meeting, 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.

This paper studies the relationship between the stock market and real activity over the long-run. For this purpose, we introduce a new annual cross-country dataset on stock market capitalization going back to 1870. These data 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. They allow us to re-assess the Rajan and Zingales (2003) "great reversals" hypothesis, and put the recent US corporate profit boom into a broader perspective. Our data go far beyond currently available series which only cover benchmark years for a cross-section of countries, or selected historical periods for some individual countries. These data are complemented by estimates of listed firms dividends and earnings, again stretching back across countries and time. Together with the extensive documentation in the Data Appendix, they provide a new resource for researchers to study the development of the stock market, corporate profitability and capital structure throughout the last century and a half.

During the first 120 years of our data, the ratio of market capitalization to GDP was constant at roughly one-third. The listed profit share was also stable, with dividends paid by listed firms accounting for about 1.3% of GDP. But after the 1980s, a sharp structural break took place. The ratio of market capitalization to GDP tripled to 100%, and the ratio of listed firms' earnings to GDP doubled. We show that these increases had nothing to do with quantities: equity issuance and new listings make no contribution to the increases in market cap. Instead, they reflect a *profit shift* within capital, with the share of listed firms' earnings in net capital income increasing from 12% in 1990 to 21% in 2015. The impact of this profit shift on capitalization was particularly large because the equity discount rate was at historically low levels. Our findings show that the recent US corporate profit boom is a historically unparalleled global phenomenon contributing to the increases in wealth across advanced economies documented by Piketty and Zucman (2014).

We start by analysing long-run trends in market capitalization, and show that the increase in stock market cap during the 1980s and 1990s was both highly unusual and reflected a common trend across the 17 economies in our sample. The cross-country average market cap to GDP ratio was constant between 1870 and the 1980s, tripled in the 1990s and remained high after 2000. This "big bang" increase took place in each of the 17 countries, and in all countries bar one (Belgium), it represents the single largest structural break in the market cap to GDP ratio during the entire 146-year historical time period. The long-run trends also underscore a rise in the global importance of the US equity market at the expense of the UK and France: while all three countries enjoyed roughly equal global market shares in the early 1900s, by the 1960s the US stock market accounted for close to 70% of the total 17-country capitalization.

When we decompose the "big bang" in market capitalization into price and quantity changes, it turns out that this entire increase was driven by higher stock prices, and none by higher issuances or new listings. Net issuance is sizeable at roughly 4%

of market cap, but is constant over time and even declines slightly after 2000. Real capital gains on equity were, on the contrary, close to zero before the 1980s but persistently high afterwards, rising to an average of 3.6% per year after 1985. We run two counterfactual scenarios, shutting off, in turn, the price and the quantity channel after 1985. If we shut off the price channel, market capitalization stagnates at its historical level of one-third of GDP. But if we shut off the quantity channel, observed stock price movements are still able to explain the entire post-1985 increase in market capitalization.

To map out the deeper underlying drivers behind this increase in stock prices and capitalization, we use the standard dynamic Gordon growth model to decompose the market cap to GDP ratio into three components: the current ratio of listed firms' dividends to GDP (the *profit share* channel), *future growth* of dividends or earnings, and the rate at which these future cashflows to shareholders are discounted (the *discount rate* channel). The drivers of the "big bang" increase in market cap should display two characteristics: first, their short-run movements should be correlated with market capitalization, showing that their variation can potentially affect market cap. Second, their long-run trend should display changes which can explain the increases in equity wealth observed in the 1980s and 1990s. We therefore tease out the relative contribution of these factors in two steps. First, we run predictive regressions to test if market capitalization is correlated with any of these three factors in the right direction. Second, we look at long-run trends in those factors which are correlated with market cap, and perform a counterfactual analysis of the changes in market cap induced by these trends.

Our analysis shows that when it comes to explaining the long-run increases in the market cap to GDP ratio, the "profit share" channel is key, the "discount rate" channel – important, and the "future growth" channel – irrelevant. Predictive regressions show that high market capitalization is correlated with high current profit share and low future equity returns, but with low, rather than high, future dividend growth. This suggests that changes in expected future profitability play little, if any, role in the dynamics of the market cap to GDP ratio.

Turning to trends in the profit share, we show that dividends and earnings of listed firms have both increased markedly as share of GDP, roughly tripling in size between 1985 and 2015. Like the big bang in stock market capitalization, these increases are historically unprecedented and part of a broad cross-country trend. Furthermore, they go above and beyond the recently documented declines in the labour shares and increases in capital income as share of GDP (Karabarbounis and Neiman, 2013). We show that the ratio of listed firms earnings to capital income has also more than doubled since 1990, and that those countries which experienced the largest capitalization increases during the big bang also recorded larger increases in the earnings-to-capital-income ratio. Changes in capital shares, on the contrary, cannot explain the cross-country differences in market capitalization growth during the big bang and show little correspondence with market cap in the historical data.

Altogether, our findings point to the redistribution of earnings within the corporate sector – for example due to increasing market power of large firms (De Loecker and Eeckhout, 2018; De Loecker, Eeckhout, and Unger, 2020) – as the key force driving the recent increases in stock market wealth.

To map these profit figures into stock prices and capitalization, we need an estimate of the equity discount rate. For this purpose we use the estimates in Kuvshinov and Zimmermann (2020), computed as the sum of the dividend-price ratio and the expected future cashflow growth rate obtained from predictive regressions. We show that the equity discount rate declined sharply shortly before the big bang and remained at historically low levels throughout the past three decades. While the initial decline in the 1970s and 1980s was driven by a low equity risk premium, the last two decades saw the falling safe rate become increasingly important. Taking stock of the underlying drivers, our counterfactual analysis shows that the "profit shift" channel is key. It accounts for 76% of the increase in market cap between the 1980s and 2015, with lower discount rates accounting for another 10%. Taken together, these two factors can account for the entire increase in the market cap to GDP ratio in the data.

The 1990s increases in market capitalization were both historically unusual and highly persistent. The historical data show that on average, a sharp run-up in market cap is more likely to be followed by a stock market crash than a structural break to a higher mean. In the last part of the paper, we explore the role of market capitalization in timing and forecasting short-run market movements and crashes. We show that market capitalization outperforms the standard price-dividend ratio metric as a return predictor, and that high capitalization predicts low rather than high growth in future dividends. The reason that market capitalization does so well is that it uses information on GDP rather than dividends to track fundamentals – enabling it to look through changes in payout policy – and that it contains issuance data which can help capture swings in investor sentiment (Baker and Wurgler, 2000). Our evidence supports Warren Buffet's view that the stock market cap to GDP ratio is "the best single measure of where valuations stand at any given moment" (Buffett and Loomis, 2001). It also means that even though the current high levels of market capitalization reflect a favourable past profit shift towards listed corporations, they do not necessarily indicate favourable future market prospects for equity investors.

Our work is related to three strands of existing literature. The first strand seeks to quantitatively document the long-run evolution of wealth in general and stock market wealth in particular. Rajan and Zingales (2003) provide cross-country estimates of stock market capitalization at selected benchmark years, and Piketty and Zucman (2014) produce annual estimates of aggregate wealth for 8 countries. But efforts to document the evolution of stock market size date back at least to the 19th century, often through surveys commissioned by wealthy financiers (Burdett, 1882; Green, 1887), and later on economic historians undertaking increasingly systematic efforts to map out trends in household wealth and its components (Hoffmann, 1965; Roe, 1971; Goldsmith, 1985). 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 of the literature seeks to understand what drives the changes in total and equity market wealth. When it comes to long-run growth, existing studies focus on the importance of quantities in the form of listings and market access (Rajan and Zingales, 2003) and savings rates (Piketty and Zucman, 2014). Studies of the recent increase in market capitalization in the US, however, focus on price rather than quantity changes. De Loecker, Eeckhout, and Unger (2020) attribute these price increases to changes in market power and McGrattan and Prescott (2005) – to the corporate tax code. In a related contemporaneous paper, Greenwald, Lettau, and Ludvigson (2019) use a model to show that most of the high post-1989 stock market growth in the US can be accounted for by a factor share shift from labour to capital income. Our paper shows that the post-1980s divergence between market capitalization and GDP is in fact a persistent long-run shift that took place not only in the US, but in nearly every advanced economy. We further show that the key driver of this shift is a redistribution of profits within capital income rather than from labour to capital.

The final strand of the literature studies the drivers of short-run volatility in stock markets and the associated empirical predictability relationships. While the dividend-price ratio is found to be a reliable and robust predictor of equity returns in the US (Cochrane, 2008) and for long-run international data (Engsted and Pedersen, 2010; Kuvshinov, 2020), wealth-based valuation ratios such as that between consumption and total (equity, housing, and human) wealth (Lettau and Ludvigson, 2002) have also been shown to predict future stock returns. We show that a relatively simple measure of wealth to income – stock market capitalization to GDP – is a highly reliable predictor of future returns, risk premia and market crashes, not only for the US but in a broad cross-section of advanced economies.

1.2 New long-run data on 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 2016. 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 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, 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 1.B.1–1.B.17 and Figures 1.B.1–1.B.17 detail the sources used for each country, and compare our estimates to others in the existing literature.

The main challenge in constructing stock market capitalization time series 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 constructing our own benchmark year estimates where necessary, 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 1.B.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.

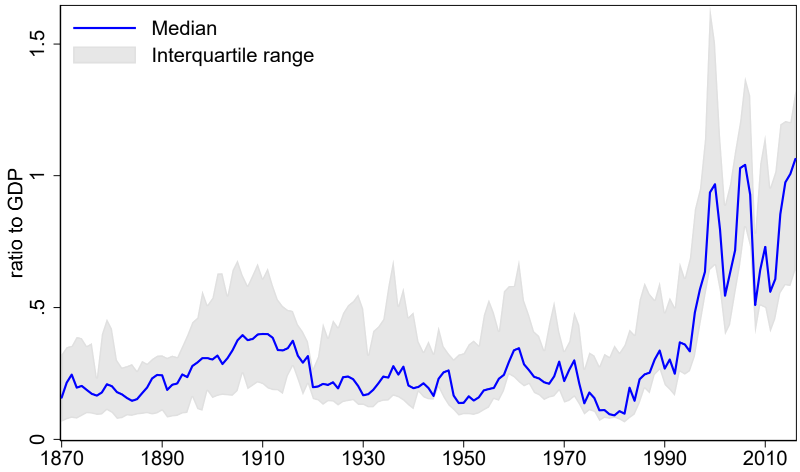
The third challenge relates to excluding foreign stocks. For most of our estimates, the foreign stock share is either well measured (e.g. in recent data) or small (as for most of the mid-20th century data), so the measurement issues mainly concern the large international stock exchanges in the early 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 largely reflects the inclusion of curb and regional exchanges in the US data, and exclusions of bonds and foreign shares for the UK (see Sylla, 2006, for a further discussion of these issues).

Finally, we complement our data on stock market capitalization with estimates of listed firms' profits and dividends which enable us to ultimately link them to the profitability of the listed corporate sector. We compute a long-run series of dividends paid by listed firms as the market capitalization times the dividend-price ratio from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019), with an additional new series for Canada. Since variation in payout ratios and means of compensating shareholders makes dividends an imperfect measure of underlying profit fundamentals (Grullon and Michaely, 2002), we complement these dividend data with estimates of listed company earnings obtained from Compustat Global and Compustat North America going back to the 1980s. The coverage of Compustat firms broadly matches that of our data, but for some of the early observations we drop country-years with insufficient data (less than 30% of total market cap) and scale the other observations by the ratio of Compustat capitalization to our aggregate capitalization estimates.

Taken together, our data provide an improved estimate of historical market capitalization that has far greater country-year coverage than existing sources. This dataset can further be linked to data on listed firms earnings, dividends and stock prices, which allow us to decompose variation in market capitalization, first, into price and quantity movements, and second, into changes in listed firms profit share, equity discount rates, and future cashflows.

Figure 1.1. 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.

1.3 The Big Bang

Figure 1.1 shows the ratio of stock market capitalization to GDP across the 17 economies in our sample between 1870 to today. The solid blue 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 until the late 1980s, the size of the stock market evolved in line with GDP, with the market cap to GDP ratio relatively stable at around one-third. This was true both across time – with the median ratio always below 0.5 – and across countries, with the interquartile range oscillating between 0.1 and 0.6 of GDP. Over the short run, the markets and the economy frequently diverged: the boom of the early 1900s saw the cap to GDP ratio roughly double, only to be undone by the turbulence of World War 1, while the modest World War 2 decline was followed by a run-up in the 1950s and another downturn during the stagflation of the 1970s. But during this first century of our data, stock market size always eventually returned to its long-run average level of one-third of GDP.

Over the last several decades however, an unprecedented market expansion took place, with market capitalization increasing markedly from the 1980s onwards. 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 post-1980s market expansion and the associated shift in the relationship between stock market size and GDP as "the big bang". The left-hand panel of Appendix Figure 1.A.1 shows that this time series pattern – a century of stability followed by a sudden and persistent increase – holds regardless of how we aggregate the individual country data: equally-weighted and GDP-weighted market cap to GDP ratio averages display the same trend as the median in Figure 1.1.

This "big bang" increase can be identified more formally using a statistical test. Figure 1.2 displays the country-level market capitalization series alongside structural break dates estimated using the Bai and Perron (2003) method and the associated break-adjusted means. The red line picks up the largest structural break in the data, while the blue line allows for multiple structural breaks. The country-level data show that the big bang is a truly global phenomenon. In every single country in our sample, the stock market grew rapidly during the 1980s and 1990s, and sustained this high level of capitalization thereafter. In all countries bar one (Belgium), this market capitalization increase is identified as the most important structural break in the entire series. When we allow for multiple structural breaks, nothing comparable comes up. For almost every country in our sample, these recent levels of market capitalization represent an all-time historical high. Based on this, we can conclude that the 1990s saw a truly unprecedented market expansion and a long-run stock market – GDP disconnect happening on a global scale.

Even though this market expansion took place in every country, it was not of equal size everywhere. While market cap in Switzerland and Finland increased almost sixfold, the increase in Belgium was relatively modest and increases in France and UK took market capitalization to previously seen historical peaks rather than an all-time historical high point. The case of Portugal is also rather unique since the 1980s market expansion mainly represented the re-emergence of the stock market after its near disappearnce in the aftermath of the Carnation Revolution of 1974. This means that when we look at the distribution of global market capitalization across countries, we may see some changes – and this exactly what our data show.

Figure 1.3 shows the share of each country's stock market in the global total of our 17 countries. It reports separate shares for the US, UK, France, Germany and Japan and lumps the other 12 countries together. In 1880 capital markets were roughly equally 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. The 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 advanced-economy market cap in 1950.

Even though the US stock market has lost some of its global importance over recent decades, its size is still comparable to that of the other 16 countries 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

Figure 1.2. Stock market cap to GDP ratio in individual countries

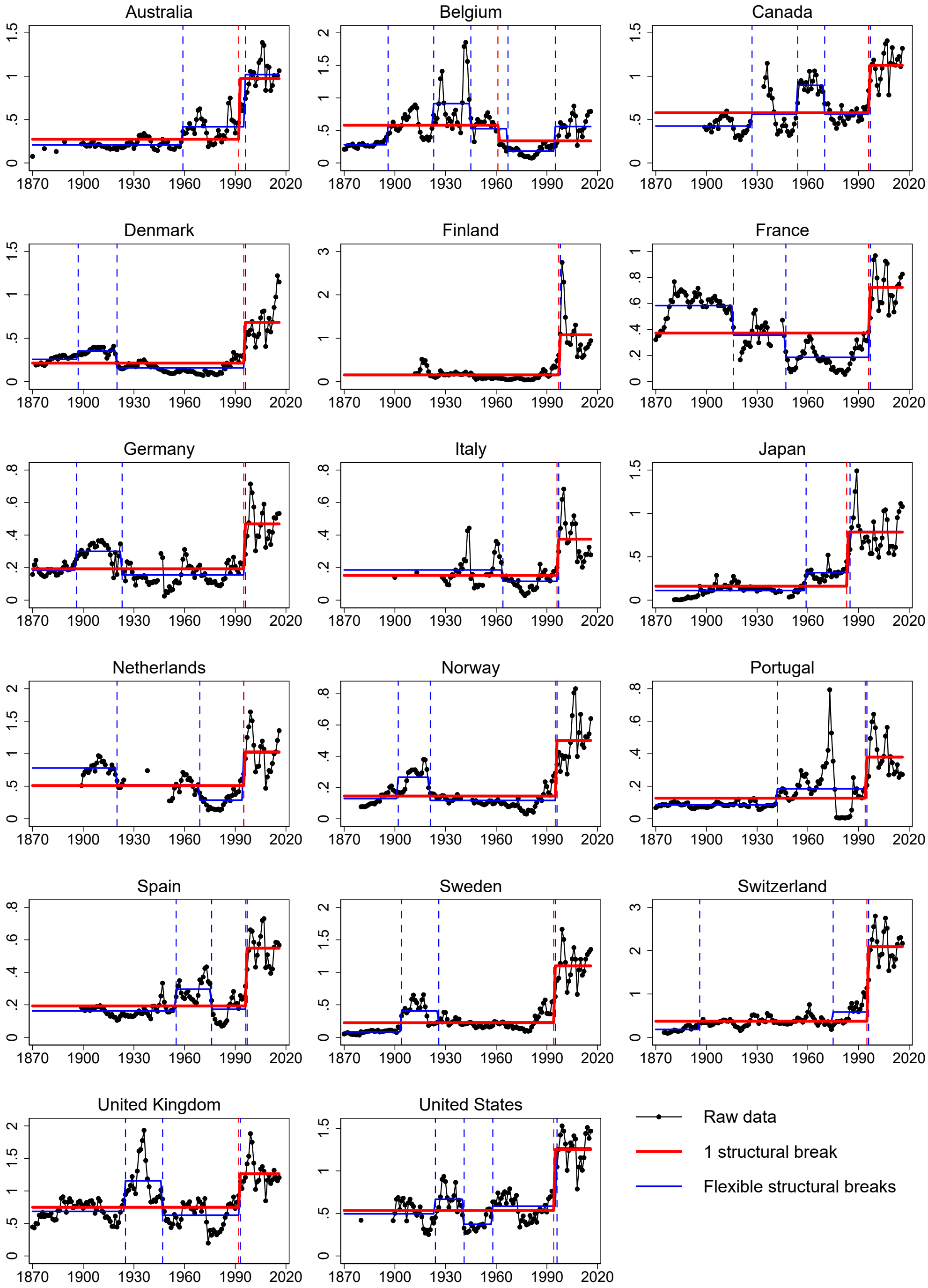
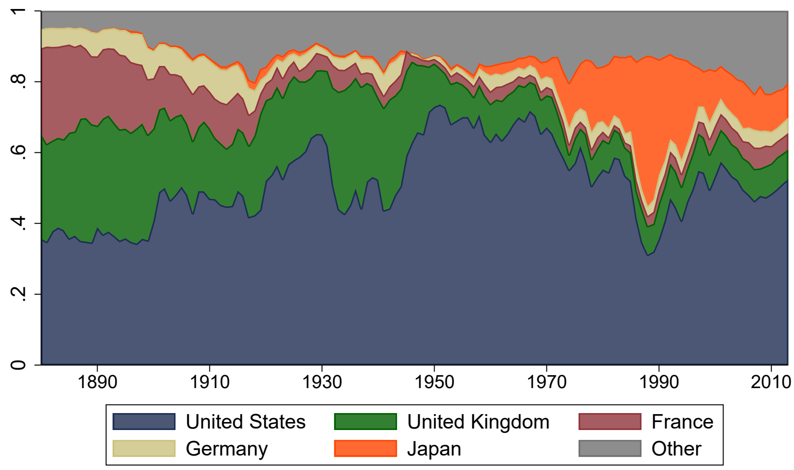


Figure 1.3. Global market capitalization shares across advanced economies



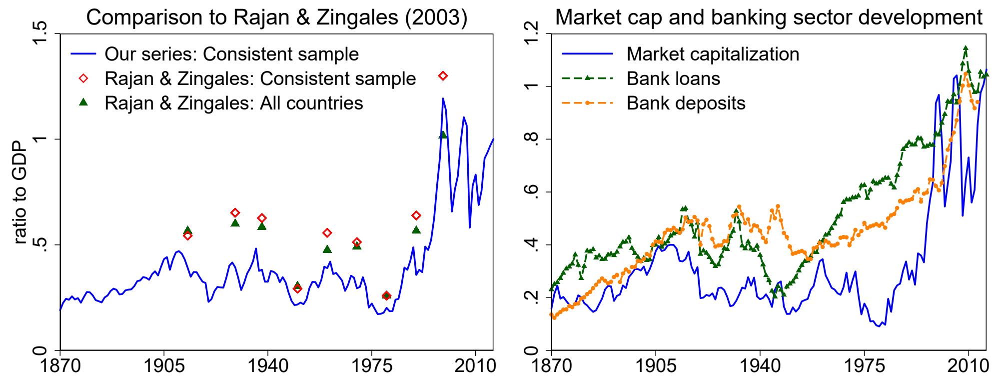
Notes: Shares of individual countries' capitalization in advanced-economy total. Capitalization shares are computed by transforming domestic stock market capitalization into US dollars using historical exchange rates and dividing it by the sum of capitalizations of all 17 countries. Shares of the United States, the United Kingdom, France, Germany and Japan are shown separately. All other countries are combined together into one joint item.

World War 2 and even temporarily catching up to the US at the peak of the Japanese stock market bubble before a dramatic collapse. Capitalization of Japanese listed companies grew from 5% of the global market in 1970 to 40% in 1989, but fell back to around 10% thereafter.

In comparison to existing literature, this "big bang" time series pattern differs somewhat from the U-shaped "great reversals" trend documented by Rajan and Zingales (2003) (henceforth RZ). The left-hand panel of Figure 1.4 compares our estimates to the benchmark-year series 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 blue 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). As shown by the red-diamond figures, the differences between our estimates and RZ are not driven by the sample of countries used. Some of the difference is attributable to improved data quality, with our data utilising new hand-collected sources and additional data assembled by economic historians and statistical agencies (for example, López, Carreras, and Tafunell, 2005; Annaert, Buelens, and De Ceuster, 2012; Waldenström, 2014). A more detailed discussion of data sources and differences to existing estimates for each country can be found in the Data Appendix.

The main reason that, up to this point, the big bang pattern has been somewhat difficult to detect, 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 choice of the benchmark year thus has a significant influence on long-

Figure 1.4. Comparison to alternative estimates and other financial indicators



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.

run market cap comparisons. Using benchmark year estimates means that a short-run temporary disconnect is easily interpreted as a long-run trend with, for example, part of the great reversals pattern driven by the temporary stock market booms of the 1910s and 1990s. Adding annual estimates and extending the horizon by the 17 years of data beyond 1999 helps establish that the increase in market capitalization in the 1980s and 1990s was a unique and persistent increase rather than a short-lived equity boom.

Is the increase in market capitalization linked to other measures of financial development or countries' legal characteristics? The right-hand side of Figure 1.4 plots two proxies of financial development – the total credit to the non-financial sector (green triangles) and total bank deposits (brown crosses, sourced from 2021) alongside stock market size. While all three indicators have grown relative to GDP over the long run, the development of the banking sector follows a very different timing to the stock market expansion: bank loans and deposits start increasing much earlier, do so at a much slower pace, and show the most pronounced increases during the early 2000s – a time when the market cap to GDP ratio had already increased two its new level.

To explore the connection between broader institutional characteristics and market size, the right-hand panel of Appendix Figure 1.A.1 displays the evolution of stock market capitalization separately for countries with common law legal systems (Britain, Canada, US and Australia) and those operating under civil law (the rest of our sample). 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. Consistent with the legal origins hypothesis, common law countries had a persistently higher market cap to GDP ratio. Nevertheless, the post-1980 increase in stock market cap "takes off" at a similar time and is similar in magnitude across both country groups.

In sum, after a century of relative stability equity markets in advanced economies expanded quickly in the 1980s and 1990s. This stock market expansion took place across all countries in our sample pretty much regardless of their institutional characteristics. To understand the drivers of this market expansion, we first decompose market capitalization movements into price and quantity changes, and see which one of these made a more important contribution to the big bang.

1.4 Prices versus Quantities

This section decomposes changes in stock market capitalization into movements in prices and quantities. Much previous research has focused on changes in the quantity of listed equity – in turn driven by deeper institutional characteristics and firms' market access (Rajan and Zingales, 2003) – as the key driver of long-run divergence between stock market size and GDP. However, we show in this section that the entire post-1980s disconnect between stock market capitalization and GDP can be explained by rising equity prices.

To fix ideas, note that total market capitalization *MCAP* at time t is the sum of market capitalizations of each individual share i, in turn calculated as the share price P_i times the quantity of shares issued Q_i:

$$MCAP_t = \sum_{i=1}^{N_t} P_{i,t} Q_{i,t} \tag{1.1}$$

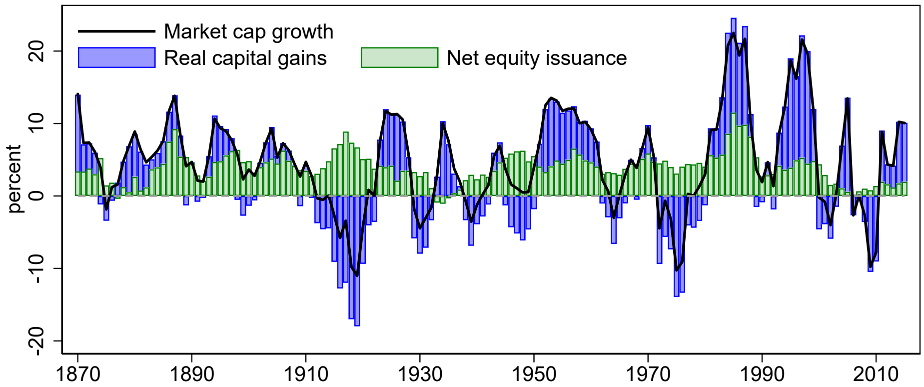
An increase in market capitalization can then come about from share issuance by listed companies (higher Q), new companies entering the listing (higher N), or higher prices of existing listings P. Put differently, aggregate market capitalization is the sum of last year's capitalization $MCAP_{t-1}$ times the capital gain during the year, and the net equity issuance consisting of IPOs and SPOs net of funds paid to retire equity:

$$MCAP_t = (1 + Capital\ gain_t) * MCAP_{t-1} + Net\ issuance_t \tag{1.2}$$

Dividing both sides by $MCAP_{t-1}$ and deflating by CPI, real market capitalization growth g^{MCAP} is the sum of the real equity capital gain cg_t and net equity issuance relative to the previous year's market cap iss_t:[1]

1. Note that in equation (1.3), the market cap growth rate g and capital gain cg can be either nominal or real, but for ease of interpretation across periods we perform the decomposition with real growth rates.

Figure 1.5. Market capitalization growth decomposition



Notes: Decomposition of real market cap growth – the year-on-year growth in nominal market cap less inflation – into real capital gains and net issuance relative to previous year's market cap. Centered 5-year moving averages. Unweighted averages of 17 countries.

$$g_t^{MCAP} = cg_t + iss_t, \tag{1.3}$$
$$\text{where } g_t^{MCAP} = \frac{MCAP_t * CPI_{t-1}}{MCAP_{t-1} * CPI_t} - 1 \text{ and } iss_t = \frac{Net\ issuance_t}{MCAP_{t-1}}$$

The decomposition in equation (1.3) is a pure accounting exercise and does not rely on any assumptions about the underlying sources of stock market wealth. The different components can be estimated directly from the data: we observe market cap growth g and capital gains cg, and calculate net issuance as the residual. This decomposition mirrors the technique used by Piketty and Zucman (2014), among others, to decompose changes in wealth-to-income ratios into savings and capital gains. The main difference to us is that Piketty and Zucman (2014) observe savings and impute capital gains, whereas we observe capital gains and impute net issuance.

Figure 1.5 shows the decomposition of market cap growth (black solid line) into prices (blue bars) and quantity changes (green bars) following equation (1.3). All variables are smoothed using five-year moving averages of annual country data. Unsurprisingly, at shorter horizons price movements are the main driver of changes in stock market wealth. Net issuance is sizeable, but stays more or less constant over time. Furthermore, these baseline estimates, if anything, understate the stability of net issuance and hence the importance of capital gains in driving the cyclical variation in stock market cap. Because we estimate net issuance as a residual from equation (1.3), part of the variation in quantities is driven by measurement error, which increases the volatility of the implied issuance series. To guard against this, we also calculate the decomposition using data on actual net issuance of listed equities, available over the long run for four countries: Finland, Germany, Switzerland and the USA. Appendix Figure 1.A.3 shows the year-by-year decomposition of market cap growth using actual net issuance. Actual issuance is even more stable than implied issuance over the long run, with part of the variation in implied issuance in Figure 1.5 driven by measurement error.

But somewhat unexpectedly, prices remain key even when it comes to long-run stock market growth. Table 1.1 compares the growth in market capitalization and its components to GDP growth over the full sample and three distinct time periods. These correspond to the initial market expansion (1870–1914) during which capitalization grew modestly by about 0.6% of GDP per year, the mid-20th century stagnation (1915–1985) during which market cap to GDP growth was around zero, and the big bang (1985–2015) during which capitalization grew by, on average, more than 2% of GDP per year – about four times the historical norm. To see whether these differences in market cap growth across these long historical periods are driven by prices or quantities, the middle rows of Table 1.1 decompose market cap growth into issuance and capital gains according to equation (1.3), and compare market cap to GDP growth.

It turns out that even over these long historical period, differences in capitalization growth rates are driven by equity prices. Equity issuance is sizeable at roughly 4% of market cap, but is remarkably stable over both short and long horizons, remaining at around 4% of market cap during the first century of our data and actually falling slightly to 3.4% of market cap during the post-1985 market expansion. Capital gains, on the contrary, are small on average – around 0.9% per year – but vary substantially across periods. The first 40 years of our data saw slightly positive capital gains of 1.2%, largely attributable to an absence of large negative shocks (Figure 1.5). It is these capital gains, and not issuance, which were the main driver of the modest pre-1914 stock market expansion. During the stagnation period between 1914 and 1985, real capital gains were negative, thanks to the shocks of the two World Wars, the Great Depression and the 1970s stagflation. These large equity price shocks, then, explain why market capitalization did not grow faster than GDP during this period. The period after 1985 saw equity capital gains of, on average, 3.6% per year persisting for three decades – a magnitude roughly four times the historical norm. As a result, the size of the stock market expanded rapidly despite net equity issuance actually slowing down from roundabout 2000 onwards.

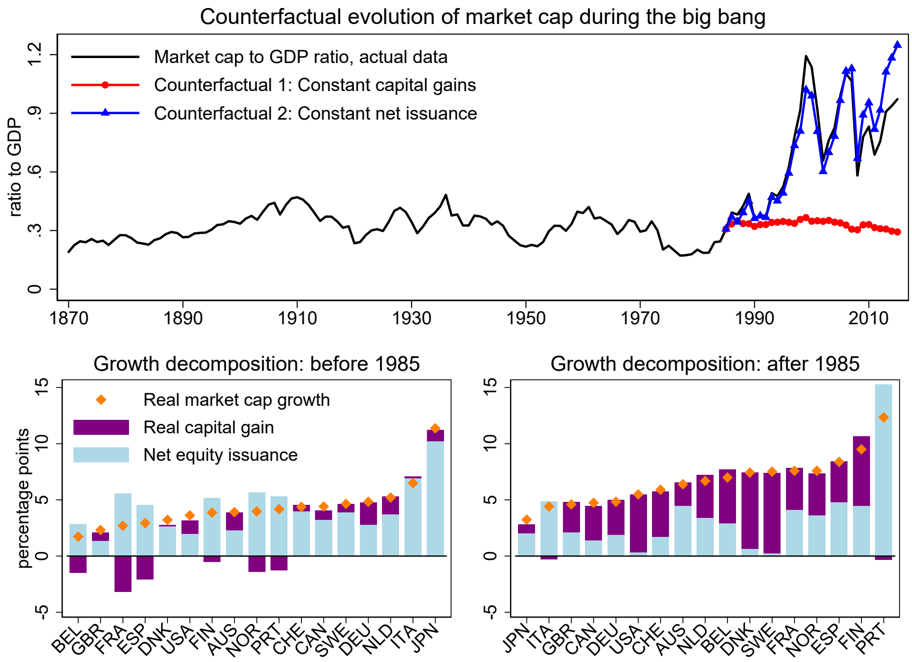
Focussing on the big bang increase in equity wealth, the top panel of Figure 1.6 displays two counterfactual market cap evolutions together with the actual data (solid black line). The first counterfactual, marked by red diamonds, shows what the market cap evolution after 1985 would have been if we fixed capital gains to their pre-1985 average. Under this scenario, all changes in the market cap to GDP ratio from 1985 onwards are attributable to net issuance and real GDP growth. 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. We

Table 1.1. Market capitalization growth over long time periods

	(1) Full sample	(2) Pre 1914	(3) 1914–1985	(4) Post 1985
Average change in *MCAP/GDP*	0.68	0.61	-0.00	2.21
Market cap growth decomposition:				
Real market cap growth ≈	4.82	5.22	3.80	6.68
Real capital gain on equity	0.85	1.18	-0.54	3.55
+ Implied issuance to market cap	3.97	3.95	4.23	3.42
Real GDP growth	2.83	2.42	3.26	2.27
Observations	2096	473	1113	510

Notes: Top row: average annual change in the market cap to GDP ratio, percent of GDP. Middle rows: decomposition of real market cap growth into real equity capital gains and net equity issuance relative to previous year's market cap. Bottom row: year-on-year growth in real GDP. All data are pooled sample averages of log growth rates.

Figure 1.6. Counterfactual and cross-country evidence



Notes: Top 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. Data are benchmarked so that the combined growth of the two counterfactuals between 1985 and 2015 equals the actual growth in observed market cap data. All data are unweighted averages of 17 countries. Bottom panel: Real market cap growth decomposed into real capital gains and net issuance relative to market cap. Averages of log growth rates for each country before and after 1985. The sum of the bars can deviate from the orange diamond because of the correlation between issuance and capital gains.

benchmark the estimates so that the combined growth under the two counterfactual scenarios equals the actual growth in market cap over 1985–2015.[2]

Counterfactual 1 shows that ignoring the post-1985 capital gains eliminates the big bang entirely from our data, with market cap growing in line with GDP and even declining slightly after 2000 due to lower net issuance. Counterfactual 2 closely follows actual data, which means that higher post-1985 capital gains can explain the entirety of the big bang. If anything without the slight slowdown in net issuance after 1985 shown in Table 1.1 column 4, the growth in market cap over recent decades would have been even stronger. Again, we confirm these findings using actual issuance data from four countries in the bottom panel of Appendix Figure 1.A.3.

The bottom panel of Figure 1.6 shows that these aggregate trends are also reflected in coutry-level data. The left-hand panel shows the average growth rate of market cap before the big bang decomposed into capital gains and issuance, and the right-hand panel shows the average growth rate during the big bang. As with cross-country averages shown in Table 1.1 and Figure 1.5, almost all of the stock market cap growth before 1985 is attributable to net issuances (light-blue bars) rather than capital gains (purple bars). After 1985, the situation changes. For around half the countries, capital gains make a larger contribution than issuances, and out of the remaining half only Portugal shows strong market cap growth backed by high issuance. This country, however, presents a rather special case: it stock market almost disappeared after the 1974 Carnation revolution, and the high issuance relative to market cap simply reflects the re-establishment of this market from a very low base.

Capital gains can also account for cross-country differences in the size of the stock market increase during the big bang. Appendix Figure 1.A.4 shows that countries with the highest real capital gains also experienced the largest increase in market cap to GDP ratios. Appendix Figure 1.A.5 additionally shows the counterfactual evolutions of market cap for each country, first fixing capital gains and then net issuance to their pre-1985 levels. In every country bar two (Australia and Portugal), the fixed-issuance counterfactual closely follows actual data whereas the fixed-capital-gain counterfactual results in little or no market cap growth, again showing that higher post-1985 capital gains are key.

In sum, the stock market growth of the last three decades was not driven by equity issuance or new listings. Rather, the entirety of the big bang and the associated long-run disconnect between stock markets and GDP is driven by a persistent stock price increase, the drivers of which we explore in the next section.

2. The benchmarking ensures that the different timing of the shocks to issuance and capital gains, and the correlation between the two, do not bias our findings. For example, after the burst of the 1980s Japanese bubble both issuance and capital gains were sharply negative, meaning that any subsequent growth took place from a very low base. Ignoring the correlation between these two shocks would overstate the counterfactual market cap growth under both scenarios. That being said, data for non-benchmarked counterfactuals show even higher market cap growth under counterfactual 2 of fixed issuance which further confirms our findings; results available from authors upon request.

1.5 What drives long-run stock market growth?

This section explores the drivers of the persistent equity price increase that underlies the post-1980s expansion in stock market wealth. We start by identifying the different drivers of market capitalization within the framework of a dynamic Gordon growth model. We first note that the ratio of market capitalization to GDP is approximately equal to the ratio of aggregate dividends paid by listed firms to GDP, D^{agg}/GDP, times the average price-dividend ratio in the economy P_t/D_t:

$$\frac{MCAP_t}{GDP_t} = \sum_{i=1}^{N_t}\left(\frac{Q_{i,t}D_{i,t}}{GDP_t} * \frac{P_{i,t}}{D_{i,t}}\right) \approx \frac{D_t^{agg}}{GDP_t} * \frac{P_t}{D_t}, \tag{1.4}$$

where $D_t^{agg} = \sum_{i=1}^{N_t} Q_{i,t}D_{i,t}$.

We can then take logs on both sides and apply the Campbell and Shiller (1988) decomposition to the log price-dividend ratio:

$$ln\left(\frac{MCAP_t}{GDP_t}\right) \approx ln\left(\frac{D_t^{agg}}{GDP_t}\right) + \mathbb{E}\left(\sum_{\tau=1}^{\infty}\rho^{\tau}(dg_{t+\tau} - r_{t+\tau})\right), \tag{1.5}$$

where dg is log real dividend growth, r is log real total return and $\rho = \frac{P/D}{1+P/D}$ is a linearisation constant. Intuitively, equation (1.5) is an approximate present value identity for the entire stock market. It tells us that the size of the stock market relative to the economy is determined by three factors.

The first factor is the *current profit share* of listed firms D^{agg}/GDP. If listed firms' profits and dividends constitute a large share of economic income, the size of the stock market – which reflects the present value of these profit and dividend streams – will also be large relative to the rest of the economy. The second factor is the *future profit growth* (or dividend growth) dg: stocks can be valued highly not only because profits are high today, but if they are expected to be high in the future. The third factor r is the *discount rate* which determines the present value of future profits. The lower the discount rate or expected return, the more valuable are future cashflows and the higher the stock prices.

We evaluate the relative importance of these three drivers using a two-step procedure. First, we look at the correlation between market capitalization and its three determinants. As we cannot measure return and dividend growth expectations directly, we follow existing literature (Campbell and Shiller, 1988; Cochrane, 2008) and run predictive regressions of future dividends and future returns on current market capitalization. If high market capitalization predicts low future stock returns or high future dividend growth, this means that r and dg are relevant drivers of fluctuations in the size of the stock market. These correlations tell us about which forces tend to drive the year-on-year growth in the market cap to GDP ratio, but they do not necessarily tell us what drives its long-run trend. To this end, in the second step we look at the trends in those variables which are correlated with market capitalization in the right direction in order to determine the contribution of these trends to the long-run evolution of market cap.

1.5.1 Market capitalization, future returns and dividends

Figure 1.7 shows the correlations of market capitalization with the three components of equation (1.5): the current share of dividends in GDP, future growth of dividends, and future returns. As in equation (1.5), all the variables are in logs, and we calculate present value discounted sums $\sum_{j=1}^{\infty} \rho^j dg_{t+j}$ and $\sum_{j=1}^{\infty} \rho^j r_{t+j}$ by taking realised returns and dividends between year t and 2015, and assuming returns or dividend growth are equal to the country-specific sample mean after 2015.[3] We also adjust market cap and dividends to GDP for structural breaks so that all variables are stationary to avoid potentially spurious correlations, and postpone the questions of time series trends and structural breaks until the second step of the analysis in Sections 1.5.2 and 1.5.3. Not adjusting for the structural breaks, however, leads to similar results.

The correlation between market capitalization, dividend share in GDP and future returns is in line with the theoretical predictions in equation (1.5): a high market cap to GDP ratio means high current listed firm dividends relative to GDP, and low future equity returns. But market capitalization and future dividend growth are, if anything, correlated in the wrong direction: high cap actually predicts low, not high future dividend growth.

Table 1.2 further tests these relationships by running predictive regressions of the following form:

$$r_{j,t+1} = \alpha_j^r + \beta^r ln(MCAP_{j,t}/GDP_{j,t}) + u_{j,t}^r \tag{1.6}$$

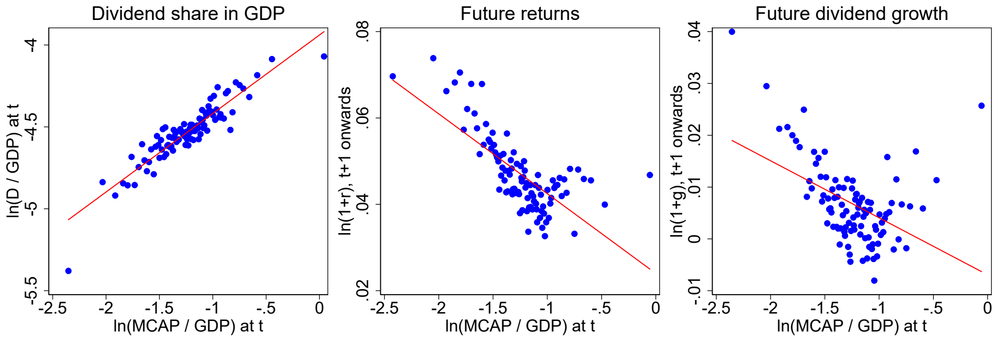
$$dg_{j,t+1} = \alpha_j^{dg} + \beta^{dg} ln(MCAP_{j,t}/GDP_{j,t}) + u_{j,t}^{dg}, \tag{1.7}$$

where j is a country index, α is a country fixed effect, and u is a Driscoll and Kraay (1998) standard error clustered by country and year, and adjusted for serial autocorrelation.

The Campbell and Shiller (1988) decomposition in equation (1.5) tells us that the regression coefficient on future returns β^r should be negative, and β^{dg} on dividend growth should be positive. In the data, β^r is negative, and statistically and economically significant. In the baseline specification in column 1, a 10 percentage point increase in the market cap to GDP ratio (around half a standard deviation)

3. In Figure 1.7, we annualise the discounted sums of r and g by multiplying them by $1-\rho$.

Figure 1.7. Correlations between market cap, current dividend share, future returns and future dividend growth



Notes: Binned scatter plots of pooled full-sample data (17 countries, 1870–2015), 100 bins. Each point represents the group specific means of both variables after controlling for country fixed effects. All variables are winsorized at 1 percent level. Market cap to GDP and dividends to GDP are adjusted for structural breaks.

Table 1.2. Stock market capitalization as a predictor of equity returns and dividends

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline		Post-1985		Year effects	
	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}
$ln(MCAP_t/GDP_t)$	-0.134***	-0.068*	-0.193***	-0.012	-0.088***	-0.059
	(0.029)	(0.038)	(0.048)	(0.064)	(0.026)	(0.045)
R^2	0.054	0.011	0.087	0.000	0.497	0.191
Observations	2076	2076	519	519	2076	2076

	(7)	(8)	(9)	(10)	(11)	(12)
	Risk premia and safe rates		5-year returns		No structural Breaks	
	er_{t+1}	r^{safe}_{t+1}	$\bar{r}_{t+1,t+5}$	$\overline{dg}_{t+1,t+5}$	r_{t+1}	dg_{t+1}
$ln(MCAP_t/GDP_t)$	-0.110***	-0.046**	-0.097***	-0.048***	-0.036***	-0.005
	(0.028)	(0.020)	(0.021)	(0.018)	(0.011)	(0.015)
R^2	0.031	0.012	0.136	0.030	0.016	0.000
Observations	2076	2076	1991	1991	2076	2076

Notes: Market cap is adjusted for structural breaks. r is total real return, dg is real dividend growth, er is excess return and r^{safe} is real government bond return, all measured in logs. Regressions with country fixed effects. Columns (5) and (6) additionally add year fixed effects. Standard errors clustered by country and year and adjusted for autocorrelation are in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

predicts 3.5 percentage point lower return one year ahead.[4] But contrary to the model's predictions, high market capitalization actually predicts negative real dividend growth, with 10 percentage points higher capitalization forecasting roughly 1.8 (0.25*(-0.068)*1.03) percentage points lower growth 1 year ahead, using the numbers in column 2.

Return predictability becomes somewhat stronger and dividend predictability weaker if we limit the sample to the post-1985 period (columns 3 and 4), and the coefficients are similar to baseline when we control for common cross-country time variation in returns, cashflows and yields through year fixed effects (columns 6 and 7). Columns 7 and 8 show that both excess equity returns and safe rates are predictable, suggesting that market capitalization responds to changes in both the real safe rate and the ex ante equity risk premium. Columns 9 and 10 show that these predictive relationships are highly persistent: 10 percentage points higher capitalization predicts a 12.7 percentage points lower cumulative real equity return and 6.2 ppts lower cumulative real dividend growth 5 years ahead. Columns 11 and 12 show that even without adjusting for structural breaks, high capitalization predicts low returns and does not predict dividend growth in the right direction, even though the predictive coefficient on future growth becomes insignificant. This mirrors the findings of Lettau and Van Nieuwerburgh (2008) who show that return predictability in the US becomes stronger after adjusting the predictor variable for structural breaks.

Under all the different regression specifications, market capitalization responds to current profit share and the discount rate channels, but not to future growth. Expected future profitability increases are therefore unlikely to be the main driver of the big bang. We next look at the trends in profit shares and discount rates to determine their relative contribution to the post-1980s increase in the market cap to GDP ratio.

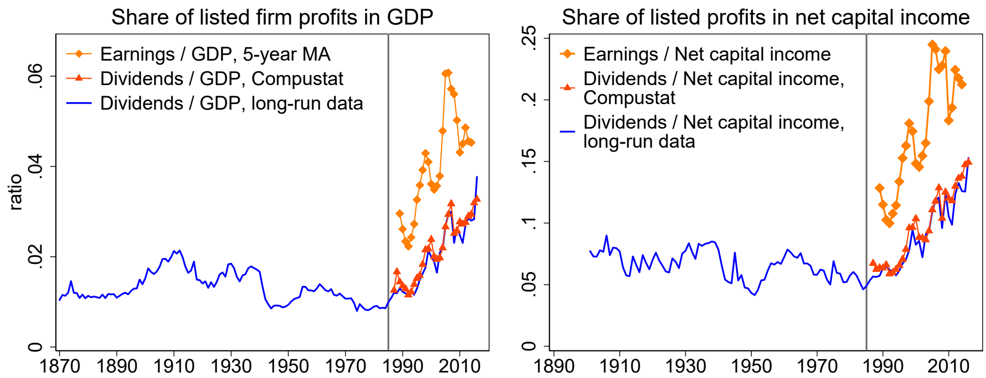
1.5.2 A profit shift towards listed firms

Profits of US corporations have increased markedly over recent decades (Barkai, 2020). This section puts these recent US trends into a cross-country historical perspective, and shows that the corporate profit boom is global, historically unprecedented, and a key driver of the post-1980s increases in market capitalization. We also show that the increase in listed firms' profits has come primarily at the expense of other forms of capital income, and is therefore likely to reflect a redistribution of earnings within the corporate sector.

We start by examining long-run trends in the share of listed firm profits in GDP. The left-hand panel of Figure 1.8 plots the long-run evolution of dividends paid by

4. A 10 percentage point increase is roughly 25% in relative terms, meaning a $0.25 \times 0.134 \times 1.048 \approx 0.035$ fall in real total return.

Figure 1.8. Listed firm profit share over the long run

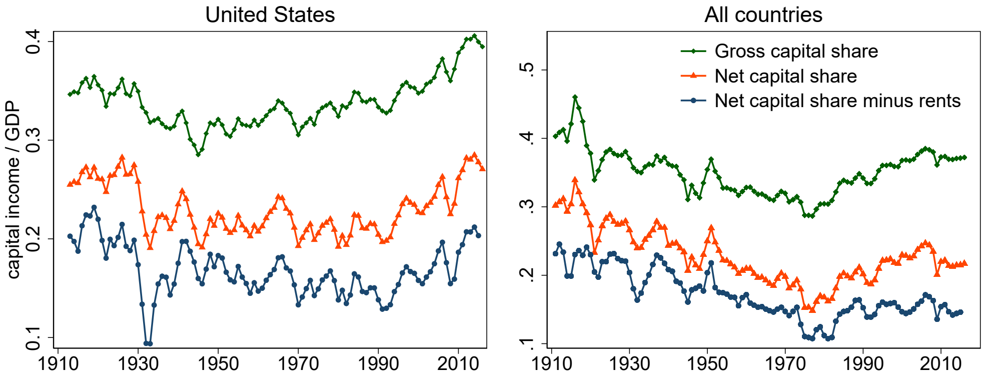


Notes: Unweighted averages of 17 countries. The dividend-to-GDP ratio for each country is calculated as the dividend yield D_t/P_t multiplied by the market cap to GDP ratio $MCAP_t/GDP_t$. Profit data are aggregated up from Compustat Global and Compustat North America and cover all listed firms with non-missing values for market cap, dividends and earnings, scaled up to match our aggregate market cap data where necessary and dropping periods with insufficient coverage. Listed firm profits are smoothed using a 5-year moving average. Data on gross and net capital income are from Bengtsson and Waldenström (2018).

listed firms relative to GDP (solid blue line) alongside the share of listed company earnings in GDP (orange diamonds). The share of dividends in GDP has roughly tripled between the 1980s and today – an increase comparable to that in the market capitalization to GDP ratio. Dividends are, however, a noisy proxy for total cashflows to shareholders and could be affected by changes in firms' payout policy, such as the increasing use of stock buybacks in the US starting in the early 1980s (Grullon and Michaely, 2002). Data on corporate earnings allow us to look through such payout policy changes and focus on the underlying profitability trends. Figure 1.8 therefore also presents the estimates of listed firm earnings for the 17 countries in our sample, computed using data in Compustat Global and Compustat North America. The red triangles in Figure 1.8 show that the corresponding Compustat dividend series matches well with our aggregate dividend data. Similarly to dividends, earnings of listed firms increased markedly during the big bang, rising from 2.5–3% of GDP to 5–6% of GDP during this period. Appendix Figure 1.A.6 shows that the recent earnings increase is also unusually large and persistent relative to those observed historically in longer-run US earnings data from Shiller (2015). Appendix Figure 1.A.7 further documents that these increases in earnings were not accompanied by significant changes in firm leverage.

The increase in listed firm profits relative to GDP suggests that there has been a redistribution of income away from other sectors of the economy towards listed firms. This redistribution could have come about from the following two sources: labour income or other forms of capital income such as profits of unlisted corporations. If it has come about at the expense of labour income, the profit shift should be accompanied by an increasing share of capital income in GDP, and the share of listed firms profits in capital income should be stable over time. The right-hand

Figure 1.9. Capital shares in a long-run perspective



Notes: Left-hand panel: data for the US. Right-hand panel: unweighted average of data for Australia, France, Germany, Sweden, UK and US. Gross and net capital income from Bengtsson and Waldenström (2018). Net capital share minus rents subtracts the net rental income from national accounts of the respective country.

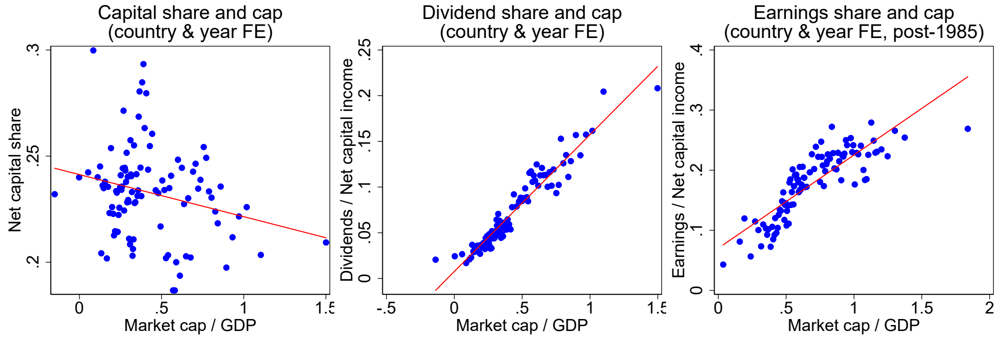
panel of Figure 1.8 shows that this is not the case. It plots the ratio of listed firms' earnings and dividends to net capital income, using newly released long-run cross-country capital share estimates from Bengtsson and Waldenström (2018). The ratio of listed profits to capital income has recorded a similar relative increase to the share in GDP, from about 10% in the 1990 to some 25% in 2015. Longer-run data on dividends paid by listed firms show that this within-capital shift is unusually large and persistent by historical standards, with the ratio of listed firms' dividends to capital income currently standing at an all-time historical high of 15%.

Karabarbounis and Neiman (2013) have shown that the share of labour income in GDP has fallen since the 1980s. In a contemporaneous paper to ours, Greenwald, Lettau, and Ludvigson (2019) argue that the decline in the labour share has been a key driver of the post-1989 increase in the US stock market capitalization. To investigate the contribution of this channel to the long-run cross-country increases in capitalization in our data, Figure 1.9 plots the long-run evolution of three different capital share measures in the US and on average across six advanced economies. The green diamond line shows the gross capital share, the orange triangle line adjusts this measure for depreciation and the blue square line additionally subtracts net rental income to give us an estimate of the non-housing net capital share – a measure that most closely tracks the income of total corporate capital. We take the gross and net capital shares from Bengtsson and Waldenström (2018), and use the balance sheet net rental income data from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) to further adjust these for rents.

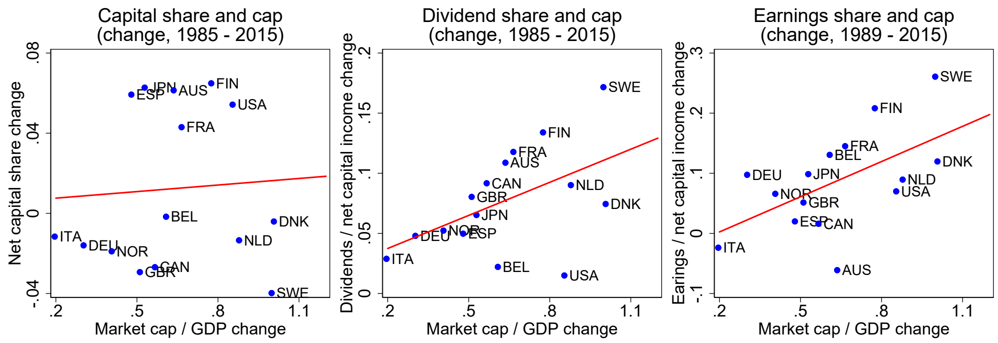
Our evidence suggests that changes in the capital share play relatively little role in explaining the big bang in stock market capitalization. While the gross US capital share has indeed increased substantially since the 1980s, the net-of-housing net capital share has remained broadly flat, in line with evidence in Rognlie (2015) and Gutiérrez and Piton (2020). Furthermore, both gross and net capital shares in other

Figure 1.10. Stock market capitalization, capital share, and profit share

(a) Correlations between market cap, capital share, and share of profits in capital income



(b) Post-1985 growth in market cap, capital share and profit share by country



Note: Capital share is net capital income / GDP, from Bengtsson and Waldenström (2018). Dividend share is listed firms' dividends / GDP. Earnings share is listed firms' earnings / GDP, averaged over 3 years (Compustat Global and North America data). Top panel: bin scatter plot, 100 bins, controlling for country and year fixed effects. Capital and dividend share for 1870–2015; earnings data for 1985–2015. Bottom panel: country-level changes in capital, dividend and earnings shares during the big bang. Earnings share change is for 1989–2015 to ensure consistency across countries.

advanced economies have, if anything, declined slightly over the long run (Figure 1.9 right-hand panel). Taken together, these trends suggest that a redistribution of profits within capital rather than between capital and labour is a key driver of the post-1980s increases in market capitalization across advanced economies.

To investigate the link between market capitalization, capital shares and profit shares more systematically, Figure 1.10 studies the correlation between these variables in the full sample (top panel) and during the big bang (bottom panel). Figure 1.10a shows a bin scatter plot of market cap against, respectively, net capital share in GDP, listed firm dividend share in net capital income, and listed firm earnings share in net capital income, divided into 100 bins. To avoid spurious correlation, the plots control for country and year fixed effects, but removing these does not affect the results. From the broad long-run cross-country perspective, capital shares are uncorrelated with market capitalization, while dividend or profit shares and market cap are strongly positively correlated. It could still be the case, however, that some of these variables are uncorrelated in their year-on-year movements but display somewhat similar patterns during the big bang.

To assess the importance of capital share and profit share changes during the big bang, we take advantage of the cross-country dimension of our data and ask if those countries which reported the largest capital or profit share increases after 1985 also observed the largest increases in market cap. Figure 1.10b shows that countries with larger increases in the capital share did not, as a rule, also record larger increases in stock market capitalization. While US capitalization and capital shares both increased, Sweden recorded both the largest cap increase and the largest *fall* in the capital share. Overall, country-level growth in the capital share and capitalization during the big bang are uncorrelated. The middle and right panels of Figure 1.10b show that, on the contrary, countries which saw the largest increases in the dividend to GDP or earnings to GDP ratios also recorded the largest increases in market capitalization.

To sum up, the big bang was accompanied by large increases in the *listed firm profit share* – the ratio of listed firms' earnings to GDP and capital income. These profit increases are comparable in size to those in market cap, occurred in a wide cross-section of countries, are historically unprecedented, and present a fundamentally different phenomenon from the well documented increases in the capital share of income. What could be driving these profit shift movements? One potential candidate is an increase in market power of large listed firms. De Loecker, Eeckhout, and Unger (2020) and De Loecker and Eeckhout (2018) show that markups and profitability of large listed firms have increased over the past 30 years, both in the US and globally. They also show that these markups are strongly correlated with firm-level market capitalization in the cross-section.[5] The associated redistribution of profits within the corporate sector is likely to be an important driver of the profit shift and increases in market capitalization during the big bang.

1.5.3 A declining discount rate

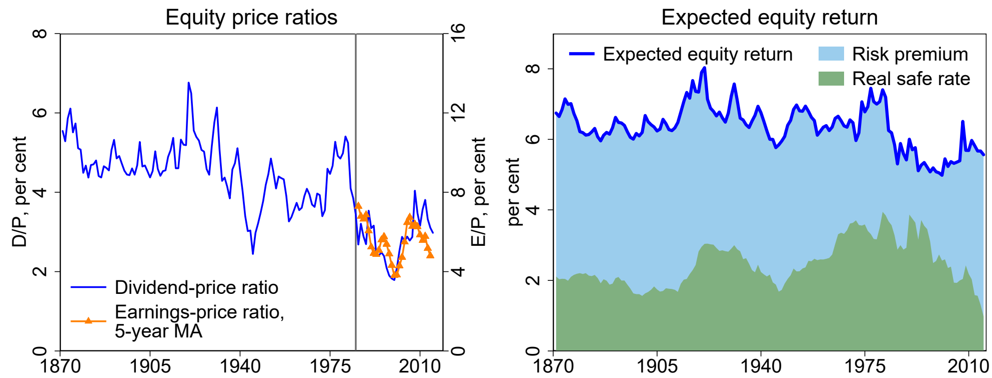
Mapping profit and cashflow increases into stock market capitalization requires an estimate of the equity discount rate. We calculate the equity discount rate r as the sum of the dividend-price ratio and expected long-run future cashflow growth:[6]

$$\mathbb{E}\left(R_{t+1}\right) \approx \mathbb{E}\left(D_{t+1}/P_t\right) + \mathbb{E}\left(\tilde{g}_{i,t+2}\right) \tag{1.8}$$

5. Philippon (2019) and Diez, Leigh, and Tambunlertchai (2018) provide further evidence of rising market power in the US and globally.

6. This is a levels version of the Campbell and Shiller (1988) decomposition derived by Blanchard (1993). Here $\tilde{g}_{i,t+2}$ is the annuity value of future dividend growth, calculated as $\tilde{g}_{t+2} = w_1\mathbb{E}g_{t+2} + w_2\mathbb{E}g_{t+3} + \ldots + w_\tau\mathbb{E}g_{t+\tau+1}$. $g_t = D_t/D_{t-1} - 1$ is the year-on-year cashflow growth and $w_t = (1+g)^{\tau-1}(r-g)/(1+r)^\tau$ are the weights, where g and r are the average dividend growth and return rates.

Figure 1.11. Trends in equity discount rates

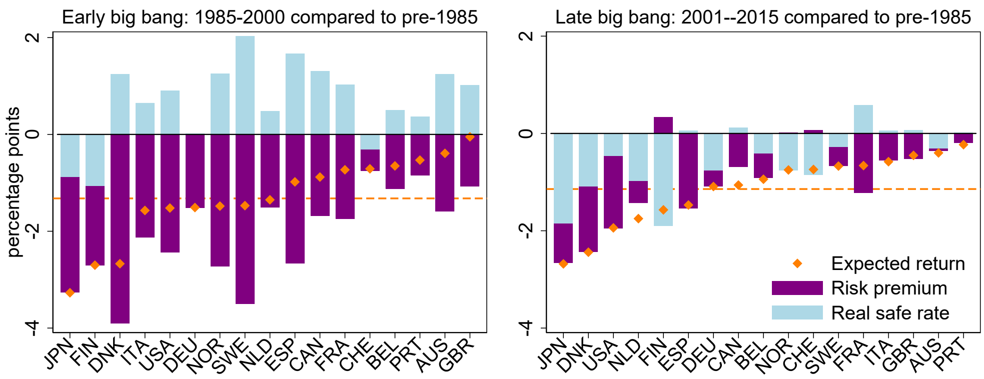


Notes: Unweighted averages of 17 countries. Left-hand panel shows the average dividend-price ratio in our long-run data and the earnings-price ratio from Compustat Global and North America. Right-hand panel shows the expected equity return, calculated as the dividend-price ratio plus expected cashflow growth. Cashflow growth is an average of the GDP growth forecast, and dividend growth forecast using the Campbell and Shiller (1988) decomposition. The safe rate is the trend long-term real government bond rate estimated using the time series filtering method of Del Negro, Giannone, Giannoni, and Tambalotti (2019), and the risk premium is the difference between expected return and safe rate. See Kuvshinov and Zimmermann (2020) for more details.

If cashflow expectations are stable over time, the equity discount rate should move in line with the dividend-price ratio. Therefore, as a first proxy for r, the left-hand panel of Figure 1.11 plots the evolution of the dividend-price ratio (solid blue line) over the long run. The dividend-price ratio has declined over the long run, falling from some 5-6% in the late 19th century to a low of 2% in 1990 before recovering slightly to around 3% today. Much of this decline slightly predates the market cap surge of the 1990s, and suggests that the profit shift of 1990s took place at a time of historically low discount rates, which means that it translated into an unusually large asset price increase. The orange triangles in Figure 1.11 also show the change in the earnings-price ratios between 1989 and 2015, computed using cap-weighted averages of the firm-level Compustat data for each of the 17 countries, and then averaged over 5-year windows and across countries. The earnings-price ratio shows a similar decline to the dividend-price ratio, which means that the dividend-price ratio decline is not simply driven by firms switching from dividends to other forms of shareholder compensation. Long-run data for the US – shown in the Appendix figure 1.A.6 – also show that these earnings-price ratio declines are also historically unusual.

The decline in dividend-price ratios could, however, simply be attributable to higher cashflow growth expectations. In this case, a higher $\tilde{g}$ in equation (1.8) would offset the decline in D/P such that expected returns hardly move. To see if this is the case, we estimate expected cashflow growth $\mathbb{E}(\tilde{g})$ and hence the expected return in equation (1.8). For details of this estimation, the reader is referred to Kuvshinov and Zimmermann (2020) which studies how expected returns and risk premia on

Figure 1.12. Drivers of falling expected returns during early and late big bang



Notes: Left-hand panel shows the difference between the average level of expected return, risk premium and safe rate during 1985–2000, and 1870–1984, for each country. Right-hand panel shows the differences between average levels in 2001–2015 and 1870–1984.

housing and equity have evolved over the long run. The solid blue line in the right-hand panel of Figure 1.11 is the average of two alternative estimates of $\mathbb{E}(g)$: first, the Campbell-Shiller decomposition of the dividend-price ratio into cashflow and discount rate news, calculated using a cross-country VAR with time-varying coefficients; and second, an estimate of the long-run growth rate of the economy. These estimates show that long-run variation in expected cashflow growth does not offset the decline in the dividend-price ratio in the left-hand panel of Figure 1.11, and the expected equity return declines after the 1970s.

The right-hand panel of Figure 1.11 further decomposes the discount rate into a risk premium and a safe rate component using the trend real safe rate estimates of Del Negro, Giannone, Giannoni, and Tambalotti (2019) (again, more details can be found in Kuvshinov and Zimmermann, 2020). The initial decline in the equity discount rate, which largely preceded the big bang, is driven by a lower equity risk premium. The increases in market cap to GDP ratios during the 1980s and 1990s were, therefore, so large partly because equity risk premia were at historically low levels. After 1990, the equity premium has increased back close to its long-run average but the discount rate has remained low and market capitalization – high due to the well-documented post-2000 decline in the safe interest rate (Holston, Laubach, and Williams, 2017). Our finding of the post-1990 increase in the equity premium is consistent with evidence of a recent divergence between risky and safe returns in the US documented by Caballero, Farhi, and Gourinchas (2017) and Farhi and Gourio (2018).

Figure 1.12 shows that these time trends are also evident at the country level. The left-hand panel of Figure 1.12 compares the level of risk premia and safe rates in the early phase of the big bang – years 1985–2000 – to the pre-1985 period. While safe rates were high by historical standards, risk premia were unusually low. The right-hand panel of Figure 1.12 compares safe rate and risk premium levels during the late stage of the big bang – years 2000–2015 – to historical averages. Unlike the early big bang period, safe rates in a number of countries are materially below the historical average, while the differences in risk premia compared to historical averages are less pronounced than during the initial stage of the big bang. In Kuvshinov and Zimmermann (2020), we argue that the long-run equity discount rate decline is primarily driven by a declining price of risk, driven by factors such as the decline in global macroeconomic volatility, consistent with evidence in Lettau, Ludvigson, and Wachter (2008) and Bianchi, Lettau, and Ludvigson (2016) who link the post-1950 US equity premium decline to, respectively, lower consumption volatility and a greater stabilising role of monetary policy. The fact that the timing of the profit shift coincided with the onset of the great moderation has, therefore, been an important factor in propping up equity market capitalization.

1.5.4 A back-of-the-envelope counterfactual

The big bang is explained by a profit shift towards listed firms – and away from other forms of capital income – at the time of historically low discount rates. How much contribution to the post-1980s trends in market capitalization does each of these two factors make? To determine this, we run a simple counterfactual exercise of the following form: we allow either the profit share or the discount rate to vary with the actual trend observed in the data, but fix all the other market cap determinants at their long-run levels. We then estimate the counterfactual market cap using a constant-growth Gordon model version of equation (1.5):

$$\frac{MCAP_t}{GDP_t} \approx \frac{D_t^{agg}}{GDP_t} * \frac{1}{\mathbb{E}(r_{t+1}) - \mathbb{E}(g_{t+1})} \tag{1.9}$$

To ascertain the relative importance of the different channels, we compare the counterfactual increase in market cap obtained by varying D/GDP or $\mathbb{E}(r)$ and keeping other factors constant to the increase observed in the data. Throughout this counterfactual exercise, we fix expected dividend growth $\mathbb{E}(g)$ at its long-run mean level of 2.6% p.a.[7]

Table 1.3 shows the actual market cap to GDP ratio before and after big bang in the top row and the counterfactual market cap levels under various scenarios in the bottom rows. Column 1 shows the actual and counterfactual market cap levels during the 1980s, the decade before the big bang. Column 2 shows the average

7. To calculate the full-sample mean, we first calculate the annuity value of real dividend growth $\tilde{g}_{t+2}$ in equation (1.8) as the discounted sum of all future dividend growth rates from $t+1$ onwards, setting the growth rate after 2015 to the sample mean. We then winsorize the data at the 1% level to reduce the influence of outliers, e.g. high growth rates from a very low base. Using non-winsorized or non-annuity based averages results in a higher g and hence makes us able to explain an even higher proportion of the big bang with the observed profit share and discount rate trends.

Table 1.3. Contribution of discount rates and dividends to the big bang

	(1)	(2)	(3)	(4)
	1980 – 1989 average	2015	Change in MCAP/GDP	Share of increase in the data
Actual MCAP/GDP	0.31	0.97	0.67	
Counterfactuals:				
Profit shift only	0.28	0.79	0.51	76%
Discount rate decline only	0.28	0.35	0.06	10%
Combined	0.28	0.97	0.68	103%

Notes: Pooled averages, annual data for 17 countries. Counterfactual market cap is calculated as $\frac{D}{GDP} * \frac{1}{\mathbb{E}(r)-\mathbb{E}(g)}$, where D/GDP are dividends paid by all listed firms relative to GDP, $\mathbb{E}(r)$ is the expected return on equities, and $\mathbb{E}(g)$ is expected real dividend growth, set to sample average of 2.6%.

level after the big bang, in 2015. Columns 3 and 4 show the resulting market cap increases.

Starting with column 1, the average level of market cap before the big bang was around one-third of GDP. During this period, dividends were around 1% of GDP and expected equity returns were on average 6.2%. Plugging these into equation (1.9) gives us a counterfactual market cap to GDP ratio of $1/(6.2-2.6) \approx 0.28$, close to the ratio of 0.31 in the data – a starting point for all our counterfactual scenarios. Row 2 of Table 1.3 allows the profit share D/GDP to increase to its 2015 average of 2.8%, while keeping the discount rate constant at 6.2%. This gives us a $2.8/(6.2-2.6) \approx 0.79$ counterfactual level of market cap in 2015, and a 51% of GDP increase in market cap during the big bang – around 76% of the total. Row 3 allows the discount rate to fall from 6.2% to 5.6% while keeping dividends-to-GDP constant at 1%, which gives us a $1/(5.6-2.6) \approx 0.35$ counterfactual 2015 market cap level and a 6% of GDP increase, around 10% of the total.

Together, these two factors amplify each other as higher cashflows are discounted at a lower rate. Allowing for both the profit shift and the discount rate decline in counterfactual 3 results in a $2.8/(5.6-2.6) \approx 0.97$ counterfactual 2015 market cap level, and an increase of 68% of GDP accounting for the entire observed increase in market cap during the big bang. Note that starting the counterfactual scenario before the 1980s would somewhat increase the relative contribution of the fall in the discount rate because expected equity returns started falling before 1980 (Figure 1.11). Starting it later would increase the contribution of the profit shift relative to that of the discount rate. But regardless of the starting date, we are always able to explain almost all of the increase in market cap during the big bang, and the profit shift remains the dominant factor.

The increase in market capitalization during the 1980s and 1990s is rather unique from a historical perspective. But the data also show many more shorter-run fluctuations and mean-reverting market movements. How representative is the observed long-run market cap increase of increases in capitalization more generally, and what can market capitalization tell us about upcoming stock market movements? We turn to address these questions in the next section.

1.6 The Buffet indicator

1.6.1 Stock market run-ups and crashes

Recent decades saw sharp and persistent increases in stock market capitalization. But from a broader historical perspective, this time period was rather unique. The first century of our data saw many short-run deviations with market capitalization rising sharply and then reverting to its long-run mean. These booms and busts include country-specific episodes such has the Scandinavian post-World-War 1 booms which ended in a series of financial crises during the early 1920s and the US roaring 1920s followed by the 1929 crash. They also include global events such as the early 1900s expansion which was ended by World War 1, and the post-World-War 2 boom which unwound during the 1970s stagflation.

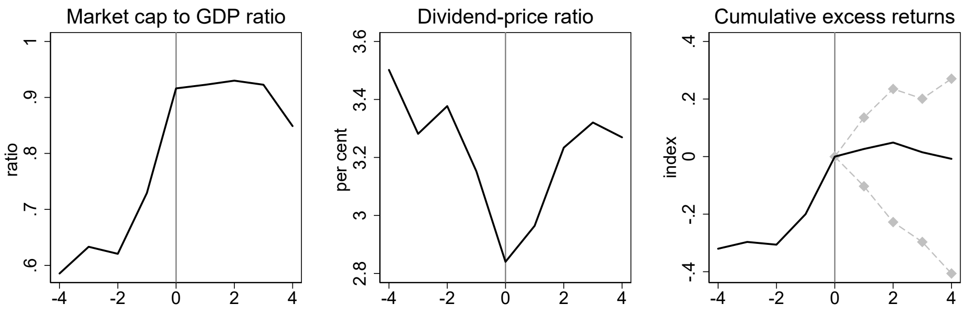
We first investigate how typical or untypical the "big bang" market expansion was by asking the following question: if we observe a sharp run-up in market cap, is it likely to be persistent or mean revert, potentially leading to a crash? We identify the run-ups in our sample using a definition similar to that of Greenwood, Shleifer, and You (2019) sector-specific equity market bubbles. More precisely, we look at sharp capitalization increases during 2 years (20% of GDP or more) which do not follow a previous downturn (at least a 10% of GDP increase over preceding 5 years). The 5-year growth requirement allows us to focus on run-ups and exclude recoveries from temporarily low market cap levels. The threshold levels are calibrated so that we get a similar number of run-ups to those examined by Greenwood, Shleifer, and You

Table 1.4. Frequency of market booms and crashes after a market cap run-up

	Number of events			*Share of run-ups ending in:*	
	Run-ups	Crashes	Structural increases	Crashes	Structural increases
Post-run-up sample	75	50	19	67%	25%
Full sample	75	150	34		

Notes: Top row shows market crashes and structural increases after a run-up in market cap, as numbers and share of total number of run-ups. A crash is a cumulative stock market return of -25% or less over 1 or 2 years. A structural increase is a structural break in market cap which results in a higher post-break mean. Bottom row shows the total number of run-ups, crashes and structural increases in the full sample.

Figure 1.13. Stock valuations and returns around sharp increases in market cap



Notes: Average market cap to GDP, dividend-price ratio and cumulative real return during and after stock market run-ups, defined as a 20% GDP or higher increase in market cap over 2 years ($t = -2$ to $t = 0$), and 10% GDP or higher increase over 5 years ($t = -5$ to $t = 0$) (75 run-ups in total). Excess return is the total stock return minus the return on government bonds, indexed to 0 at $t = 0$.

(2019).[8] We then look at whether these run-ups coincide with a structural increase in market cap – a structural break leading to a higher mean – within a ±2 year window, or are followed by a crash – a -25% or lower cumulative real equity return over 1 or 2 years within 4 years.

Together, we identify 75 such run-up episodes in our long-run sample. The top row of Table 1.4 shows how many of these run-ups are followed by a crash, and how many coincide with a structural increase in market cap. Out of the 75 run-ups, 50 are followed by market crashes and 19 by structural cap increases. Put differently, a run-up has a 67% chance of ending up as a crash and mean reverting, and a 25% chance of being persistent as in the case of the big bang. In the full sample (Table 1.4 bottom row), there are 150 crashes and 34 structural increases in capitalization. This means that around one-third of crashes and two-thirds of the structural increases occur after sharp increases in market capitalization. An observed market cap run-up, therefore, signals increasing odds of one of these two types of events. Still, even conditional on observing a run-up, mean-reversion is the norm and persistent market expansions similar to the big bang are the exception.

Figure 1.13 plots the average level of market capitalization, dividend-price ratio and cumulative return in excess of government bonds during the market cap run-up. During an average run-up, stock market capitalization increases sharply, by around 30% of GDP. This increase is accompanied by rising stock valuation metrics and high returns: the dividend-price ratio falls by 0.6 ppts – around one-sixth of its long-run mean – and excess returns during the last 2 years of the run-up average 15% per year. After the run-up however, market performance is substandard. The market cap to GDP ratio stops growing and then falls by 10% of GDP, dividend-price ratios increase and excess returns are on average zero. Still, the market cap increases do

8. Applying the exact Greenwood, Shleifer, and You (2019) definition to our aggregate return index gives fewer run-ups – around 30 – but does not change the results.

Table 1.5. Predicting equity market crashes

	(1)	(2)	(3)	(4)	(5)
$ln(MCAP_{t-1}/GDP_{t-1})$	0.60***		0.34**	2.04***	1.74***
	(0.19)		(0.16)	(0.40)	(0.41)
$\Delta_3\ ln(MCAP_{t-1}/GDP_{t-1})$		1.13***	1.17***	0.43**	0.38*
		(0.17)	(0.19)	(0.21)	(0.20)
$ln(P_{t-1}/D_{t-1})$					0.47***
					(0.17)
Country fixed effects				✓	✓
ROC	0.59	0.67	0.68	0.73	0.75
Number of Crashes	142	139	139	139	138
Observations	2139	2044	1853	2044	2027

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.

not fully mean revert, reflecting the fact that some run-ups result in persistently higher capitalization levels. The run-up aftermath is also associated with high equity market risk: within the bottom quartile of the post-run up outcomes (bottom dashed line in Figure 1.13 right panel), excess equity returns record a 40% cumulative loss. The patterns of sharply rising and then partially mean-reverting market cap, and high but then stagnant returns persist through several alternative run-up definitions shown in the Appendix Figure 1.A.8.

To more formally assess the downside risk of market cap run-ups, we test whether high levels or sharp increases in market capitalization help predict an equity market crash using the following logit model:

$$Prob(C_{j,t} = 1) = \Lambda\left[ln(MCAP_{j,t-1}/GDP_{j,t-1}), X_{j,t-1}, \beta\right], \tag{1.10}$$

where C is a crash dummy equal to 1 if there is a -25% or lower equity return in year t or, cumulatively, in years t and $t+1$, X are other predictors, β is the estimated parameter vector, Λ is the logistic distribution function, and j and t are country and time indices.

Table 1.5 reports the estimated β coefficients and standard errors for the market crash prediction. Consistent with the stylised facts in Figure 1.13, 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 in a joint regression with both variables (column 3), and when controlling for country fixed effects (column 4) and the price-dividend ratio (column 5). The ROC statistic in Table 1.5 compares our predictive model to a random sorting of observations into crash and non-crash episodes, showing that it outperforms the random sorting ROC of 0.5 by a considerable factor.[9] Appendix Table 1.A.1 shows that high, or growing, stock market cap predicts high crash risk across different time periods and crash definitions.

1.6.2 Capitalization as a valuation metric

Table 1.5 shows that market capitalization can be used to predict market crashes, and Table 1.2 shows that it can be used to predict future market returns more generally. But how does market capitalization compare to alternative valuation metrics? Warren Buffet famously called stock market capitalization "the best single measure of where valuations stand at any given moment" (Buffett and Loomis, 2001). To test whether this is the case, we run a horse-race between market cap and the price-dividend ratio – which has been shown to be a reliable predictor of stock returns in the US and internationally (Cochrane, 2008; Engsted and Pedersen, 2010; Kuvshinov, 2020) – by estimating the following predictive regressions:

$$r_{j,t+1} = \alpha_j^r + \beta^{mcap} ln(MCAP_{j,t}/GDP_{j,t}) + \beta^{pd} ln(P_{j,t}/D_{j,t}) + u_{j,t}^r \quad (1.11)$$

$$dg_{j,t+1} = \alpha_j^{dg} + \zeta^{mcap} ln(MCAP_{j,t}/GDP_{j,t}) + \zeta^{pd} ln(P_{j,t}/D_{j,t}) + u_{j,t}^{dg} \quad (1.12)$$

As before, we use a cross-country panel with standard errors adjusted for country and time clustering and autocorrelation, and with both predictor variables adjusted for structural breaks using the Bai and Perron (2003) method.

Table 1.6 shows that both market capitalization and the price-dividend ratio predict future returns (column 1), and hence subsequent mean-reversion in the market following high levels of market cap or stock prices. The market cap to GDP ratio remains an economically and statistically significant predictor across the full range of alternative specifications which, respectively, restrict the sample to the post-1985 period, add year fixed effects, predict excess returns, 5-year-ahead returns or do not adjust for structural breaks; while the price-dividend ratio becomes less powerful when it comes to predicting excess returns (column 7), long-run returns (column 9) or if we do not adjust the data for structural breaks (column 11). Across the different specifications, the coefficient on market cap is similar in size to the unconditional coefficient which does not control for the price-dividend ratio shown in Table 1.2.

Furthermore, while the price-dividend ratio responds strongly to changes in future fundamentals, market capitalization forecasts fundamentals in the wrong direction. Table 1.6 column 2 shows that a one-quarter relative increase in the price-dividend ratio predicts 6.8 ppts ($0.25 \times 0.265 \times 1.003$) higher dividend growth one year ahead, while a one-quarter increase in the market cap to GDP ratio predicts 3.2 ppts ($0.25 \times (-0.124) \times 1.003$) lower growth. The negative dividend growth predictability remains under the alternative regression specifications in columns 4, 6,

9. For further details on the application of ROC curves to the financial extreme event analysis, see Schularick and Taylor (2012).

Table 1.6. Horse race between market cap and the price-dividend ratio

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline		Post-1985		Year effects	
	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}	r_{t+1}	dg_{t+1}
$ln(MCAP_t/GDP_t)$	-0.114***	-0.124***	-0.181***	-0.038	-0.077***	-0.096*
	(0.032)	(0.040)	(0.044)	(0.069)	(0.025)	(0.053)
$ln(P_t/D_t)$	-0.096***	0.265***	-0.109***	0.248***	-0.083***	0.290***
	(0.025)	(0.033)	(0.033)	(0.051)	(0.019)	(0.033)
R^2	0.072	0.118	0.106	0.083	0.508	0.292
Observations	2076	2076	519	519	2076	2076

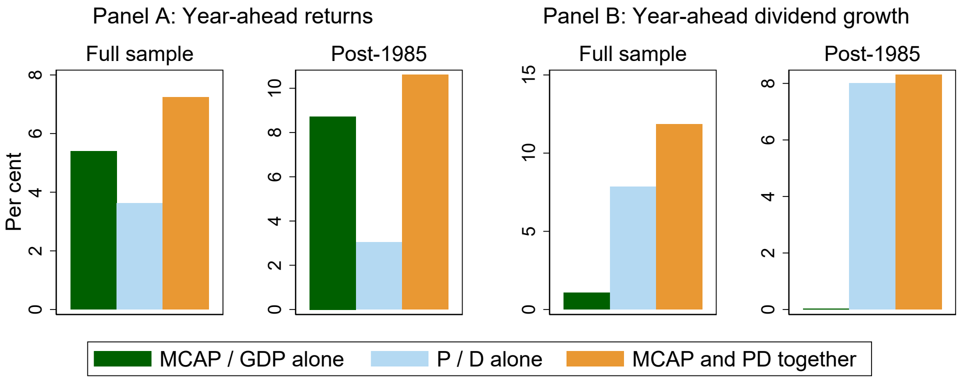
	(7)	(8)	(9)	(10)	(11)	(12)
	Risk premia and safe rates		5-year returns		No structural breaks	
	er_{t+1}	r^{safe}_{t+1}	$\bar{r}_{t+1,t+5}$	$\overline{dg}_{t+1,t+5}$	r_{t+1}	dg_{t+1}
$ln(MCAP_t/GDP_t)$	-0.096***	-0.036**	-0.086***	-0.077***	-0.026**	-0.043***
	(0.031)	(0.017)	(0.022)	(0.021)	(0.013)	(0.015)
$ln(P_t/D_t)$	-0.065*	-0.046**	-0.057***	0.142***	-0.035	0.138***
	(0.033)	(0.022)	(0.018)	(0.015)	(0.022)	(0.021)
R^2	0.039	0.020	0.167	0.205	0.021	0.066
Observations	2076	2076	1991	1991	2076	2076

Notes: Market cap and price-dividend ratio are adjusted for structural breaks. r is total real return, dg is real dividend growth, er is excess return and r^{safe} is real government bond return, all measured in logs. Regressions with country fixed effects. Columns (5) and (6) additionally add year fixed effects. Standard errors clustered by country and year and adjusted for autocorrelation are in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

10 and 12, but becomes somewhat weaker after 1985. These findings can be understood in light of the importance of "profit shifts" – or increases in *current* listed firms dividends to GDP – in driving market capitalization changes. High market capitalization is associated with high current dividend payments, but some of these higher payments eventually mean revert, resulting in the negative future dividend growth.

Figure 1.14 shows the R^2 statistics for forecasting year-ahead returns and dividends using market cap alone (dark-green bars), price-dividend ratio alone (light-blue bars) and both metrics together (orange bars, equivalent to the specification in Table 1.6), both for the full sample and the post-1985 period. The return R^2 of market cap alone is higher than that of the price-dividend ratio, and the R^2 of both metrics together is only slightly higher than that of market capitalization alone, with the predictive power of market cap becoming stronger and that of the price-dividend ratio – weaker – after 1985. When it comes to predicting dividend growth, on the contrary, market capitalization alone has almost no predictive power, whereas the

Figure 1.14. Predictive power of market capitalization and the price-dividend ratio



Note: Comparison of R^2 statistics from predictive regressions. The dependent (y) variables are the year-ahead log real total return (Panel A) and log real dividend growth (Panel B). Different coloured bars correspond to different sets of predictors.

price-dividend ratio has considerable predictive power similar to that of the two metrics together.

Why does market capitalization do so well at predicting returns? Compared to the price-dividend ratio, the market cap to GDP ratio carries two theoretical advantages. First, the denominator of the ratio is GDP rather than dividends, allowing it to look through changes in dividend policy which can obscure the underlying predictive relationships (Chen, Da, and Priestley, 2012). Second, capitalization includes quantities as well as prices, allowing it to potentially better capture movements in market sentiment and investors timing the market to issue stocks when valuations are high (Baker and Wurgler, 2000). To test the importance of these two channels, Table 1.7 tests the predictive power of the stock price to GDP ratio and of equity issuance relative to GDP.

The top panel of Table 1.7 predicts future stock returns and dividends using today's price-GDP ratio, constructed following Rangvid (2006) as the ratio of the nominal stock price index to nominal GDP.[10] The price-GDP ratio is a statistically and economically significant predictor of future stock returns. A one standard deviation increase in the price-GDP ratio forecasts 5.4 percentage points lower real equity returns one year ahead.[11] Like the market cap to GDP ratio, high stock prices relative to GDP also forecast low rather than high dividend growth. The bottom panel of Table 1.7 regresses future returns and dividends on the three-year back-

10. 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. We use separate trends before and after 1985 to account for the big bang structural break.

11. The standard deviation of the detrended log price-GDP ratio is 0.65, giving a $0.65 \times (-0.079) \times 1.048 = 0.054$ predicted return decline.

Table 1.7. Price-GDP ratio and issuance as predictors of equity returns and dividends

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline		5-year returns		No structural breaks	
	r_{t+1}	dg_{t+1}	$\bar{r}_{t+1,t+5}$	$\overline{dg}_{t+1,t+5}$	r_{t+1}	dg_{t+1}
Panel 1: *Price-GDP ratio as a predictor*						
$ln(P_t/GDP_t)$	-0.079***	-0.105***	-0.085***	-0.081***	-0.046***	-0.051**
	(0.022)	(0.022)	(0.016)	(0.015)	(0.017)	(0.022)
$ln(P_t/D_t)$	-0.092***	0.270***	-0.041*	0.159***	-0.034*	0.125***
	(0.029)	(0.034)	(0.021)	(0.018)	(0.020)	(0.020)
R^2	0.070	0.126	0.261	0.285	0.037	0.065
Observations	2205	2205	2119	2119	2205	2205
Panel 2: *Implied equity issuance as a predictor*						
Implied issuance / GDP	-0.579	-0.135	-0.321*	-0.148	-0.598*	-0.146
	(0.358)	(0.309)	(0.175)	(0.212)	(0.324)	(0.313)
$ln(P_t/D_t)$	-0.145***	0.209***	-0.092***	0.104***	-0.051***	0.104***
	(0.027)	(0.044)	(0.021)	(0.018)	(0.018)	(0.019)
R^2	0.048	0.067	0.089	0.095	0.020	0.045
Observations	1990	1990	1907	1907	1990	1990

Notes: The price-dividend ratio is adjusted for structural breaks, and the price-GDP ratio is detrended using separate trends for before and after 1985. The no structural breaks price-GDP ratio in columns (5)–(6) uses one trend for the whole sample. r is total real return, dg is real dividend growth, er is excess return and r^{safe} is real government bond return, all measured in logs. Regressions with country fixed effects. Standard errors clustered by country and year and adjusted for autocorrelation are in parentheses. *, **, ***: Significant at 10%, 5% and 1% levels respectively.

ward moving average of implied issuance to GDP. In line with the sentiment literature (Baker and Wurgler, 2000), high equity issuance predicts low future equity returns. But the magnitude of the predicted return decrease is relatively small – around 0.6 percentage points for every 1 percentage point increase in issuance to GDP ($0.01 \times 0.58 \times 1.048$) – and insignificant under some specifications. This suggests that swings in issuance and the associated sentiment variation play a relatively modest role in driving the short-run variation in market cap.

Our evidence confirms Warren Buffet's intuition that market capitalization is a good metric of where stock market valuations stand. It also tells us that a string of favourable past shocks – such as the profit shift occurring in the 1990s – does not necessarily mean favourable prospects for the market going forward. On the contrary, these favourable shocks and high market cap are often followed by mean reversion – both in terms of low returns, and lower growth in future dividends.

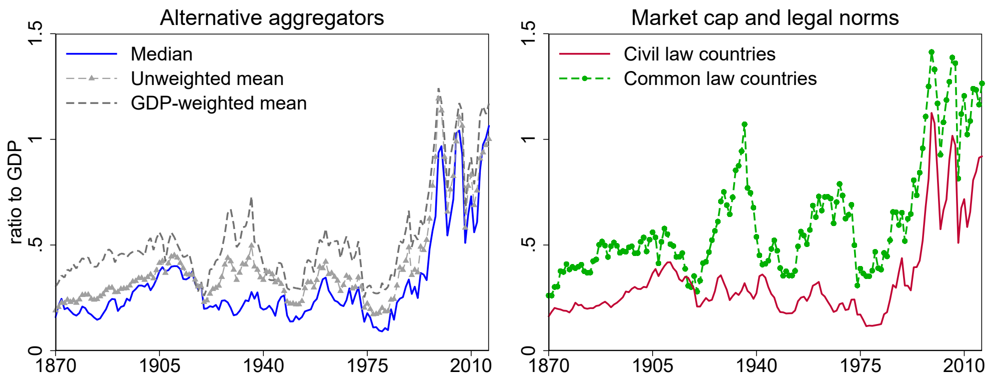
1.7 Conclusion

This paper has introduced a new dataset of stock market capitalization in 17 advanced economies covering years 1870 to 2016. A deeper exploration of the data has revealed several novel facts. The ratio of stock market size to real activity follows a hockey-stick-like pattern: flat for the first century of our sample followed by sharp persistent increases over the past three decades. The post-1980s expansion in stock market size is historically unprecedented, but it is not driven by higher equity issuance or new listings. Instead, it represents a persistent increase in stock prices which took place alongside stagnating economic growth. We show that these increases are driven by a profit shift – a marked rise in listed firms' profits as a share of both GDP and capital income. The existence of this profit shift is consistent with the broader trend of increasing market power of large firms at an increasingly uneven distribution of corporate earnings in the US and globally (De Loecker and Eeckhout, 2018; De Loecker, Eeckhout, and Unger, 2020). Because these high market values reflect a distributional shift within current income rather than a high future growth potential, they do not generally signal favourable near-term prospects for the economy. On the contrary, we show that high levels of market capitalization are typically a sign of brewing trouble, predicting low returns, low growth, and a high probability of a stock market crash.

Appendix 1.A Additional results

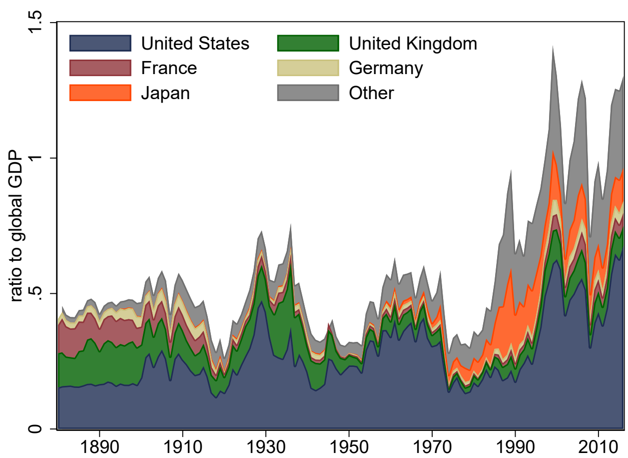
1.A.1 Trends in market capitalization

Figure 1.A.1. Alternative aggregators and legal norms



Notes: Right-hand panel: median market cap to GDP ratios for two groups of countries. Common law countries are Australia, Canada, the UK and the US. Civil law countries are all other countries in our dataset.

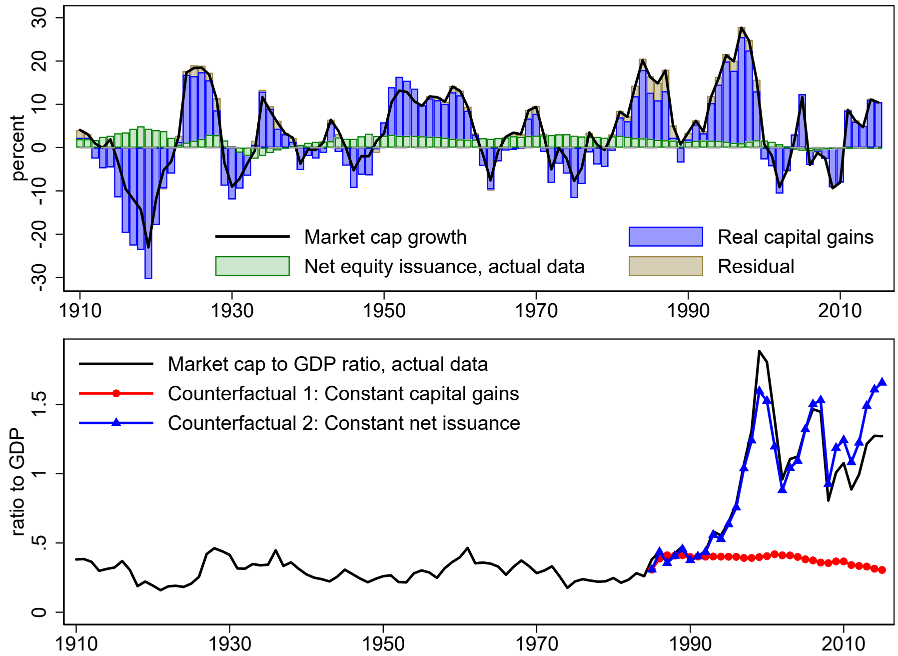
Figure 1.A.2. Global market capitalization



Notes: The ratio of advanced-economy market capitalization to advanced-economy GDP (both 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 total advanced-economy GDP.

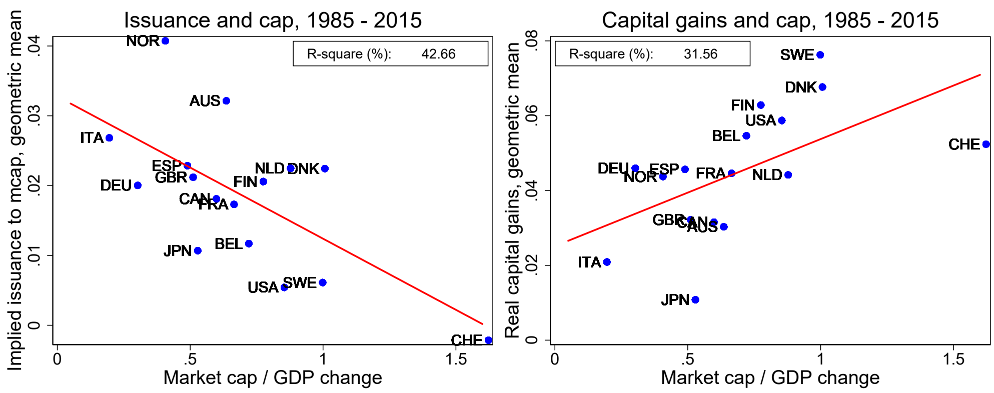
1.A.2 Prices versus Quantities: additional material

Figure 1.A.3. Decomposition trends and counterfactual: actual net issuance data



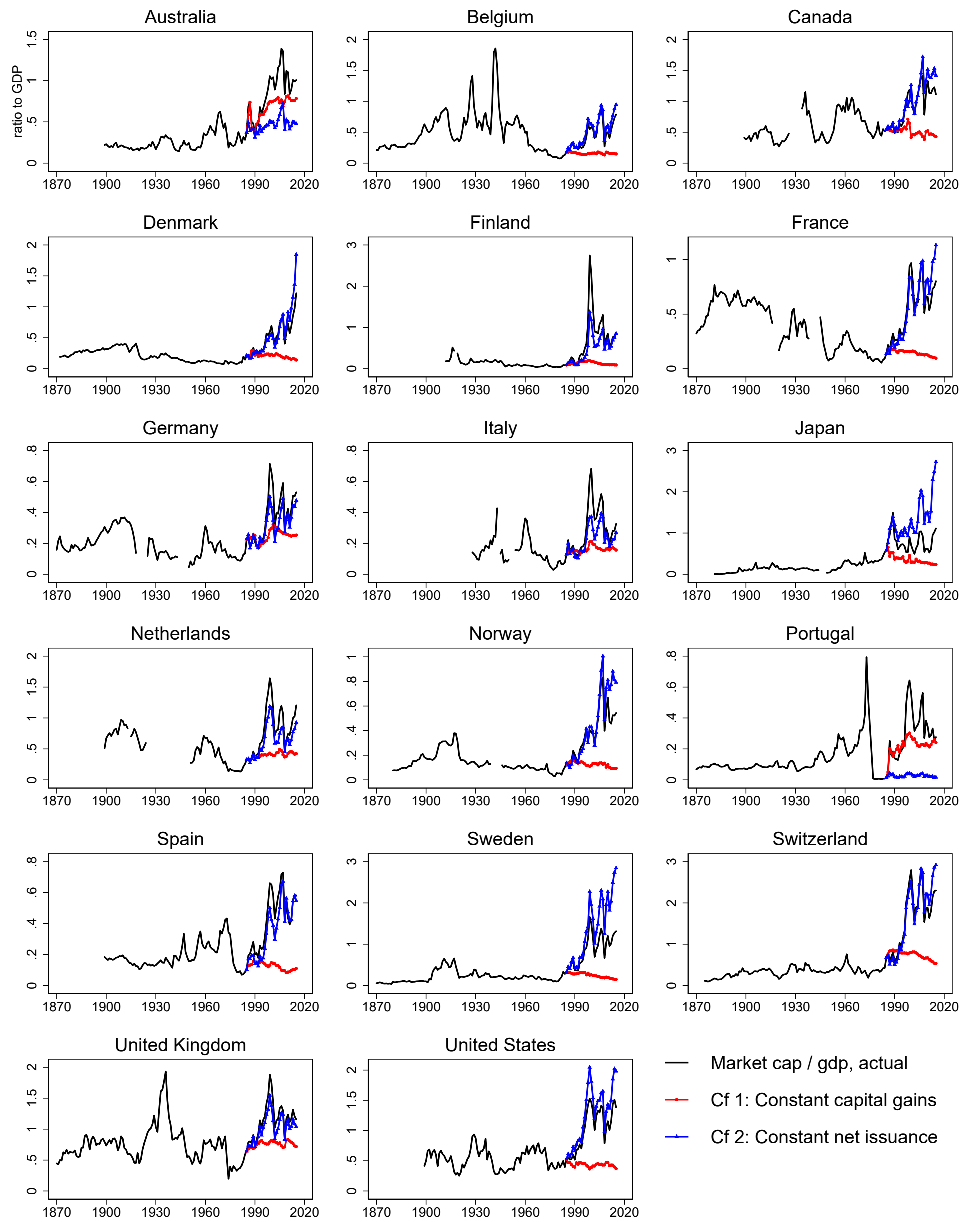
Notes: Unweighted averages of 4 countries: Finland, Germany, Switzerland and USA. Top panel: decomposition of real market cap growth into capital gains, net issuance and a residual; using actual net issuance and capital gains data. Centered 5-year moving averages. Bottom panel: counterfactual market 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. Data benchmarked so that the combined growth of the two counterfactuals between 1985 and 2015 equals the actual growth in observed market cap data.

Figure 1.A.4. Post-1985 growth in market cap, issuance and capital gains by country



Notes: Country-level changes in market cap and average real capital gains and net issuance relative to previous year's market cap between 1985 and 2015. We omit Portgual for illustrative purposes, but including it in the graph does not change the correlations.

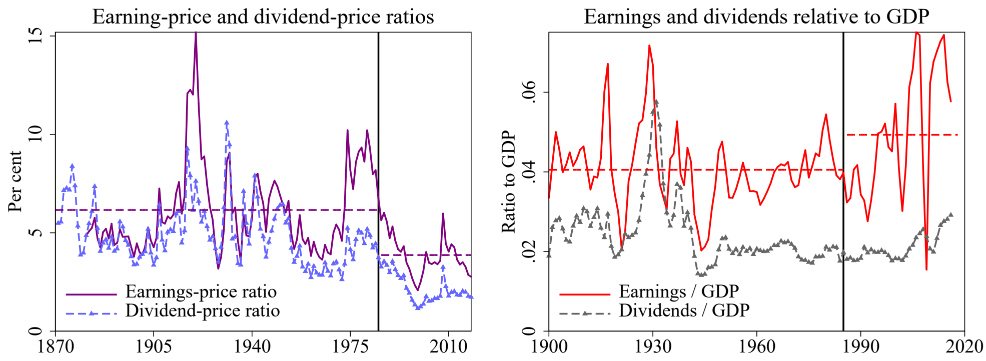
Figure 1.A.5. Counterfactual evolution of market cap in individual countries



Notes: 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 for the specific country. Constant net issuance counterfactual forces net issuance relative to market cap during 1985–2015 to equal the pre-1985 average. Data are benchmarked so that the combined growth of the two counterfactuals between 1985 and 2015 equals the actual growth in observed market cap data.

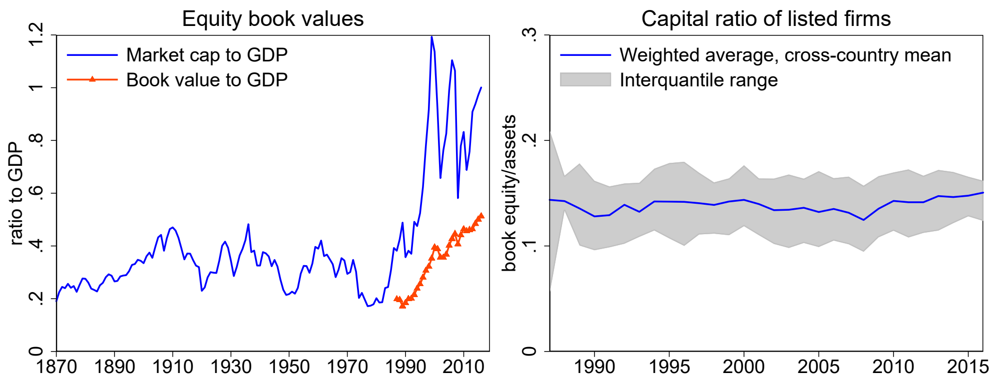
1.A.3 Drivers of long-run stock market growth: additional material

Figure 1.A.6. Earnings and dividends of listed firms in the US



Notes: Left-hand panel: dividend-price ratio and the cyclically adjusted total return earnings-price ratio (inverse of P/E10 CAPE) from Shiller (2015), December values. Right-hand panel: earnings to GDP calculated as market cap to GDP times the earnings-price ratio. Black vertical line in 1985 signifies the big bang. Dashed horizontal lines show the averages of the earning-price ratio series (left-hand panel) and earnings-GDP series (right-hand panel) before and after the big bang.

Figure 1.A.7. Book equity values and firm leverage

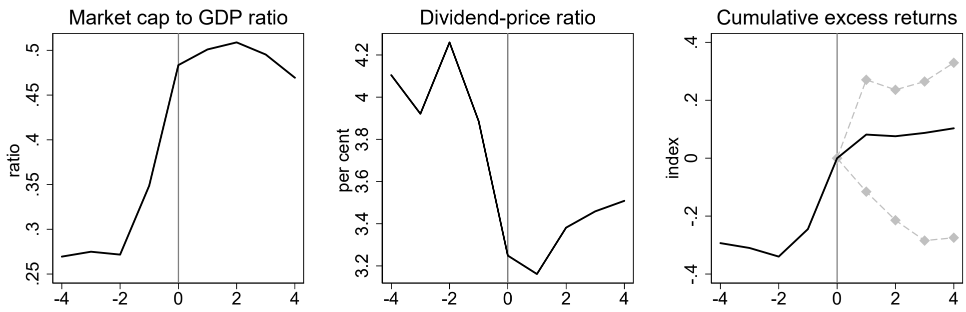


Notes: Unweighted averages of 17 countries. Left-hand panel: market and book value of listed firm equity. Right-hand panel: capital ratio of listed firms. The capital ratio is calculated by dividing book equity by total firm assets. Book equity and total assets data are aggregated up from Compustat Global and Compustat North America and cover all listed firms with non-missing values for market cap, dividends and earnings, scaled up to match our aggregate market cap data where necessary. We drop country-year observations at the beginning of the sample with less than than 30% of market cap covered. The data include both financial and non-financial firms.

1.A.4 Buffet indicator: additional material

Figure 1.A.8. Alternative definitions of stock market run-ups

(a) Relative market cap run-ups



(b) Stock return run-ups

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 50% or more relative increase in market cap to GDP over 2 years, and 25% or more relative increase over 5 years. Panel (b): run-up defined as a cumulative real total return of 80% or more over 2 years, and 40% or more over 5 years. Excess returns indexed to 0 at $t = 0$.

Table 1.A.1. Predicting equity market crashes: alternative specifications

	(1) Pre 1945	(2) Post 1945	(3) No wars	(4) Decade FEs	(5) Large crashes	(6) 1-year crashes	(7) 3-year crashes	(8) MCAP crashes
$ln(MCAP_{t-1}/GDP_{t-1})$	3.31***	2.16***	2.55***	2.08***	3.98***	1.95***	2.72***	3.25***
	(1.19)	(0.39)	(0.44)	(0.36)	(0.55)	(0.31)	(0.46)	(0.56)
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
ROC	0.75	0.73	0.74	0.78	0.88	0.71	0.74	0.78
Number of Crashes	45	97	121	142	32	102	118	150
Observations	953	1186	1943	2072	1887	2139	2139	2139
	(1) Pre 1945	(2) Post 1945	(3) No wars	(4) Decade FEs	(5) Large crashes	(6) 1-year crashes	(7) 3-year crashes	(8) MCAP crashes
Δ_3 $ln(MCAP_{t-1}/GDP_{t-1})$	1.35***	1.06***	1.27***	0.94***	1.39***	0.70***	1.51***	1.24***
	(0.49)	(0.16)	(0.19)	(0.16)	(0.26)	(0.15)	(0.22)	(0.22)
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
ROC	0.71	0.69	0.70	0.76	0.79	0.64	0.72	0.70
Number of Crashes	44	95	118	139	31	100	111	146
Observations	886	1158	1853	2003	1803	2044	2044	2044

Note: Dependent variable is the equity market crash dummy at time *t*. In columns (1)–(4), crash is defined as real equity returns falling by more than 25% in one year or within a two year window, with no crashes in the two previous years. Column (1) restricts the panel to observations before 1945. Column (2) only includes observations after 1945. Column (3) removes observations from the world wars. Column (4) reports estimates with decade fixed effects. Columns (5) to (8) 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, keeping the same thresholds (25% drop). All estimates are based on logit estimations with country fixed effects and country clustered standard errors. *, **, ***: Significant at 10%, 5% and 1% levels respectively. Standard errors in parentheses.

Appendix 1.B 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

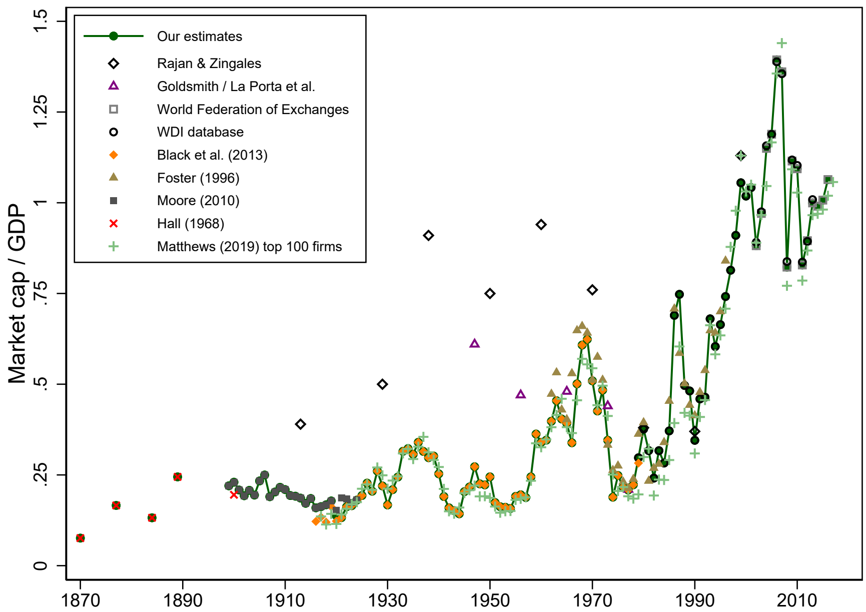
Table 1.B.1 documents the sources of our stock market capitalization data for Australia, and Figure 1.B.1 plots the resulting series alongside alternative existing estimates. The Australian securities market has generally been dominated by two major stock exchanges, located in Sydney and Melbourne. Hall (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, Kirkwood, Williams, and Rai (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, Kirkwood, Williams, and Rai (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

Table 1.B.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, Kirkwood, Williams, and Rai (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 1.B.1. Australia: alternative stock market cap estimates



companies (again, up to the 1970s). These two biases do, however, largely seem to 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.[12]

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 1.B.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, Kirkwood, Williams, and Rai (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.

12. 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 1.B.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 1.B.2. Belgium: alternative stock market cap estimates

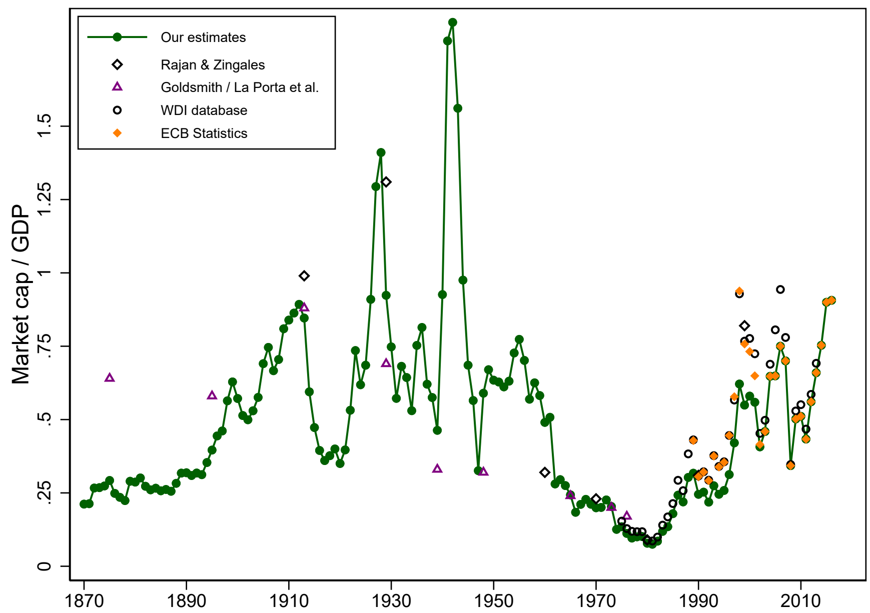


Table 1.B.2 documents the sources of our stock market capitalization data for Belgium, and Figure 1.B.2 plots the resulting series alongside alternative existing estimates. The data cover the Brussels stock exchange, which was the dominant stock exchange throughout the data coverage period in our paper, 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.

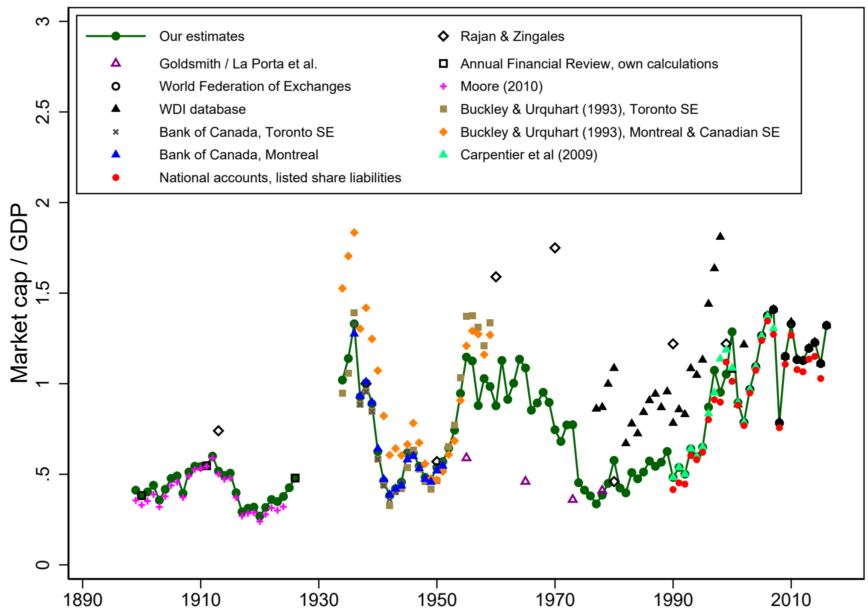
Canada

Table 1.B.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 (2020), 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 (2020), 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 1.B.3 documents the sources of our stock market capitalization data for Canada, and Figure 1.B.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 1.B.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 1.B.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 1.B.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.[13] 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 1.B.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-

13. 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 1.B.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 (2020), 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 1.B.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.[14] 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 1.B.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 1.B.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.

14. 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

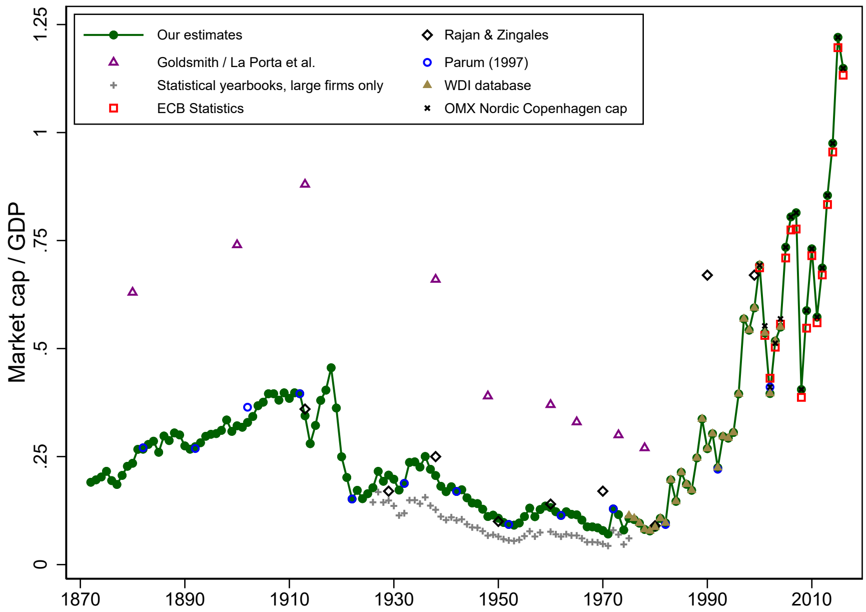
Table 1.B.4. Data sources: Denmark

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 1.B.4 documents the sources of our stock market capitalization data for Denmark, and Figure 1.B.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 1.B.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 1.B.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 1.B.5. Finland: alternative stock market cap estimates

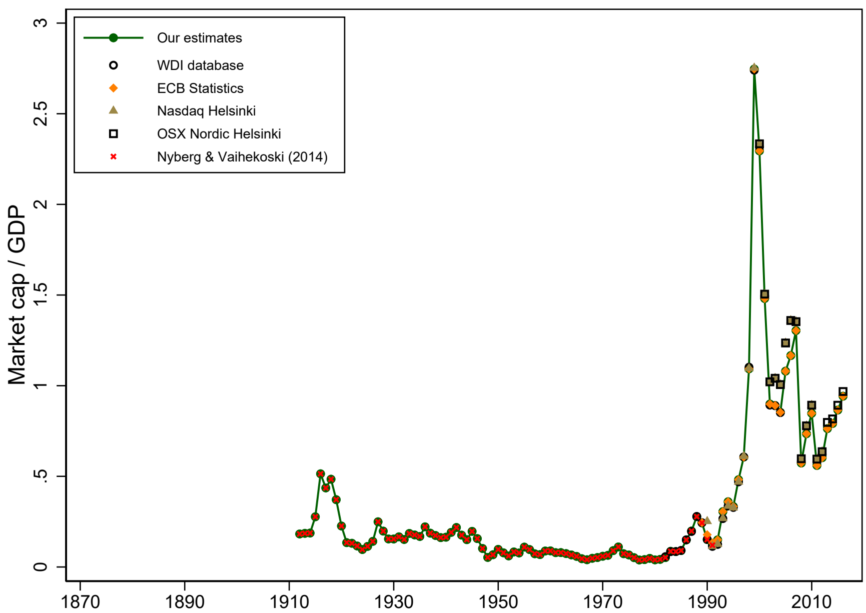


Table 1.B.5 documents the sources of our stock market capitalization data for Finland, and Figure 1.B.5 plots the resulting series alongside alternative existing estimates. The long-run data come from Nyberg and Vaihekoski (2014a), who have compiled a database of returns and capitalization on all stocks 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 1.B.5. The modern data from the ECB series are very close to Helsinki stock exchange capitalization estimates from Nasdaq and OMX Nordic (Figure 1.B.5).

We are grateful to Mika Vaihekoski for sharing data and providing help and support in locating the sources for Finland.

France

Table 1.B.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 1.B.6. France: alternative stock market cap estimates

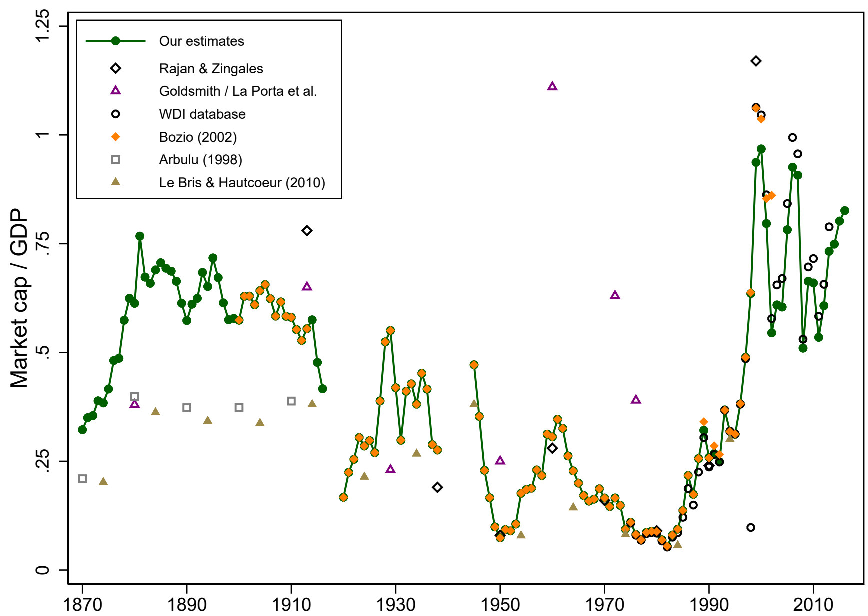


Table 1.B.6 documents the sources of our stock market capitalization data for France, and Figure 1.B.6 plots the resulting series alongside alternative existing estimates. Most of the data are drawn from the comprehensive study of Bozio (2002), which estimated the total capitalization of French shares 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.[15] 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 1.B.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.

15. 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.

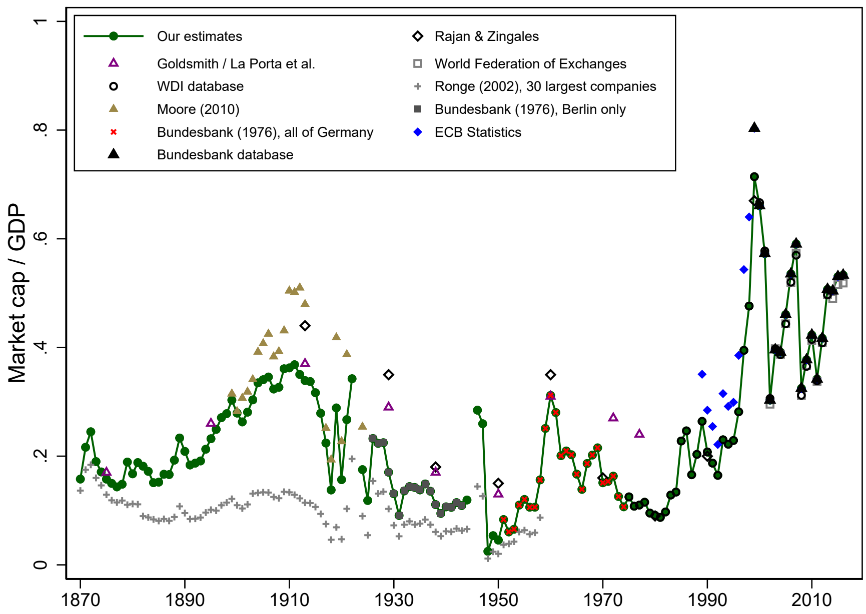
Germany

Table 1.B.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 1.B.7 documents the sources of our stock market capitalization data for Germany, and Figure 1.B.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 1.B.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 1.B.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 1.B.8. Italy: alternative stock market cap estimates

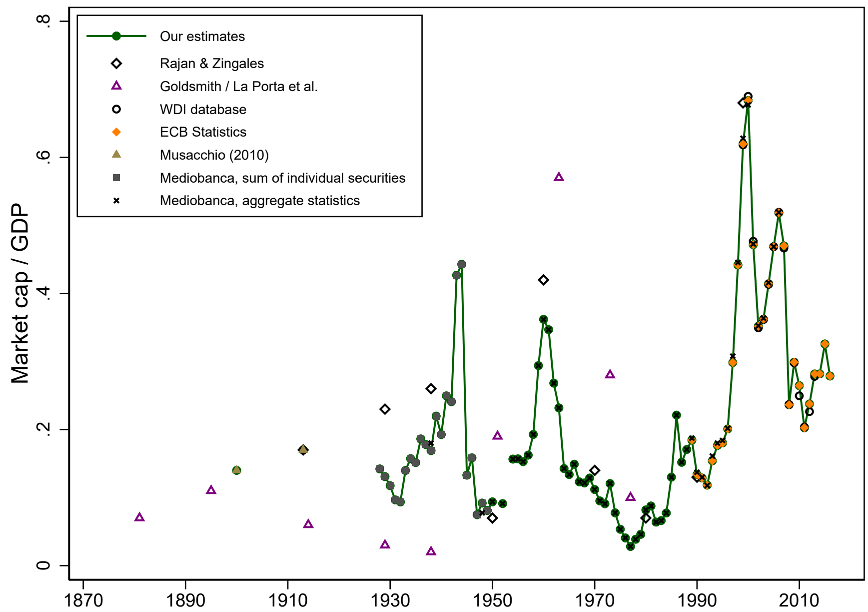


Table 1.B.8 documents the sources of our stock market capitalization data for Italy, and Figure 1.B.8 plots the resulting series alongside alternative existing estimates. Most of the data are sourced from the *Indici e Dati* publication, Mediobanca (Various years), which presents various aggregate and 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 1.B.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 1.B.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, Kuvshinov, Schularick, and Taylor (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, Kuvshinov, Schularick, and Taylor (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 1.B.9. Japan: alternative stock market cap estimates

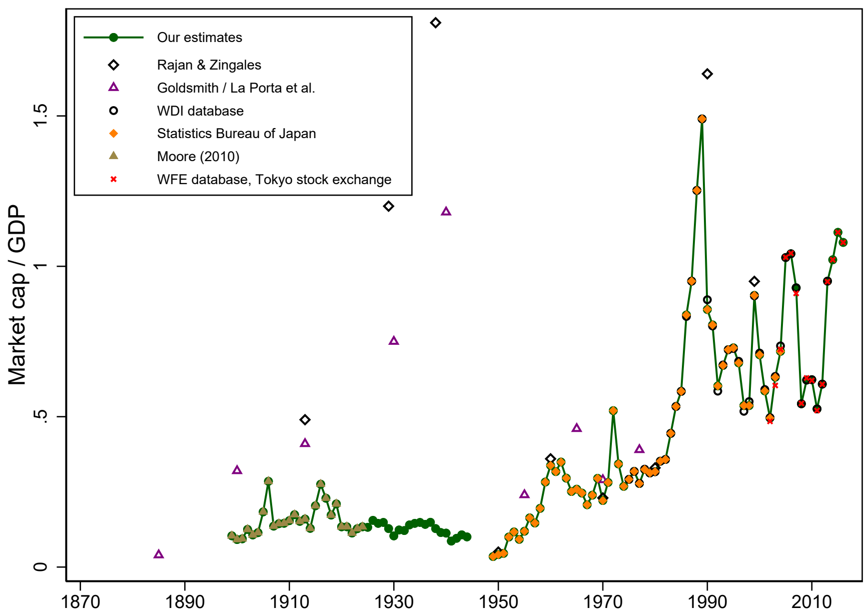


Table 1.B.9 documents the sources of our stock market capitalization data for Japan, and Figure 1.B.9 plots the resulting series alongside alternative existing estimates. For the early historical period, our main source are the Moore (2010b) estimates of the total capitalization of the Tokyo stock exchange. 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, Kuvshinov, Schularick, and Taylor (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 in market cap 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 1.B.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 1.B.10. Netherlands: alternative stock market cap estimates

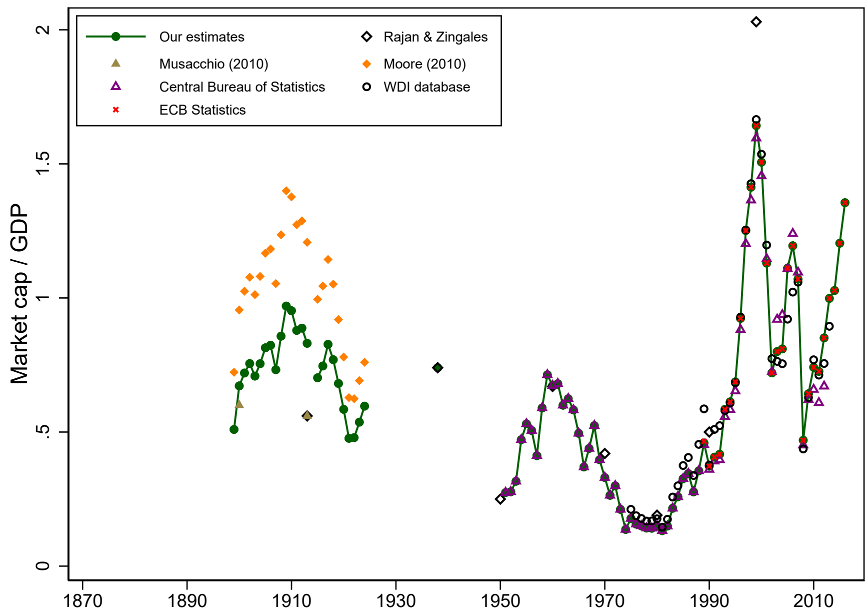


Table 1.B.10 documents the sources of our stock market capitalization data for the Netherlands, and Figure 1.B.10 plots the resulting series alongside alternative existing estimates. In the early period, our main source are the Moore (2010b) estimates of the total capitalization of the Amsterdam stock 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

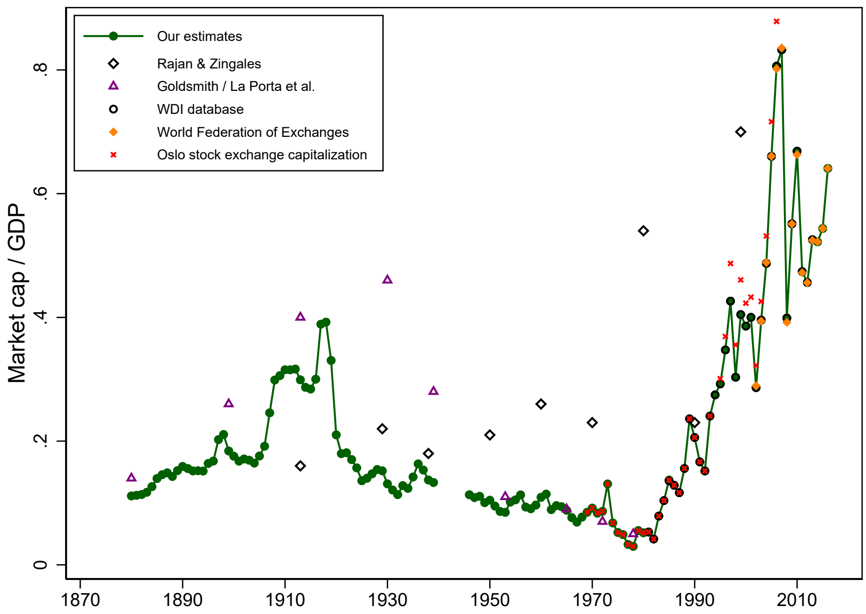
Table 1.B.11. Data sources: Norway

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 1.B.11 documents the sources of our stock market capitalization data for Norway, and Figure 1.B.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 1.B.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 1.B.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 1.B.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 1.B.12. Portugal: alternative stock market cap estimates

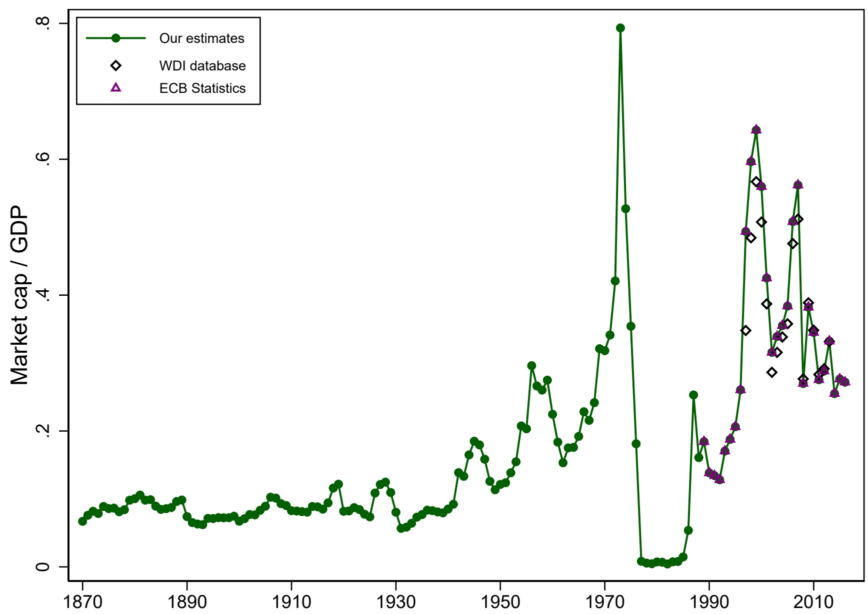


Table 1.B.12 documents the sources of our stock market capitalization data for Portugal, and Figure 1.B.12 plots the resulting series alongside alternative existing estimates. Very few estimates of the Portuguese market capitalization exist, particularly for the period before 1990. Therefore we construct 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 1.B.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, Kuvshinov, Schularick, and Taylor (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 1.B.13. Spain: alternative stock market cap estimates

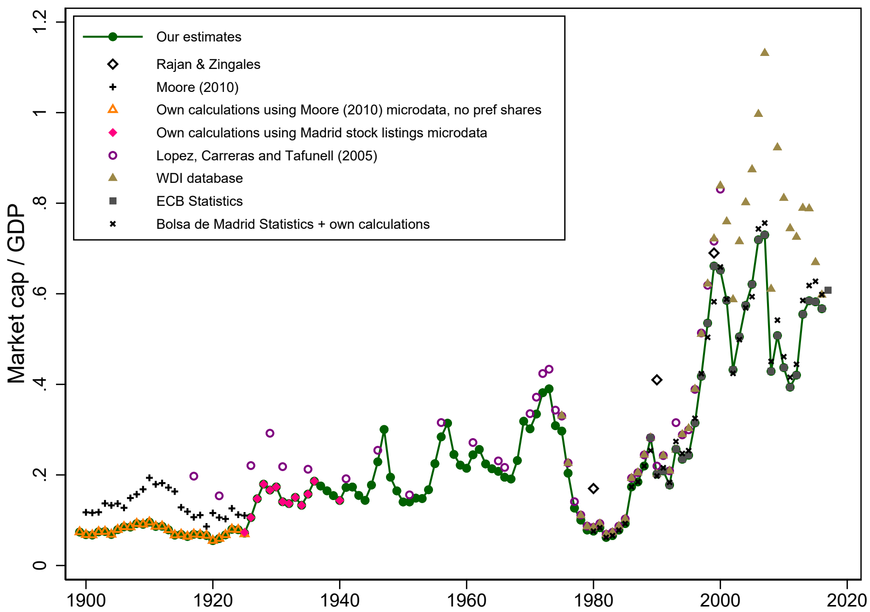


Table 1.B.13 documents the sources of our stock market capitalization data for Spain, and Figure 1.B.13 plots the resulting series alongside alternative existing estimates. For the early historical period, we construct estimates of total capitalization of or-

dinary 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, Kuvshinov, Schularick, and Taylor (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 1.B.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 1.B.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 1.B.14. Sweden: alternative stock market cap estimates

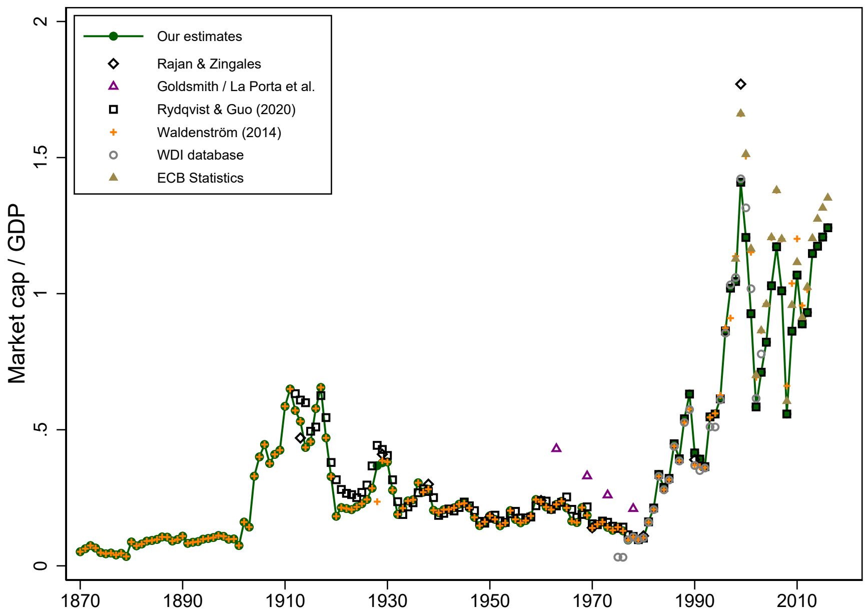


Table 1.B.14 documents the sources of our stock market capitalization data for Sweden, and Figure 1.B.14 plots the resulting series alongside alternative existing estimates. The main source for our series are the data compiled by Waldenström (2014), who put together a long-run series of Swedish stock 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 1.B.15. Data sources: Switzerland

Year	Data source
1875–1970	Total market capitalization of all Swiss firms listed in Zurich, own estimates. For 1875–1898 and 1925–1970, we digitise the stock listings of the Zurich exchange (Kursblatt der Züricher Effektenbörse), complemented by data on individual company accounts, and compute the sum of capitalizations of all Swiss companies. For 1899–1925, we use microdata helpfully shared by Lyndon Moore (Moore, 2010a,b). We scale up the pre-1899 series to match the Lyndon Moore estimates. To match the WDI market cap value in 1975, we further scale up the annual cap growth rate by 1 ppt.
1971–1974	1970 stock market cap extrapolated forward using net issuance data from the Swiss National Bank *Capital Market statistics*, with growth rates adjusted up by 1% to match the 1975 WDI cap value.
1975–1979	Total capitalization of all Swiss listed firms, shares listed on Swiss exchanges, from the *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 1.B.15. Switzerland: alternative stock market cap estimates

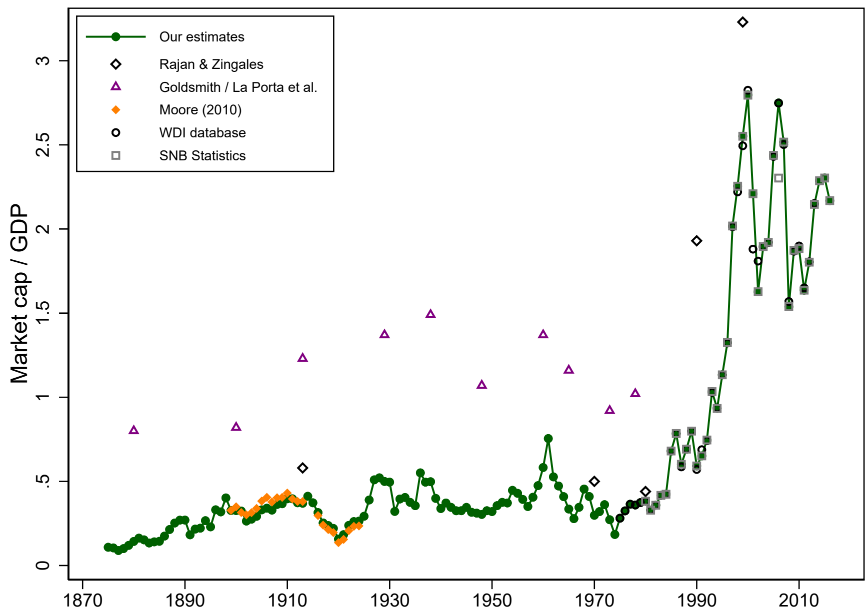


Table 1.B.15 documents the sources of our stock market capitalization data for Switzerland, and Figure 1.B.15 plots the resulting series alongside alternative existing estimates. The early estimates of the Swiss stock market capitalization are based

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, Kuvshinov, Schularick, and Taylor (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

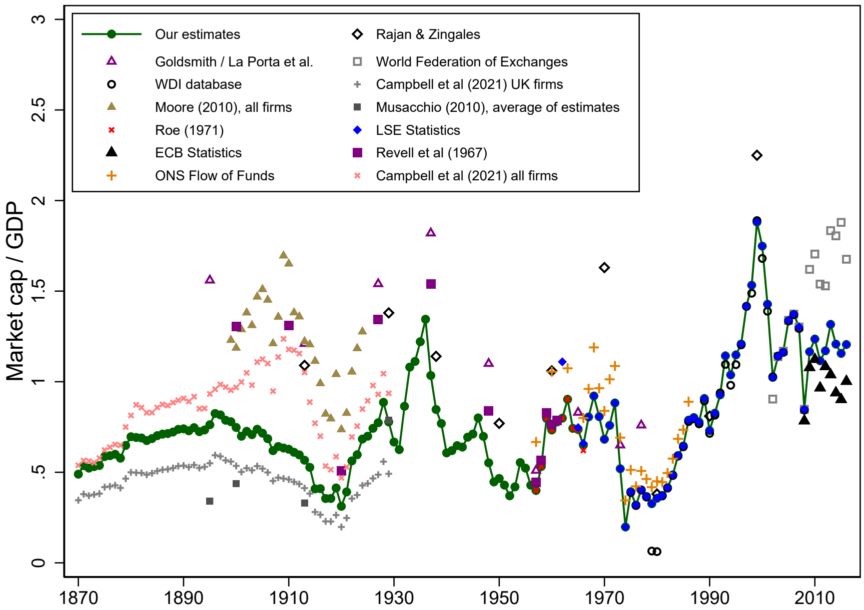
Table 1.B.16. Data sources: United Kingdom

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 1.B.16 documents the sources of our stock market capitalization data for the United Kingdom, and Figure 1.B.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 1.B.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 1.B.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).[16]

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 1.4 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 1.B.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 1.B.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 1.B.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 gener-

16. Musacchio (2010) recognises the difficulty of estimating the UK stock market capitalization precisely, and offers a range of estimates.

ally close to our data and the estimates from the LSE, but overall seem somewhat 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 1.B.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

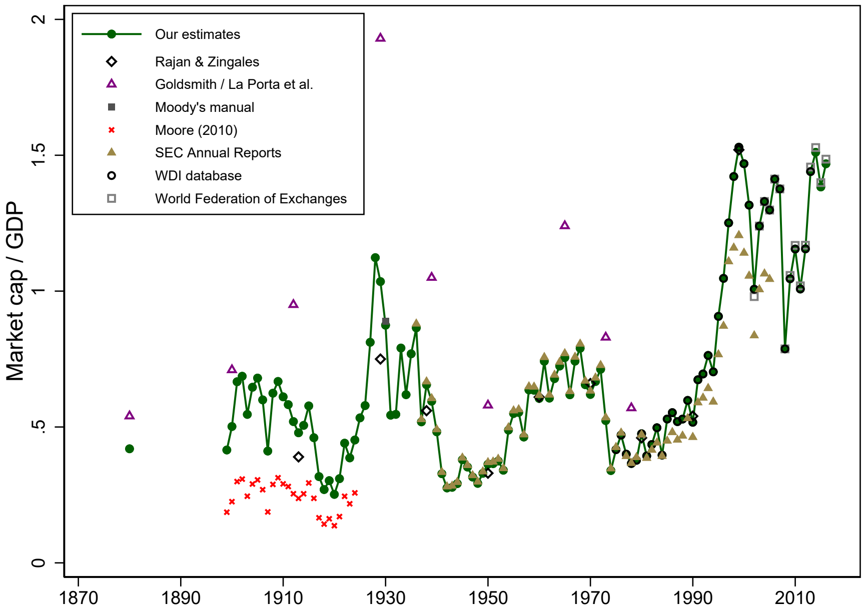
Table 1.B.17. Data sources: United States

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 1.B.17 documents the sources of our stock market capitalization data for the United States, and Figure 1.B.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 1.B.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 1.B.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.

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.

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.

Bank of Japan. 1966. *Hundred-Year Statistics of the Japanese Economy*.

Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75 (5): 2421–2463.

Barkai, Simcha, and Seth G. Benzell. 2018. "70 Years of US Corporate Profits." Stigler Center for the Study of the Economy and the State, New Working Paper Series 22.

Bengtsson, Erik, and Daniel Waldenström. 2018. "Capital Shares and Income Inequality: Evidence from the Long Run." *Journal of Economic History* 78 (3): 712–743.

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.

Blanchard, Olivier J. 1993. "Movements in the Equity Premium." *Brookings Papers on Economic Activity* 1993 (2): 75–138.

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.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2017. "Rents, Technical Change, and Risk Premia Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares." *American Economic Review: Papers & Proceedings* 107 (5): 614–620.

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): 287–310.

Central Bureau of Statistics. 2010. *111 Jaar Statistiek in Tijdreeksen, 1899–2010*. Den Haag: Statistics Netherlands.

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.

De Loecker, Jan, and Jan Eeckhout. 2018. “Global Market Power.” NBER Working Paper 24768.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. “The Rise of Market Power and the Macroeconomic Implications.” *Quarterly Journal of Economics* 135 (2): 561–644.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2019. “Global Trends in Interest Rates.” *Journal of International Economics* 118: 248–262.

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.

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.

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.

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.” *Brookings Papers on Economic Activity* 2018 (2): 147–250.

Feldstein, Martin, and Lawrence Summers. 1977. “Is the Rate of Profit Falling?” *Brookings Papers on Economic Activity* 1977 (1): 211–228.

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. 2019. “How the Wealth Was Won: Factors Shares as Market Fundamentals.” NBER Working Paper 25769.

Greenwood, Robin, Andrei Shleifer, and Yang You. 2019. “Bubbles for Fama.” *Journal of Financial Economics* 131 (1): 20–43.

Grossman, Richard S. 2002. “New Indices of British Equity Prices, 1870–1913.” *Journal of Economic History* 62 (1): 121–146.

Grullon, Gustavo, and Roni Michaely. 2002. “Dividends, Share Repurchases, and the Substitution Hypothesis.” *Journal of Finance* 57 (4): 1649–1684.

Gutiérrez, Germán, and Sophie Piton. 2020. “Revisiting the Global Decline of the (Non-housing) Labor Share.” *American Economic Review: Insights* 2 (3): 321–338.

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* 134 (3): 1225–1298.

Jordà, Òscar, Björn Richter, Moritz Schularick, and Alan M. Taylor. 2021. "Bank Capital Redux: Solvency, Liquidity, and Crisis." *Review of Economic Studies* 88 (1): 260–286.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual* 31 (1): 213–263.

Karabarbounis, Loukas, and Brent Neiman. 2013. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.

Kuvshinov, Dmitry. 2020. "The Co-Movement Puzzle." Working Paper.

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2020. "The Expected Return on Risky Assets: International Long-run Evidence." Working Paper.

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* 52 (3): 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, 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.

Lettau, Martin, and Stijn Van Nieuwerburgh. 2008. "Reconciling the Return Predictability Evidence." *Review of Financial Studies* 21 (4): 1607–1652.

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."

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.

Nordhaus, William D. 1974. "The Falling Share of Profits." *Brookings Papers on Economic Activity* 1974 (1): 169–217.

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.

Philippon, Thomas. 2019. *The Great Reversal: How America Gave Up on Free Markets*. 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.

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.

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): 1029–1061.

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. 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, 223–293.

Chapter 2

The Expected Return on Risky Assets: International Long-run Evidence*

Joint with Dmitry Kuvshinov

2.1 Introduction

Safe interest rates have declined markedly over the past 30 years (Holston, Laubach, and Williams, 2017). But households and firms cannot raise funds at the government borrowing rate. This makes the expected risky return – the sum of the safe rate and the market risk premium – a key input into most economic decisions. But despite some evidence of recent divergence between expected risky and safe returns (Caballero, Farhi, and Gourinchas, 2017b), we know little about their joint evolution over the long run.

This paper studies long-run trends in the expected risky return and its relationship with the safe rate. We use new data from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) to estimate the expected return directly, as the sum of expected yield and long-run cashflow growth. While previous studies have focussed on the US stock market (Blanchard, 1993; Fama and French, 2002), our estimates cover 17 countries, two major asset classes – equity and housing – and 145 years. Our direct expected return estimate has several advantages relative to averages of past realised returns examined in much previous work (for example, Jordà, Knoll, Kuvshinov, Schularick, and Taylor, 2019). First, it looks through large information surprises and unexpected shocks which tend to dominate the realised return series (Elton, 1999). The volatility of our expected return measure is an order of magnitude lower than that of realised returns. Second, it gives us a forward-looking estimate of the

* We would like to thank Marco Del Negro, Kinda Hachem, Björn Richter, Moritz Schularick, Alp Simsek, Alan Taylor, and the seminar participants at Bonn, CREI, the BGSE Webinar on Asset Prices, Finance and Macroeconomics and the Verein für Socialpolitik 2020 annual meeting for helpful comments and suggestions.

rate of return required by potential investors in risky projects. It is this rate, rather than returns realised by investors who already hold the asset, which characterises the financing conditions of households and firms.

We find that the expected risky return has been declining steadily throughout the last 145 years. This decline is largely unrelated to movements in the real safe rate and, as a consequence, the risk premium exhibits large secular variation. The risky-safe rate disconnect carries important implications for asset pricing theory. Standard theory puts forward two key drivers of expected returns: growth and risk. These entail opposing predictions for the relationship between risky and safe rates. While the growth channel pushes risky and safe rates in the same direction by affecting the general willingness to save, the risk channel pushes safe rates and risk premia in opposite directions with ambiguous overall effect on the risky rate.

We show that risk premia and safe rates are, in general, strongly negatively correlated. This suggests that risk, rather than growth, is the key driver of expected risky and safe returns over the long run. Consistent with this view, we show that secular movements in the risk premium can be explained by changes in macroeconomic risk in ways consistent with standard theory (Lettau, Ludvigson, and Wachter, 2008). We document that consumption volatility halved between 1870 and 1990, rationalising the observed 50% decline in the risk premium. After 1990, risk premia increased somewhat – a trend that can be rationalised by sharp increases in macroeconomic tail risks and higher co-movement between risky asset returns and the macroeconomy. Existing literature puts risk as a key driver of short-run variation in expected returns (Cochrane, 2017; Pflueger, Siriwardane, and Sunderam, 2020). Our findings suggest that risk is key not only in the short-run, but also when it comes to long-run trends in both risky and safe interest rates.

We start by documenting the trend in the risky asset yield – the average of the dividend- and rent-price ratios and a common proxy for the expected return. The risky yield has fallen from 6.5% in the 1870s to 3.3% in 2015. This decline holds across countries, assets, and alternative yield measures such as earnings yields, and is somewhat stronger for housing than for equity. It means that valuations of risky assets relative to fundamentals have doubled over the long run, and are now at an all-time historical high. But risky asset yields are only an imperfect proxy for expected returns. Low asset yields could mean that the discount rate – i.e. the expected return – is unusually low, but also that future cashflow growth is unusally high (Campbell and Shiller, 1988).

Our expected return estimate is based on the dynamic Gordon growth model, and is equal to the sum of the expected yield and long-run cashflow growth for the respective asset class (Gordon, 1962; Blanchard, 1993). To estimate these expectations, we follow the standard practice in the literature and forecast future yields and cashflows using today's asset price data using a flexible VAR specification building on Golez and Koudijs (2017). We find that a little over half of the asset yield variation corresponds to predictable changes in future returns, with the rest accounted for by predictable changes in future cashflows. Consequently, our analysis assigns more than half of the long-run asset yield decline to a lower expected return. The results are similar under several alternative forecasting methods: forecasting long-run cashflow growth directly rather than through a VAR as in Blanchard (1993), using a GDP growth forecast in place of cashflows as in Farhi and Gourio (2018), and assuming constant cashflow growth as in Fama and French (2002).

Our baseline estimate of the expected risky return shows a steady and gradual long-run decline, from about 8% in the 1870s to 6% in 2015. This decline is evident in almost every country in our sample, including the US, and holds up under various alternative methods for estimating expected cashflow growth. The decline is somewhat stronger for housing than for equity owing to the larger rental yield decline and weaker rental growth predictability. Our baseline estimate of the decline is, however, conservative, since it assumes above-average long-run growth in profits and rents despite a slowing growth in GDP and productivity (Fernald, 2015). Bringing cashflow growth expectations in line with GDP growth reduces the current expected return estimate to 5%, with long-run declines of up to 5 percentage points observed in individual countries.

Regardless of the method we use, expected returns remain substantially above estimates of the trend real safe rate which are currently close to zero. At the same time, expected returns were already low in the 1980s, a period when safe rates were high. This suggests that the ex ante risk premium displays substantial secular movements. To study these movements, we first estimate the trend real safe rate using the Bayesian time series model of Del Negro, Giannone, Giannoni, and Tambalotti (2019), extending their estimates of the trend long-term government bond yield to the 17 countries in our sample. We then calculate the risk premium as the difference between our expected return measure and the trend real interest rate on long-term government bonds.

We find that the steady decline in the expected return masks sharply different trends in risk premia and safe rates. Between 1870 and 1990, risk premia more than halved from 6% to 2.5%, but the safe rate actually increased by 1.5 percentage points, offsetting a large part of the decline in the risk premium. After 1990, safe rates fell sharply towards zero, but an almost equally sharp increase in the risk premium ensured that the expected return decline was, again, modest at less than 1 percentage point. As before, these findings hold across countries and under alternative methods for calculating risky and safe rates. We use long-run trends in corporate bond spreads as an alternative proxy for the economy wide risk premium to make sure that our results are not dependent on the estimation of future cashflow growth. The corporate bond premium also follows a U-shaped long-run trend.

What are the drivers of observed trends in expected risky and safe returns? The large secular variation in the risk premium, and the differential trends displayed by risky and safe returns suggest that changes in risk – which should push safe rates and risk premia in opposite directions – are key. Consistent with the importance of the risk channel, we show that risky and safe rates are disconnected and safe rates and risk premia are strongly negatively correlated, to the extent that a 1 percentage point increase in the safe rate implies a 0.8–1 percentage point fall in the risk premium and no corresponding movement in the risky return. These relationships hold over time, across asset classes, in changes as well as in levels, and across different regression specifications and definitions of expected return. Realised risky and safe returns are also only weakly correlated, and ex post risk premia and safe rates are negatively correlated. But the extent of the risky-safe rate disconnect is both stronger and more stable over time for expected than for realised returns.[1]

Asset pricing theory predicts that reductions in macroeconomic risk should increase risk tolerance and reduce the risk premium, while also reducing the desire for precautionary savings and hence increasing the safe rate (Cochrane, 2009). We show that indeed, consumption volatility – proxied as the annual standard deviation of real consumption growth – more than halved between 1870 and 1990, helping explain the observed decline in the risk premium and increase in the safe rate. In line with theory, the magnitude of the decline in consumption volatility broadly matches that of the risk premium decline.

After 1990 however, risk premia increased while consumption volatility remained low. We argue that one reason for this is that despite lower volatility, large negative falls in consumption are much more likely – i.e. macroeconomic *tail risk* has increased. Consistent with this view, we show that the recent decades saw an increase in the probability of systemic banking crises alongside a lower skew and higher kurtosis of realised GDP growth. Other factors which may have increased the price of risk include changes in cross-sectional risk tolerance through population ageing – with older households favouring savings in safer assets (Kopecky and Taylor, 2020) – and increasing safe asset scarcity (Caballero and Farhi, 2014). In addition, we document an increased comovement of risky assets and the business cycle after 1980, while safe assets have started to co-move less. These changes in the quantity of risk offer an additional channel for the recent rise in the risk premium.

If risk is an important driver of long-run risky and safe rate trends, we would expect to see some divergence between expected risky and safe returns on one hand, and the rate of economic growth on the other. As a consequence, the ex ante risky and safe r – g gaps may be time varying. Indeed, in the early part of our sample, safe rates and growth were increasing while expected returns were falling, leading to a fall in the ex ante risky $r-g$ gap alongside a rise in the gap between safe returns and growth. Over recent decades, safe rates have declined sharply, while both risky

1. Existing literature has offered several competing explanations for the low and time-varying co-movement between realised stock and bond returns in the US (Baele, Bekaert, and Inghelbrecht, 2010; Campbell, Pflueger, and Viceira, 2020). Our findings suggest that while the overall lack of co-movement is largely driven by differences in expected returns and ex ante risk appetite, its time-varying nature is most likely driven by unanticipated shocks and information surprises.

rates and growth have only declined a little. This led to a sharp fall in the safe $r-g$ gap, which now stands at a negative 1%, and relatively little change in the risky $r-g$ gap which remains close to its historical average of 4%.

These secular movements carry important implications for the dynamics of capital accumulation, public debt and wealth inequality. In line with Blanchard (2019), low safe $r-g$ gaps – both now and historically – suggest that the cost of financing public debt is generally low, but the high risky $r-g$ gap points to a high opportunity cost of public borrowing in the form of crowded out private investment. A positive gap between expected risky returns and growth also implies that our selection of advanced economies remain dynamically efficient (Barro, 2020), with little evidence of diminishing returns to capital despite the sharp increases in wealth-to-income ratios observed over recent decades (Piketty and Zucman, 2014). The fact that returns on risky wealth are likely to remain substantially above income growth in the foreseeable future also means that equilibrium levels of wealth inequality are likely to remain high (Piketty, 2014).

Our findings relate to three strands of existing literature. The first strand studies trends in risky and safe returns. The consensus is that safe rates have declined in recent decades (Holston, Laubach, and Williams, 2017) and over the longer run (Del Negro, Giannone, Giannoni, and Tambalotti, 2019; Schmelzing, 2020), and that there is evidence of a declining equity premium in the US (Blanchard, 1993; Jagannathan, McGrattan, and Scherbina, 2000; Fama and French, 2002). We show that the expected return decline goes further back in time and is more widespread, that risky and safe rates follow markedly different long-run trends, and that a recent increase in the risk premium has kept expected returns high despite low safe rates across advanced economies. Our estimation of the risky and safe returns also contributes to the extensive return predictability literature (Cochrane, 2008). We confirm the findings in Kuvshinov (2020) that both equity and housing returns and cashflows are predictable, and show that the strength of these predictability relationships varies substantially over time, in line with evidence in Chen (2009) and Golez and Koudijs (2018).

The third strand focuses on the relationship between risky and safe returns and their underlying drivers. There is evidence that ex ante risky and safe returns have diverged recently (Gomme, Ravikumar, and Rupert, 2015; Caballero, Farhi, and Gourinchas, 2017a; Farhi and Gourio, 2018), that co-movement between realised risky and safe returns is low and time-varying (Shiller and Beltratti, 1992; Baele, Bekaert, and Inghelbrecht, 2010; David and Veronesi, 2013; Campbell, Sunderam, and Viceira, 2017; Song, 2017), that variation in risk perceptions is an important driver of short-run movements in risky and safe returns and the business cycle (Caballero and Simsek, 2020; Pflueger, Siriwardane, and Sunderam, 2020), and that gaps between realised risky returns and growth are large and time varying (Jordà, Knoll, Kuvshinov, Schularick, and Taylor, 2019). Lettau, Ludvigson, and Wachter (2008) and Bianchi, Lettau, and Ludvigson (2016) link a structural decline in the US equity premium during the 1990s to changes in macroeconomic risk in the form of, respectively, lower consumption volatility and more stable conduct of monetary policy. We show that the risky-safe rate disconnect goes far beyond the recent data, that the co-movement across expected risky and safe returns is both lower and more stable than that across realised returns, and that risk is a key driver not just of short run movements, but also long-run trends in risky and safe rates.

2.2 Measuring expected returns

Expected return is the amount of compensation investors demand for holding risky assets. Today's price of a risky asset i, P_i, should equal the present value of expected future cashflows CF_i discounted at this expected rate of return $\mathbb{E}(R_i)$:

$$P_{i,t} = \mathbb{E}\left[\sum_{\tau=1}^{\infty} \frac{CF_{i,t+\tau}}{(1+R_{i,t+1})^{\tau}}\right], \tag{2.1}$$

A low discount rate means that asset prices are high, and expected future returns are low. Discount rates vary because investor willingness to save and bear risk varies. The expected return is, therefore, equal to the sum of the safe rate R^{safe} – which depends on the investor willingness to save – and the ex ante risk premium RP – which depends on the investor willingness to hold risky as opposed to safe assets:

$$\mathbb{E}\left(R_{i,t+1}\right) = \mathbb{E}\left(R^{safe}_{t+1}\right) + \mathbb{E}\left(RP_{i,t+1}\right) \tag{2.2}$$

Expected returns differ considerably from realised returns. Expected return is an ex ante measure of what a potential investor would demand to entice her to hold risky assets. Realised return is the income of an investor who already holds the asset and reflects any unanticipated shocks to asset valuations as well as the ex ante expected return. These unanticipated shocks can arise from new information I or other unanticipated changes in the supply and demand for funds giving rise to an unexpected return ϵ (Elton, 1999):

$$R_{i,t+1} = \mathbb{E}_t\left(R_{i,t+1}\right) + I_{i,t+1} + \epsilon_{i,t+1} \tag{2.3}$$

On the margin, it is expected rather than realised returns which reflect investor willingness to finance risky projects and hence drive the financing decisions of households and firms. But since realised returns R contain information on expected returns, one could in principle us it as a proxy for $\mathbb{E}(R)$. Empirically, however, realised returns offer a rather poor proxy for expected returns because their variation is driven by the unexpected components I and ϵ, rather than the expected return (Elton, 1999). Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) show that variation in realised equity and housing returns – even over periods stretching to multiple decades – is dominated by unanticipated large shocks.

We may expect the variation in I and ϵ to average out over very long time horizons, meaning that realised returns can provide a useful proxy of the *average level* of the expected return. But because the variation in realised returns – even across decades – is both sizeable, with standard deviations of 10–20 ppts per year, and largely driven by unanticipated shocks, realised returns are generally unsuitable for mapping out the *trend* in the expected return. The noise in I and ϵ aside, trends in realised returns yield a fundamentally biased estimate of the expected return trend. Whenever expected returns and discount rates *decline*, the present value of cashflows – and hence asset prices and realised returns – *increase*, driving expected and realised returns in opposite directions. This bias is far from hypothetical: as Fama and French (2002) show, a sharp decline in the ex ante US equity premium after 1950 boosted realised returns such that the ex post equity premium remained broadly flat. The above-mentioned drawbacks also apply to estimates of the rate of return on capital in the national accounts, which are inherently backward-looking and include any unanticipated shocks to capital income and wealth.

In this paper, instead of eliciting proxies of expected returns using the noisy realised return data, we seek to measure the expected return directly. We do this for the two major classes of risky assets – equity and housing – across 17 countries over the time period 1870–2015. To construct a direct expected return estimate for each of these two assets, we follow the literature and use a linearised version of the present value relationship in equation (2.1) – the dynamic Gordon growth model – derived, for example, by Blanchard (1993):

$$\mathbb{E}\left(R_{i,t+1}\right) \approx \mathbb{E}\left(CF_{i,t+1}/P_{i,t}\right) + \mathbb{E}\left(\tilde{g}_{i,t+2}\right) \tag{2.4}$$

Above, expected returns on asset i are the sum of the expected *asset yield* $CF_{i,t+1}/P_{i,t}$ – the dividend- or the rent-price ratio – and expected long-run *cashflow growth* in dividends or rents $\mathbb{E}\left(\tilde{g}_{t+2}\right)$.[2]

Our main task in this paper is, therefore, to calculate the two components of the direct expected return estimate: the expected asset yield $\mathbb{E}(CF_{i,t+1}/P_{i,t})$ and expected cashflow growth $\mathbb{E}\left(\tilde{g}\right)$. Since we do not have data on actual investor expectations for our extensive sample period, we instead use empirical forecasts of the yield and cashflow growth derived by exploiting the correlations between current and future yields, cashflows and returns. Our baseline approach uses theory and the present value identity in (2.1) to guide both our choice of predictors and the forecasting technique. For this, we consider the log-linearised version of the present value identity derived by Campbell and Shiller (1988):

2. $\tilde{g}_{i,t+2}$ is the annuity value of future cashflow growth, which is a weighted average of expected future cashflow growth rates calculated as $\tilde{g}_{i,t+2} = w_{i,1}\mathbb{E}g_{i,t+2} + w_{i,2}\mathbb{E}g_{i,t+3} + \ldots + w_{i,\tau}\mathbb{E}g_{i,t+\tau+1}$. Here $g_{i,t} = CF_{i,t}/CF_{i,t-1} - 1$ is the year-on-year cashflow growth, and the weights are $w_{i,t} = (1+g_i)^{\tau-1}(r_i - g_i)/(1+r_i)^{\tau}$, where g_i and r_i are the average cashflow growth and return rates for asset i.

$$dp_{i,t} \approx \mathbb{E}\sum_{s=0}^{\infty}\rho_i^s r_{i,t+1+s} - \mathbb{E}\sum_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s} \quad (2.5)$$

Here $\mathbb{E}\sum\limits_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s}$ is the present value of expected future log dividend growth rates, with $exp\left[(1-\rho_i)\mathbb{E}\sum\limits_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s}\right] - 1 \approx \tilde{g}_{i,t+2}$ in the level linearisation in (2.4). $\mathbb{E}\sum\limits_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s}$ is the present value of expected future log returns, and $\rho_i = \frac{P_i/CF_i}{1+P_i/CF_i}$ is a linearisation constant. Equation (2.5) is similar to the level linearisation in (2.4): both tell us that asset yields will be high, and asset prices will be low whenever the expected return is high or expected future cashflows are low. But equation (2.5) has the relative advantage that we can estimate its different components directly from the data in a way that respects their co-dependence induced by the present value identity.

For our estimation, we follow the standard procedure in the literature (see, for example Golez and Koudijs, 2017), and run a VAR in three variables $[r_{i,t}, dg_{i,t}, dp_{i,t}] \equiv z_{i,t}$ for each of the two asset classes i, equity and housing. The VAR is estimated using 6-equation GMM accounting for time and cross-sectional dependence in standard errors and respecting the present value moment constraints. The 6 equations capture the 9 moment conditions (from 3 X 3 variables) subject to 3 restrictions imposed by the present value identity. Existing studies show that the strength of such predictability relationships can change materially over time (Chen, 2009). To account for such time variation, we estimate the VAR using rolling 40-year windows, the same time window as in Blanchard (1993).

$$\begin{aligned}
\text{VAR:} \quad & z_{i,t} = A_{i,T}z_{i,t-1} + u_{i,t}, \qquad z_{i,t} \equiv [r_{i,t}, dg_{i,t}, dp_{i,t}] \\
\text{Moment conditions:} \quad & E[(z_{i,t+1} - A_{i,T}z_{i,t}) \otimes z_{i,t}] = 0 \\
\text{Restrictions:} \quad & (e1' - e2' + \rho_i e3')A_{i,T} = e3'
\end{aligned}$$

Here, $e1$ and $e2$ are the first two columns of the identity matrix I, and T is the 40-year rolling time period under consideration.

The long-run forecasts for discount rate news $\mathbb{E}\sum\limits_{s=0}^{\infty}\rho_i^s r_{i,t+1+s}$, cashflow news $\mathbb{E}\sum\limits_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s}$, and the year-ahead forecast for $dp_{i,t+1}$ can then be estimated as follows:

$$\begin{aligned}
\text{Discount rate news}: \quad & \hat{\mathbb{E}}\sum_{s=0}^{\infty}\rho_i^s r_{i,t+1+s} = e1'\hat{A}_{i,T}(I-\rho_i\hat{A}_{i,T})^{-1}z_{i,t} \\
\text{Cashflow news}: \quad & -\hat{\mathbb{E}}\sum_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s} = e2'\hat{A}_{i,T}(I-\rho_i\hat{A}_{i,T})^{-1}z_{i,t} \\
\text{Expected yield}: \quad & \hat{\mathbb{E}}\, dp_{i,t+1} = \hat{a}_{3,i,T}z_{i,t},
\end{aligned}$$

where $\hat{a}_{3,i,T}$ is the third row of the estimated $\hat{A}_{i,T}$ coefficient matrix. We use the VAR estimated over the window $[t-40, t]$ to calculate the forecasts for $t+1$, and use the VAR estimated for the window 1870–1910 to produce forecasts for the first 40 years of our sample.

The Campbell-Shiller decomposition gives us an estimate of the expected log asset yield and the present value of the expected log future cashflow growth, both demeaned at the country level. To convert these into the estimate of the expected return level in equation (2.4), we add back the country-specific means for the *dp* and *dg* variables and convert the present value cashflow growth estimate to annual year-on-year equivalent by multiplying it by $1-\rho_i$:

$$\hat{\mathbb{E}}\left(R_{i,t+1}\right) = \underbrace{exp\left[\hat{\mathbb{E}}dp_{i,t+1} + \overline{dp}_{i,j}\right]}_{\mathbb{E}(CF/P)} + \underbrace{exp\left[(1-\rho_i)\hat{\mathbb{E}}\sum_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s}\right] + \overline{DG_{i,j}} - 1}_{\mathbb{E}(\tilde{g})} \tag{2.6}$$

Above, indices i and j refer to asset and country respectively, and $DG_{i,t+1} = CF_{i,t+1}/CF_{i,t} - 1$ refers to absolute rather than log cashflow growth. This gives us an estimate of the level of the expected return at $t+1$. Adding back the mean of log cashflow growth instead would give us an estimate of expected log return, which displays the same trend as the absolute return but a lower sample mean. Note also that replacing the forecast $\hat{\mathbb{E}}dp_{i,t+1}$ with observed $dp_{i,t+1}$ gives us the discount rate news component in the Campbell-Shiller decomposition in equation (2.5), which – since yields are very persistent and hence differences between $\hat{\mathbb{E}}dp_{i,t+1}$ and $dp_{i,t+1}$ are small – displays a very similar trend to our baseline expected return estimate.

The intuition behind our estimation is the following. We can observe the trend in the risky asset yield *dp*, but we do not know if it is driven by trending expected returns or cashflows. The predictive regressions allow us to decompose year-on-year variation in *dp* into future cashflow and discount rate movements. From equation (2.5), the variance of dp_i is the sum of variances of discount rate and cashflow news. The VAR allows us to estimate these variance shares.[3] We then apply these variance

3. This variance decomposition can be directly estimated from the VAR as follows:

$$\text{Var}(dp_{i,t}) = e3'\Gamma e3 = \underbrace{e1'\hat{A}_{i,T}(I-\rho_i\hat{A}_{i,T})^{-1}\Gamma e3}_{\text{Discount rates}} - \underbrace{e2'\hat{A}_{i,T}(I-\rho_i\hat{A}_{i,T})^{-1}\Gamma e3}_{\text{Cashflows}}, \quad \Gamma = \mathbb{E}(zz')$$

shares to the overall trend in the yield to determine the relative contribution of trend changes in expected returns and cashflows. If these shares are, say, 50-50, yields are equally good at predicting future cashflows and returns and hence a 1 percentage point trend decline in the yield will be interpreted as a 0.5 ppts trend decline in the expected return and a 0.5 ppts trend increase in the expected cashflow growth rate. If the shares are 100-0 and yields only forecast returns and not cashflows, a 1 percentage point trend decline in the yield will mean a 1 percentage point trend decline in the expected return.

We estimate the VAR separately for the two asset classes in our study, equity and housing, with the expected risky return equal to the average of the two:

$$R_{t+1}^{risky} = \left[\hat{\mathbb{E}}\left(R_{t+1}^{\text{equity}}\right) + \hat{\mathbb{E}}\left(R_{t+1}^{\text{housing}}\right)\right]/2 \tag{2.7}$$

As well as the baseline Campbell-Shiller forecast, we also consider several alternative expected cashflow growth estimates. First, we draw on Blanchard (1993) and directly forecast the annuity value of future cashflow growth $\tilde{g}_{t+2}$ in equation (2.4) using the time t risky asset yield. Second, we set expected cashflow growth to a constant country-specific mean, similarly to Fama and French (2002). Third, similarly to Farhi and Gourio (2018), we use the long-run expected GDP growth as a proxy for cashflow growth in the Gordon model. To do this, we forecast the annuity value of real GDP growth using two lags of real GDP growth, the term premium and the dividend-price ratio.[4]

The final step is to decompose expected returns into the ex ante risk premium and safe rate according to equation (2.2). To do this, we take a best-practice off-the-shelf estimate of the trend long-term real safe rate from Del Negro, Giannone, Giannoni, and Tambalotti (2019). Their method is based on a Bayesian VAR model and allows us to extract slow-moving real safe rate trends from cross-country long-run data on short-term rates, long-term rates and inflation. Del Negro, Giannone, Giannoni, and Tambalotti (2019) compute safe rate estimates for 7 advanced economies, and we use their method to extend their estimates to our sample of 17 countries. As with expected returns, we also check our results against alternative safe rate estimates: a short-term rather than long-term real safe rate, and the natural rate estimates of Holston, Laubach, and Williams (2017).

We calculate the ex ante risk premium as the difference between the expected risky return and the trend real safe rate. As a further robustness check, we compare our ex ante risk premium estimates for the housing and equity market to a direct forward-looking risk premium measure from the corporate bond market –

4. Similar to $\tilde{g}_{i,t+2}$ in equation (2.4) (see footnote 2), we estimate the annuity value of real GDP growth from the year ahead onward, discounted at the country average real risky return rate r. To calculate the annuity value, we compute expected growth after 2015 using the OECD Economic Outlook forecast for GDP in 2060.

the yield-to-maturity spread between corporate and government bonds. The credit spread measure does not depend on assumptions about cashflow growth, inflation expectations and the specific modelling techniques used to estimate the expected risky return and the trend real safe rate. However, it is based on a relatively narrow market segment, and may not fully capture risk premium movements in the broader macroeconomy.[5]

Our estimates of expected risky returns and risk premia cover 17 countries – Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, UK and USA – on an annual basis between 1870 and 2015. The raw data on realised returns, risky asset yields and cashflows come from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019), extended to also include Canadian equities. The equity series are generally market cap weighted averages of the dividend-price ratio and dividend growth for all listed shares. In some of the earlier sample periods, value-weighted blue-chip indices covering a smaller number of shares are used. The dividend-price ratio is measured as dividends paid over the course of the year divided by the year-end share price.

The housing 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. These data are complemented by estimates of the real safe rate based on short and long term government bond yields, and corporate bond spread data from Kuvshinov (2020). Returns and growth rates are deflated using the consumer price data from the latest vintage of the Jordà-Schularick-Taylor macro-history database (Jordà, Schularick, and Taylor, 2017). The dividend growth series is affected by several outliers mostly relating to near-zero dividend payments during war time and their subsequent resumption, which leads to very high growth rates that could bias the average expected return calculation in (2.4). To deal with these, use the same procedure as Kuvshinov (2020) and winsorize dividend and rental growth at 1% level, adjusting the yield and total return series accordingly.

Table 2.1 shows the basic summary statistics of expected returns, risk premia and safe rates, and Figure 2.1 plots the corresponding expected return levels and their correlation with realised returns. Expected returns on both housing and equity are high – around 6.5% p.a. – and considerably above those earned on safe investments. The average levels of expected returns are similar to those of realised returns, but the volatility is an order of magnitude lower. This shows that – as suggested by Elton (1999) – variation in year-on-year realised returns is primarily driven by unanticipated shocks to asset valuations rather than the expected return, making realised

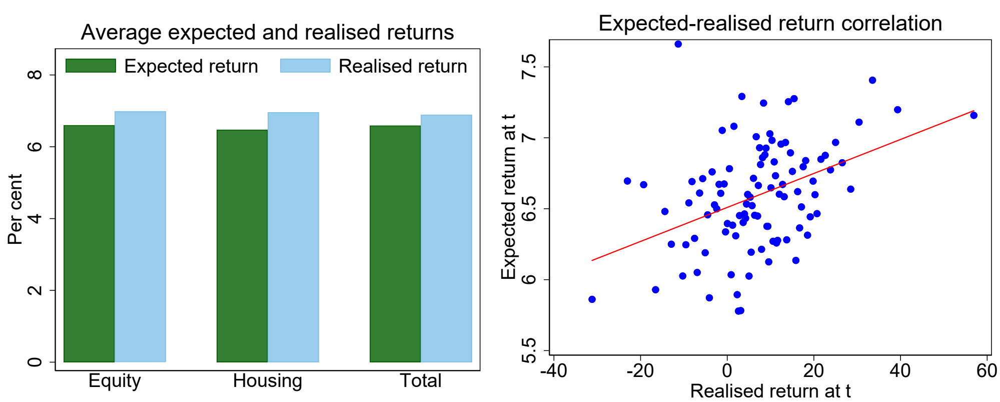
5. Changes in corporate bond spreads can also reflect time-varying credit quality or default probabilities. However, Giesecke, Longstaff, Schaefer, and Strebulaev (2011) show that in long-run US data, variation in default rates is not an important driver of changes in credit spreads.

Table 2.1. Expected risky return and its components

	(1) Equity	(2) Housing	(3) Total
Expected return	6.62 (2.05)	6.47 (1.83)	6.58 (1.60)
Components of the expected return:			
Ex ante risk premium	4.06 (2.11)	3.97 (1.94)	4.03 (1.70)
Ex ante safe rate	2.55 (1.13)	2.51 (1.14)	2.55 (1.13)
Realised return	7.12 (21.28)	7.14 (9.66)	7.12 (12.72)
Observations	1759	1645	1759

Notes: Unweighted arithmetic averages of annual country-specific data, 17 countries, 1870–2015. Annual standard deviation in parentheses. The expected return is the sum of the expected yield and expected cash-flow growth obtained using a predictability VAR. The risk premium is the difference between the expected return and ex ante safe rate. Ex ante real safe rate is estimated using a Bayesian VAR with slow-moving trends as in Del Negro, Giannone, Giannoni, and Tambalotti (2019). For Canada, we use equity data to a proxy for total risky returns and risk premia.

Figure 2.1. Levels and correlations between expected and realised returns



Notes: Pooled data for 17 countries. Left-hand panel: average levels of expected and realised risky returns, computed as the average of equity and housing. Right-hand panel: binned scatter plots of the correlation between realised and expected risky return in the specific country and year.

returns a realively poor expected return proxy. The right-hand panel of Figure 2.1 shows a binned scatter plot of expected versus realised returns for all the country-year observations in the sample split into 100 bins for each measure. Consistent with equation (2.3), expected and realised returns are positively correlated. But as discussed above, the correlation is far from perfect, with realised return bins of between -40% and +60% p.a. corresponding to expected return bins of between 5.5% and 7.5% p.a.

2.3 Long-run trends in expected returns and risk premia

2.3.1 Trends in risky asset yields

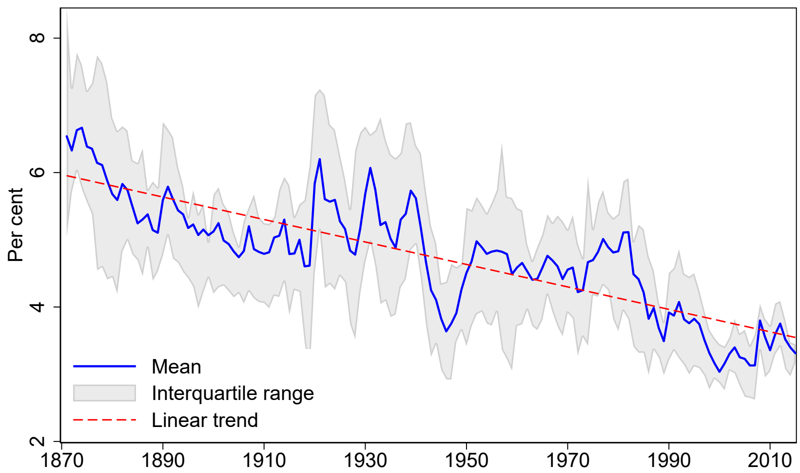
Figure 2.2 shows the evolution of the risky yield – the average of the the dividend-price and rent-price ratios – over time. The solid blue line is the median yield across the 17 countries in our sample, the shaded area indicates the interquartile range, and the dashed red line – the implied long-run linear trend. Over the last 145 years, risky asset yields have undergone a steady decline. Altogether, the average yield fell from 6.5% in the 1870s to 3.3% in 2015, meaning that risky asset valuations relative to fundamentals have doubled over the long run.

Figure 2.3 shows that this decline is evident across both asset classes and across different countries. The left-hand panel of Figure 2.3 shows that both equity and housing yields have fallen substantially. Much of the early decline in the aggregate yields is driven by housing, whereas the sharp drop during the 1980s is mostly attributable to rising equity valuations. The right-hand panel of Figure 2.3 plots the total change in the risky yield (average of equity and housing) over the full sample in individual countries. In all but two countries this change is negative, with some countries documenting drops of around 4 percentage points. No country registers a large yield increase.

Appendix Figures 2.A.1–2.A.3 show that the downward trend in yields is robust to different methods of calculating the yield and alternative groupings of countries. Appendix Figure 2.A.1 compares our benchmark housing yield estimates to those that can be obtained from national accounts data as the ratio of rents paid minus non-tax, non-utility housing costs to housing wealth. The national accounts estimates are slightly lower throughout the sample but display a very similar long-run trend to our baseline series.

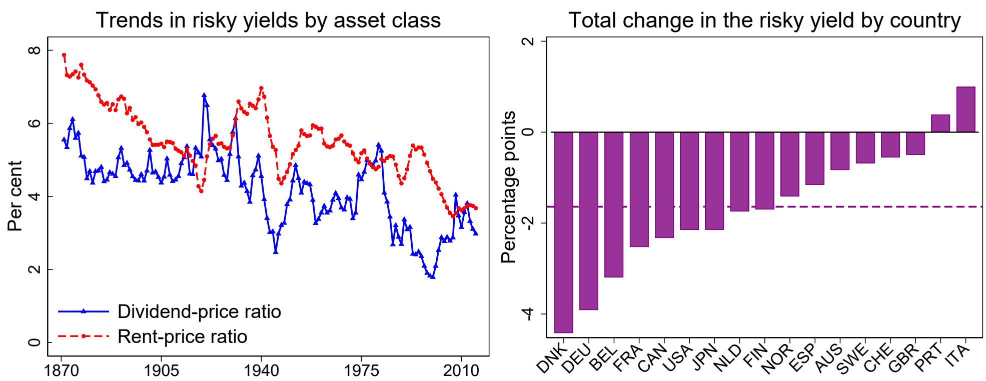
The main concern with our dividend yield estimates is that they underestimate total cashflows to shareholders – especially in recent data – because they do not account for stock buybacks, which have become an increasingly important form of shareholder compensation in the US (Grullon and Michaely, 2002). The left-hand panel of Appendix Figure 2.A.2 shows that the total earnings yield in the US – which is unaffected by buybacks – displays a similar long-run decline to the dividend yield. Kuvshinov and Zimmermann (2020) further show that in post-1990 Compustat data

Figure 2.2. The risky asset yield



Notes: Data for 17 countries. The yield is the average of the dividend-price and rent-price ratios. The solid line and the shaded area are, respectively, the mean and interquartile range of the individual country data in each year. The dashed line represents the linear trend.

Figure 2.3. Changes in yields by asset and country



Notes: The left-hand panel shows unweighted averages of 17 countries. The right-hand panel shows the difference between the average yield in the last decade of the sample and the first decade of the sample, with countries ordered from the most negative to the most positive change and dashed line showing the average change in the yield across countries.

Table 2.2. Return and cashflow predictability through time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Equity				Housing			
	1910	1940	1980	2010	1910	1940	1980	2010
Predictive coefficient on dp_t:								
r_{t+1}	0.06***	0.05**	0.15***	0.05***	0.13***	0.07***	0.07***	0.05***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
g_{t+1}	-0.09***	-0.14***	-0.03	-0.10***	0.00	-0.03***	-0.05***	-0.02***
	(0.02)	(-0.02)	(0.04)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Variance decomposition of dp_t:								
DR news	51	29	82	35	97	48	41	78
CF news	49	71	18	65	3	52	59	22
Observations	545	648	667	693	304	507	577	649

Note: Rolling window VAR estimates for years $t - 40$ to t. VAR estimated using GMM subject to present value moment constraints, 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, demeaned at country level. 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. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

covering our cross-section of countries, earnings- and dividend-price ratios follow similar trends. Finally, the left-hand panel of Appendix Figure 2.A.3 shows that the downward trend in the yield is not a result of time-varying sample composition, and also holds across different time-invariant samples where we, for example, limit the sample to only include those countries where we have data going back to the 1870s.

2.3.2 Trends in expected returns

The decline in the risky asset yield does not necessarily mean that expected returns have declined. It could, instead, be driven by higher expected cashflow growth. To assess whether this is the case, we follow Golez and Koudijs (2017) and run a predictability VAR in three variables: log real total returns, log real cashflow growth and log of the asset yield. If movements in yields are driven by changes in expected cashflow growth, yields should predict future cashflows. If changes in expected returns are important, yields should predict future returns. We run a separate VAR for the two asset classes, equity and housing. Because the strength of the predictability relationships can change over time, we estimate the VAR over 40-year rolling windows. After establishing the relative importance of variation in expected returns and cashflows, we use the VAR to construct a long-run forecast of real cashflow growth which we use as a proxy for expected cashflows $\mathbb{E}\left(\tilde{g}_{i,t+2}\right)$ in our expected return calculation in equation (2.4). Section 2.2 spells out the method in more detail.

Table 2.2 shows the most important VAR coefficients (top panel) and the relative strength of cashflow and return predictability (bottom panel) for selected benchmark periods. Appendix Table 2.A.2 shows all the estimated VAR coefficients for these benchmark periods. The top row of Table 2.2 displays the predictive coefficient of year-ahead realised total return on today's dividend-price ratio (columns 1–4) or rent-price ratio (columns 5–8). The second row shows predictive coefficients on the log real dividend and rental growth. The VAR for 1910 uses the data for 1870–1910, for 1940 – the data for 1900–1940, and so on. For both equity and housing, yields generally forecast both future returns and cashflows. The magnitudes are statistically significant and economically large. For equity, a 1 percentage point increase in the dividend-price ratio generally predicts 1–3 percentage points lower returns one year ahead, and 1–3 percentage points higher real dividend growth.[6] For housing, a 1 percentage point lower rent-price ratio forecasts 1.5–3 percentage point lower returns and 0–1.5 percentage points higher real rental growth.[7]

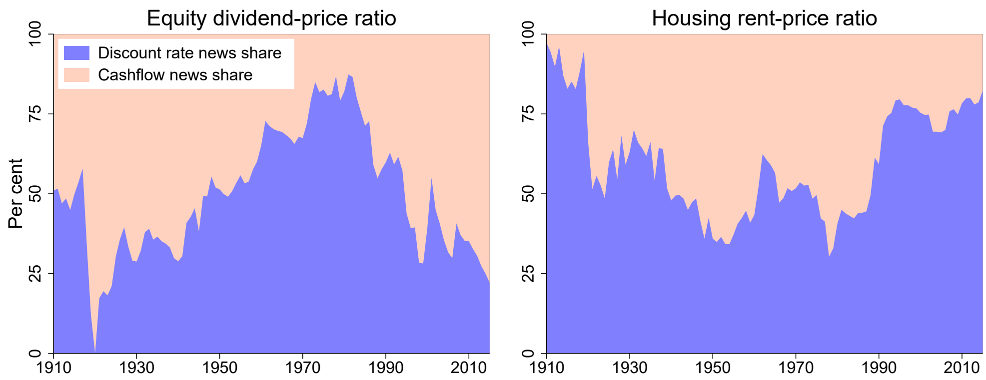
Because both cashflows and returns are predictable, the long-run decline in yields is likely to be attributable to a mixture of higher expected cashflows and lower expected returns. The key question is – how much of each? To determine this, the bottom panel of Table 2.2 and Figure 2.4 compare the relative strength of return and cashflow growth predictability by decomposing the variation in the dividend-price ratio into discount rate news (expected returns) and cashflow news. For example, Table 2.2 column 1 shows that during 1870–1910, around 50% of the variation in the dividend-price ratio was attributable to changes in future dividends, and the other 50% – to changes in future equity returns. This means that if the dividend-price is 2 percentage points above its long-run mean during this period, around half of that will be attributed to below-mean expected cashflows and the other half to above-mean expected returns. Correspondingly, our expected return estimate will equal the expected yield (in practice, very close to the actual yield) plus average cashflow growth in the sample minus about half of the distance between the yield and its long-run mean (equation (2.6)).

Figure 2.4 shows that the relative importance of cashflow and discount rate news varies substantially over time and across asset classes. The discount rate news share was high for equity during much of the 20th century, in particular during the 1970s and 1980s, a period when the dividend-price ratio registered sharp falls (Figure 2.3). For the rent-price ratio, the discount rate share is always high but is somewhat

6. A 1 percentage point increase in the dividend-price ratio is roughly a 25% relative increase, meaning that year-ahead returns are expected to fall by 1.07 (mean return) * 0.06 (regression coefficient) * 0.25 $\approx$ 1.6 percentage points, for case of the the 1910 VAR.

7. Because both dividend- and rent-price ratios are very persistent, (see the predictive coefficients of $dp_t + 1$ on dp_t in the Appendix Table 2.A.2) these return and cashflow growth increases tend to cumulate over time, such that together long-run cashflow and discount rate innovations explain the variance of the dividend- and rent-price ratios, as also shown in the decomposition in the bottom panel of Table 2.2.

Figure 2.4. Variance decomposition of the dividend- and rent-price ratios through time



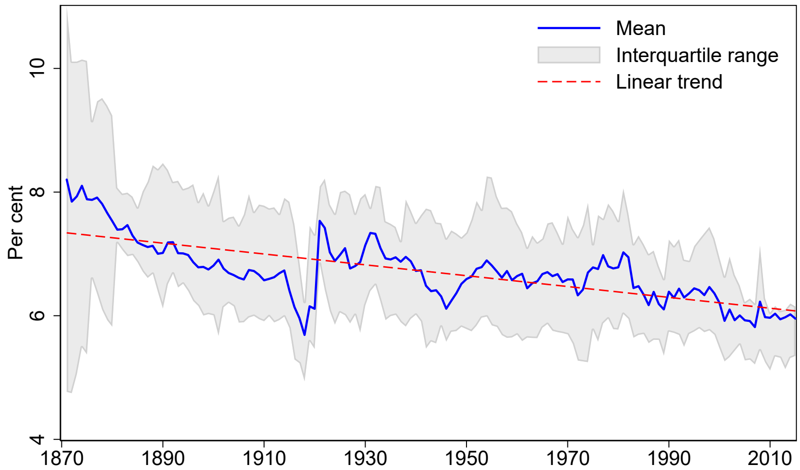
Notes: Estimates using a 40-year rolling window VAR in real returns, cashflows and valuations. Discount rate news share is the proportion of variation in $dp_{i,t}$ of asset i attributable to changes in the long-run VAR forecast of return, $\mathbb{E}_t \sum_{k=0}^{\infty} \rho_i^k r_{i,t+k+1}$. Cashflow news share is the proportion attributable to changes in the long-run VAR forecast of cashflow growth $\mathbb{E}_t \sum_{k=0}^{\infty} \rho_i^k dg_{i,t+k+1}$.

higher today than in the mid-20th century. The current discount rate news share is around 25% for equity and 75% for housing. This means that around one-quarter of the difference between the current and sample-average dividend-price ratio will be interpreted as lower expected returns and three-quarters – as higher expected cashflows, with the converse true for housing. As we subsequently show in Figure 2.7, this makes our estimates of the expected return decline somewhat conservative, especially for equities, since it assumes that the high cashflow growth of the recent decades can be sustained into the future, pushing up today's estimates of the expected return.

As a final step, we calculate the expected risky return as the sum of the yield and the VAR long-run cashflow growth forecast for the specific asset class. Appendix Figure 2.A.5 shows that our cashflow forecasts follow a similar trend to the annuity-valued growth in realised cashflows $\tilde{g}$, while also looking through some of the booms and busts in the realised growth data. Over the long run, both expected dividend and rent growth have increased, but the magnitude of these increases is smaller than that of the declines in the dividend-price and rent-price ratios. This means the sum of the yield and cashflow growth expectations – the expected return – has declined over the long run.

Figure 2.5 shows the evolution of the expected risky return – the average of the equity and housing series – over time. The solid line shows the cross-country average, the shaded area – the interquartile range, and the dashed line shows the linear trend. Individual country series can be found in the Appendix Figure 2.A.6. Over the last 145 years, expected risky returns have been in steady decline, falling from close to 8% p.a. in the 1870s to roughly 6% p.a. in 2015. The pace of this decline has been rather gradual and stable over time, with somewhat sharper drops recorded in the

Figure 2.5. The expected return on risky assets



Notes: Data for 17 countries. The expected risky return is the average of expected returns on equity and housing, each computed as the sum of the yield and long-run cashflow growth forecasts from a VAR in returns, cashflows and asset yields. The solid line and the shaded area are, respectively, the mean and interquartile range of the individual country data in each year; the dashed line is the linear trend.

late 19th century and after 1980. The dip in the expected returns around World War I is largely attributable to unusually low wartime rent levels which resulted in low yields.[8]

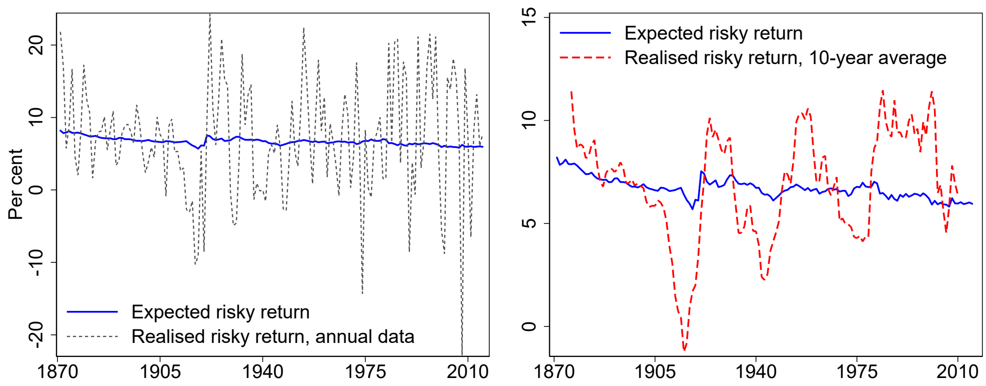
Appendix Figures 2.A.2 and 2.A.3 show that, similarly to asset yields, the downward aggregate trend in expected returns is not affected by the definition of the yield – for example, using earnings yields instead of dividend yields for the US equity market – or time-varying sample composition, with aggregate expected return estimates basically unchanged as we switch between different time-invariant country groupings.[9]

Figure 2.6 compares our expected return estimates to annual and 10-year-average realised returns. Even though there is some correspondence between the

8. The risky yield in Figure 2.2 does not show a comparative dip because the dividend-price ratio increased during World War I, offsetting the housing yield decline. But because during this time period equity return predictability was weak and housing return predictability was strong (Table 2.2), the increase in the dividend-price ratios does not translate to higher expected equity returns whereas low rent-price ratios translate to lower housing returns, resulting in an overall wartime dip in the expected risky return.

9. The main advantage of earnings yields is that they look through changes in how earnings are distributed – such as a switch from dividends to stock buybacks – and focus on underlying profitability. We follow Fama and French (2002) and calculate the earnings-based expected equity return for the US stock market as the sum of the dividend-price ratio and expected earnings growth. Earnings growth expectations are computed as the 40-year rolling window forecast of annuity-value earnings growth using today's earnings-price ratio. To further correct any buyback-related bias, we fix the changes in the US dividend-price ratio to equal those in the earnings-price after 1982, since using buybacks to compensate shareholders was relatively rare before a change in the SEC regulations in 1982 (Grullon and Michaely, 2002).

Figure 2.6. Trends in expected and realised returns



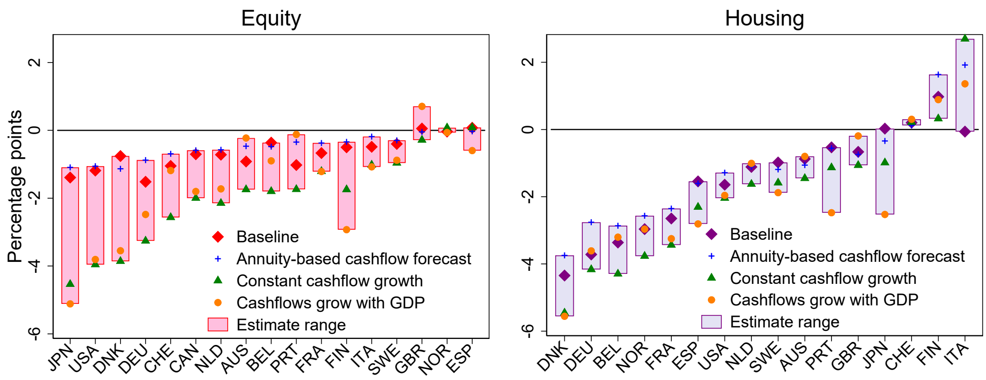
Notes: Data for 17 countries. The expected risky return is the average of expected returns on equity and housing. The realised risky return is the average of total real returns on equity and housing. 10-year average realised return is a centered rolling-window mean of realised returns between years $t - 4$ and $t + 5$.

two measures, realised returns are simply too volatile to elicit a sensible estimate of the expected return trend, even when averaged over decadal periods. Put differently, if one observes a low trend realised return level – such as during the two World Wars and the 1970s stagflation – or a high trend realised return such as the late 19th century or the 1990s period – it is much more likely to correspond to unexpected good or bad news, or other unexpected shocks to asset values rather than changes in the expected return.

Figure 2.7 investigates whether the long-run expected return decline persists through several alternative estimation methods, across countries and across asset classes. Each dot shows the difference between the average expected return in the last and first decade of the sample for selected method, asset and country, and the bars show the range of the alternative estimates. Starting with the baseline estimates (red and purple diamonds), these show a near-universal decline across both equity and housing, and in the vast majority of countries in our sample. The magnitude of the decline is larger for housing than for equity, with expected housing returns in some countries falling by roughly 4 percentage points over the long run.

Turning to the different estimation methods, the blue crosses directly forecast the annuity value of dividend or rental growth directly as in Blanchard (1993), using today's dividend or rental yield as the predictor. This direct forecast imposes less structure on the data than the VAR, but is relatively inefficient and has to assume a value for cashflow growth after 2015 in order to estimate the annuity-valued growth rate $\tilde{g}$. The green triangles in Figure 2.7 instead assume constant cashflow growth, as in the study of Fama and French (2002). This estimate allows us to abstract from estimation errors in future cashflows, but may be biased since future cashflows are positively correlated with risky yields in the data. The orange circles assume that long-run cashflows grow at the same rate as GDP as in Farhi and Gourio (2018). To

Figure 2.7. Long-run change in the expected return by country, asset class and estimation method

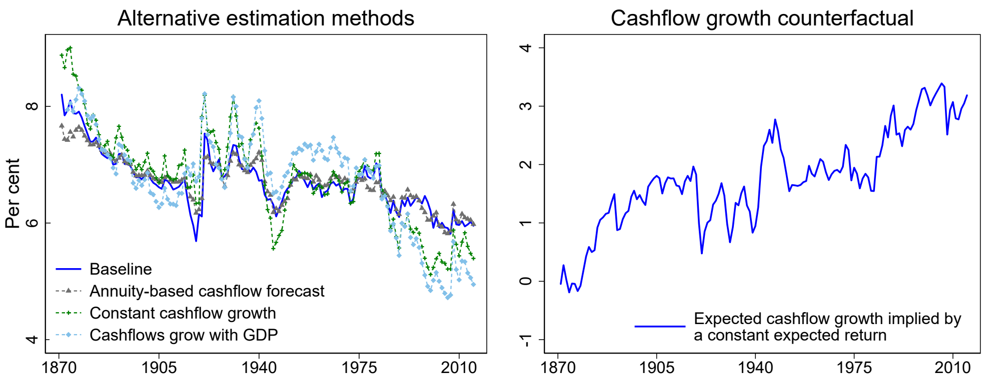


Notes: Difference between expected returns in the last and first decade of the sample for each country and asset class. The expected return is the sum of the expected yield and expected cashflow growth obtained using a predictability VAR. The "annuity-based cashflow forecast" specification uses this year's yield to directly forecast the year-ahead annuity value of cashflow growth for each asset class, as well as next period's yield, similarly to Blanchard (1993). The "constant cashflow growth" specification sets expected cashflow growth equal the sample average for each country, as in Fama and French (2002). The "cashflows grow with GDP" specification sets expected cashflow growth equal to expected long-run GDP growth as in Farhi and Gourio (2018), with the GDP growth rate, term premium and the dividend-price ratio as predictor variables.

this end, we forecast the annuity value of GDP growth – with post-2015 growth rates tied down by the *OECD Economic Outlook* GDP forecast for 2060 – using two lags of the current real GDP growth, the term premium and the dividend-price ratio, and use this value as a proxy for $\mathbb{E}(\tilde{g})$ for both housing and equity. This method has the advantage of keeping our estimates in line with long-run productivity trends, but ignores long-run changes in factor shares which can drive a wedge between growth in capital income and GDP (Karabarbounis and Neiman, 2014).

The range of alternative estimates in Figure 2.7 shows that our baseline expected return measure is, if anything, rather conservative. While the annuity-based cashflow forecast measure tends to be quite similar, assuming that cashflows are constant or grow with GDP results in substantially more pronounced declines in expected returns, particularly for equities. For some countries, asset classes and measures, the resulting decline in expected returns more than doubles. The reason for this is the following. Under our baseline estimates, every 1 percentage point of the long-run decline in the yield is partially offset by a 0.25–0.75 ppt increase in expected cashflow growth. The justification for this offset is that yields have strong predictive power for future cashflows in the VAR (Table 2.2). If we assume that cashflow growth is constant or follows GDP, this offset is no longer there. Indeed, recent decades have seen rising risky asset valuations and profits at the same time as GDP growth has declined (Greenwald, Lettau, and Ludvigson, 2019; Kuvshinov and Zimmermann, 2020). These higher cashflows are then associated with higher profit and rent shares in GDP rather than higher GDP growth (Rognlie, 2015; Barkai, 2020). Assuming these profit and rent share increases cannot continue indefinitely would

Figure 2.8. Alternative measures of expected returns



Notes: Unweighted averages of 17 countries. The "annuity-based cashflow forecast" specification uses this year's yield to directly forecast the year-ahead annuity value of cashflow growth for each asset class, as well as next period's yield, similarly to Blanchard (1993). The "constant cashflow growth" specification sets expected cashflow growth equal the sample average for each country, as in Fama and French (2002). The "cashflows grow with GDP" specification sets expected cashflow growth equal to expected long-run GDP growth as in Farhi and Gourio (2018), with the GDP growth rate, term premium and the dividend-price ratio as predictor variables. The cashflow growth counterfactual assumes expected returns are constant and equal to the sample mean, and then backs out the counterfactual expected cashflow growth as the difference between the constant expected return and the yield.

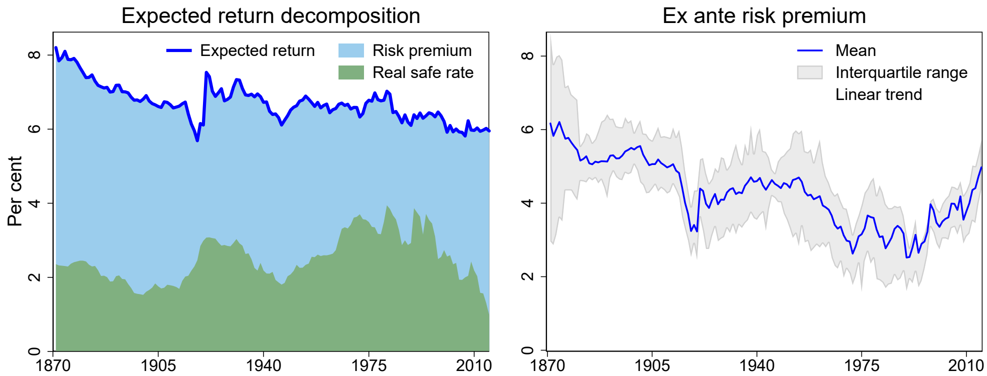
bring our long-run cashflow growth forecast closer to GDP growth, and result in larger expected return declines similar to the corresponding orange-circle estimates in Figure 2.7.

The left-hand panel of Figure 2.8 compares the time series evolution of expected returns under the four alternative estimates discussed above. The time trend is similar across all measures. The baseline and annuity-based forecast estimates are more stable over time, with movements in the dividend- and rent-price ratios partially offset by changing cashflow growth expectations. Assuming constant or GDP-driven cashflow growth results in larger long-run declines in expected returns, of up to 4 percentage points on average. As an additional check on our estimates, we ask the following question: if expected returns were in fact constant, by how much would expected long-run cashflow growth have to increase to justify this? The answer is, quite a lot. The right-hand panel of Figure 2.8 shows the counterfactual cashflow growth necessary for a constant expected return equal to our estimated sample average. This counterfactual growth displays a pronounced upward time trend, increasing from around zero to 3 percentage points per year. This counterfactual increase is difficult to rationalise in light of the relatively modest variation in trend real GDP growth and factor shares throughout our sample.

2.3.3 Trends in safe rates and risk premia

Does the decline in expected risky returns simply mirror the well-documented fall in the natural safe rate, or does it represent a separate and distinct phenomenon?

Figure 2.9. Expected returns, safe rates and risk premia



Notes: The left-hand panel shows unweighted averages of 17 countries. The expected return is the sum of the expected yield and expected cashflow growth obtained using a predictability VAR. The risk premium is the difference between the expected return and ex ante safe rate. Ex ante real safe rate is estimated using a Bayesian VAR with slow-moving trends as in Del Negro, Giannone, Giannoni, and Tambalotti (2019). In the right-hand panel, the solid line is the cross-country mean and the shaded area is the interquartile range of individual country data. The dashed line represents the linear trend.

Table 2.3. Expected returns and risk premia through time

	(1)	(2)	(3)	(4)	(5)
		Level		Absolute change	
	1880	1990	2015	1880-1990	1990-2015
Expected risky return	7.58	6.39	5.95	-1.19	-0.44
Components of the expected return:					
Ex ante risk premium	5.40	2.52	4.97	-2.88	2.45
Ex ante safe rate	2.19	3.87	0.98	1.68	-2.89

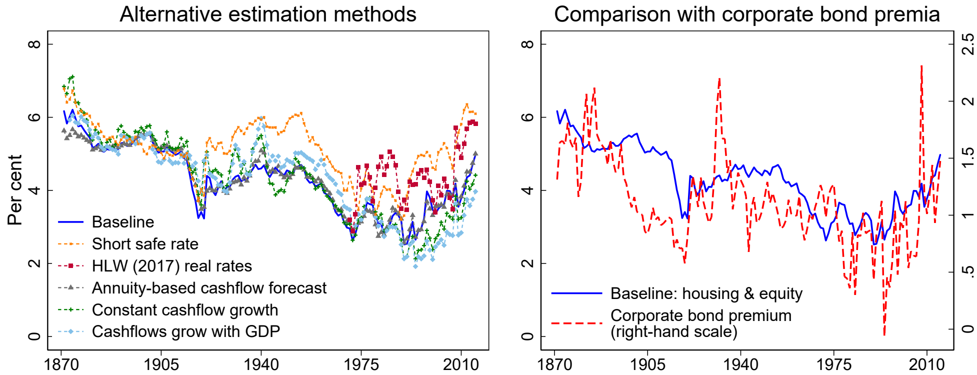
Notes: Unweighted averages of all cross-country observations during the specific time period. The expected return is the sum of the expected yield and expected cashflow growth obtained using a predictability VAR. The risk premium is the difference between the expected return and ex ante safe rate. Ex ante real safe rate is estimated using a Bayesian VAR with slow-moving trends as in Del Negro, Giannone, Giannoni, and Tambalotti (2019).

To assess this, we decompose our aggregate expected return measure into a risk premium and a safe rate component. To measure the ex ante safe rate, we estimate the trend real long-term government bond yield by applying the Bayesian VAR model of Del Negro, Giannone, Giannoni, and Tambalotti (2019) to our data series. The risk premium is computed as the difference between the expected risky return and the safe rate. Figure 2.9 displays the evolution of cross-country average expected return decomposed into these two components (left-hand panel), and the time trend in the ex ante risk premium (right-hand panel). Table 2.3 shows the levels of the expected risky return and its components for selected benchmark years.

The trend in expected returns is largely unrelated to long-run movements in the safe rate: while expected returns show a steady modest decline, safe rates follow a hump shape, increasing up to the 1980s and falling sharply thereafter. Table 2.3 shows that between 1880 and 1990, expected returns fell by 1.2 percentage points, while the safe rate actually increased by 1.7 ppts. After 1990, safe rates declined dramatically by some 2.9 ppts but expected returns only fell by 0.4 ppts. This means that the ex ante risk risk premium exhibits large movements at secular frequency. The long-run trend in the risk premium follows a U shape: a high of 6% in the 1870s followed by a sharp decline to 2.5% in 1990, and an increase to 5% thereafter. Taking stock of these long-run movements, in the late 19th century expected risky returns and risk premia were at historically high levels, while real safe rates were close to their historical average of around 2%. Today, safe rates are at their all-time historical low and approaching negative territory. Expected risky returns are also low, but are substantially higher than safe rates at some 6% thanks to the large positive risk premium.

To check the accuracy of the risk premium trends, the left-hand panel of Figure 2.10 re-estimates these under alternative assumptions for expected cashflow growth and real safe rate. All the estimates follow the same U-shape pattern as the baseline risk premium measure in the right-hand panel of Figure 2.9, with risk premium measures based on short-term interest rates registering larger long-run variation with sharper drops during the 1970s and 1980s and more pronounced increases thereafter. The right-hand panel of Figure 2.10 compares our equity and housing risk premium to the yield-to-maturity credit spread between long-term corporate and government bonds sourced from Kuvshinov (2020). The bond spread is a direct measure of the difference between ex ante discount rates on risky and safe bonds, which gives us a risk premium proxy that does not rely on assumptions about cash-flow growth and inflation expectations embedded in our baseline measure. The level of the corporate bond spread is lower than that of the equity and housing premium, largely owing to the lower riskiness of this asset class. But the credit spread follows the same U-shape trend as our baseline risk premium measure, while also showing notable spikes during the two major global financial crises in the 1930s and 2008-09. As a final check, Appendix Figure 2.A.4 shows that similarly to expected returns,

Figure 2.10. Alternative measures of risk premia



Notes: Unweighted averages of 17 countries. The baseline risk premium estimate uses the safe rate in Del Negro, Giannone, Giannoni, and Tambalotti (2019) and expected cashflow growth in the predictability VAR. The annuity-based, constant and GDP-based cashflow growth forecast measures use alternative expected return estimates which forecast cashflows directly, assume they are constant or use a GDP growth forecast instead. The "short safe rate" specification uses the real ex ante short-term rate instead of the long safe bond yield. The "HLW (2017) real rates" specification uses the safe interest rate estimates of Holston, Laubach, and Williams (2017) from 1971 onwards. The corporate bond spread measures the difference between the 10-year corporate bond and government bond yields, from Kuvshinov (2020).

using alternative cross-country or cross-asset weighting schemes generally results in a somewhat stronger downward trend for the risk premium.

2.4 Drivers of expected returns

What are the drivers of observed trends in expected returns and risk premia? Expected risky returns are the sum of the safe interest rate, reflecting a general willingness to save, and a risk correction reflecting the ex ante market risk premium for holding risky assets. The standard consumption-based model offers more precise expressions for both of these terms, linking the desire to save to expected future growth of consumption and the willingness to bear risk to macroeconomic volatility and the asset's consumption beta. For an investor with power utility $u(c) = c^{1-\gamma}/(1-\gamma)$ and fixed relative risk aversion γ, the expected risky return is then determined as follows (see, for example, Cochrane, 2009):

$$\mathbb{E}(R^{risky}_{t+1}) = \underbrace{\rho + \underbrace{\gamma\mathbb{E}[g^c_{t+1}]}_{\substack{\text{consumption}\\ \text{smoothing}}} \underbrace{-0.5\gamma^2 Var(g^c_{t+1})}_{\substack{\text{precautionary}\\ \text{savings}}}}_{R^{safe}} + \underbrace{\underbrace{\gamma Var(g^c_{t+1})}_{\substack{\text{price of}\\ \text{risk, } \Lambda}} \underbrace{\beta_{R,g^c}}_{\substack{\text{quantity}\\ \text{of risk}}}}_{\text{Risk premium}} \tag{2.8}$$

Above, ρ is the rate of impatience, g^c is consumption growth and $\beta_{R,g^c} = cov(R, g^c)/var(g^c)$ measures the co-movement of asset returns and consumption. Equation (2.8) tells us that expected risky and safe returns will be high if investors expect high future economic growth, increasing their desire to bring forward consumption. If macroeconomic risk is high (a high $Var(g^c)$), investors will want to save more to insure against future consumption movements, driving down the safe rate via the precautionary saving motive. They will also prefer to save in safe assets which provide a better hedge against consumption risk, driving up the risk premium. Risk premia will also be high if risky assets provide a poor hedge against consumption risk (high β_{R,g^c}).

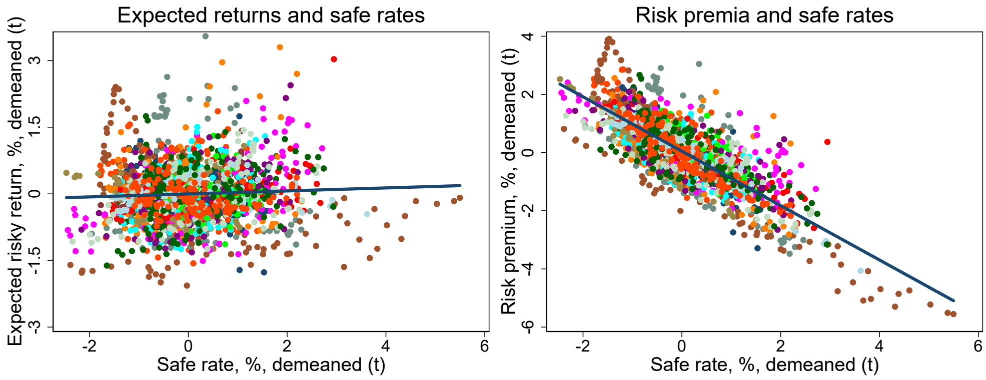
Equation (2.8) allows us to divide up the potential drivers of expected returns and risk premia into two distinct channels, which have the opposite effect on risky versus safe rates. The *growth channel* – changes in $\mathbb{E}[g^c]$ – pushes risky and safe rates in the same direction: high future growth makes investors unwilling to save, in either safe or risky assets. The *risk channel* – or changes in $Var(g^c)$ – pushes safe rates and risk premia in opposite directions, and entails that risky and safe rates are disconnected. High macroeconomic volatility or low risk tolerance will tend to both reduce the safe rate and increase the risk premium. Note that the delineation of risky and safe rate drivers into these two channels applies to a much broader class of models including, for example, more sophisticated consumption-based theories such as those with long-run risk (Bansal and Yaron, 2004), models with time varying risk perception (Pflueger, Siriwardane, and Sunderam, 2020), and models allowing for changes in the relative supply as well as demand for risky assets (Krishnamurthy and Vissing-Jorgensen, 2012; Caballero and Simsek, 2020). All of these theories essentially provide additional means through which the price of risk Λ in equation (2.8) can vary.

Section 2.3.3 shows that trends in risky and safe rates are disconnected, with risk premium and safe rate trends typically moving in opposite directions. This suggests that the risk channel plays an important role in driving long-run trends in expected risky and safe returns. The next section examines this proposition further by looking more closely at the correlations between risky returns, safe rates and risk premia.

2.4.1 The risky-safe rate disconnect

Figure 2.11 and Table 2.4 analyse the co-movement between expected risky returns, real safe rates and risk premia. The the left-hand panel of Figure 2.11 shows a scatterplot of expected returns and safe interest rates and the right-hand panel shows the corresponding graph for risk premia and safe rates. The individual observations are demeaned and coloured by country. Table 2.4 regresses, respectively, the expected risky return (top panel) and the risk premium (bottom panel) on the level of the safe rate. Column 1 runs the full-sample country fixed effects regression, column 2 adds year fixed effects, column 3 considers 5-year changes, and column 4 abstracts from cashflow growth expectations by taking the yield component of expected return only. Columns 5 and 6 run separate regressions for equity and housing, and column 8 correlates realised risky and safe returns.

Figure 2.11. Correlation between expected returns, safe rates and risk premia



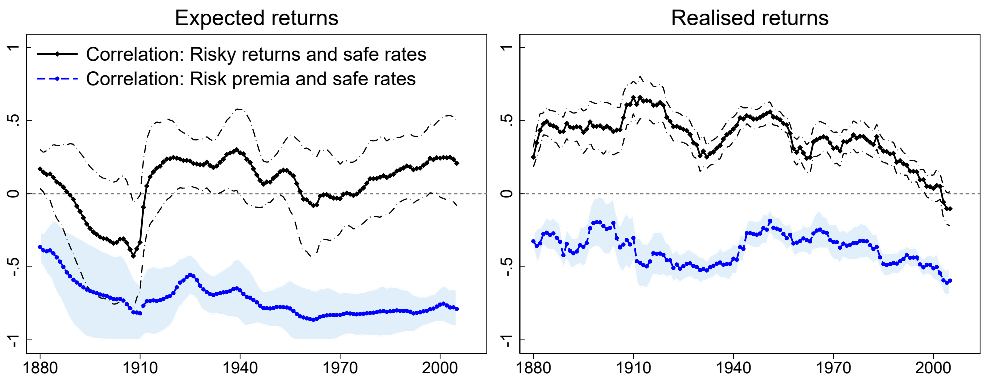
Notes: Scatterplots of expected returns, safe rates and risk premia. Fitted regression lines illustrate the correlation between safe rates and expected returns (left-hand panel) and risk premia (right-hand panel). The individual observations are demeaned and colored by country.

Table 2.4. Co-movement of expected returns, safe rates and risk premia

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Expected return on risky assets						
	Baseline	Year effects	5-year changes	Yield only	Equity	Housing	Realised returns
Safe rates	0.03 (0.08)	0.04 (0.07)	0.18*** (0.04)	0.04 (0.13)	0.08 (0.05)	0.01 (0.13)	0.38*** (0.06)
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects		✓					
R^2	0.00	0.31	0.04	0.00	0.01	0.00	0.11
Observations	1882	1882	1750	1906	2273	1817	1890
	Dependent variable: Risk premia on risky assets						
	Baseline	Year effects	5-year changes	Yield only	Equity	Housing	Realised returns
Safe rates	-0.97*** (0.08)	-0.96*** (0.07)	-0.82*** (0.04)	-0.96*** (0.13)	-0.92*** (0.05)	-0.99*** (0.13)	-0.57*** (0.07)
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects		✓					
R^2	0.68	0.78	0.48	0.40	0.61	0.46	0.20
Observations	1882	1882	1750	1906	2273	1817	1890

Notes: Regressions of expected returns or risk premia on the safe rate. The expected return and risk premium is defined as the average of the corresponding series for housing and equity. Ex ante real safe rate is estimated using a Bayesian VAR with slow-moving trends as in Del Negro, Giannone, Giannoni, and Tambalotti (2019). Column (1) includes only country fixed effects and Column (2) adds year fixed effects. Column (3) considers 5-year changes, and column (4) uses yield *dp* only as a proxy for expected return. Columns (5) and (6) consider housing and equity returns separately. Realised risky return is the average of total real returns on housing and equity; realised safe return is the real government bond return. Standard errors are clustered by country and year. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

Figure 2.12. Correlation between expected returns, safe rates and risk premia through time



Notes: Left-hand panel: Pairwise correlation coefficients between the expected risky return and safe rate, and ex ante risk premium and safe rate. Right-hand panel: pairwise correlation coefficients between realised risky returns, risk premia and real government bond returns. Rolling centered 20-year windows. Shaded areas are 90% confidence intervals, using country-clustered standard errors. Underlying data are demeaned at country level.

The correlation between expected returns and safe rates is very weak and close to zero for all regression specifications. As a consequence, risk premia and safe rates are strongly negatively correlated. The magnitude of this negative correlation is rather stark: under the baseline specification in Table 2.4 column 1, an increase in the safe rate brings an almost one-for-one decline in the ex ante risk premium and vice versa, such that the expected return remains broadly unchanged. This lack of correlation between expected returns and risk premia, and the strong negative correlation between safe rates and risk premia, hold up under all the alternative specifications in columns 2–4 of Table 2.4, and hold separately for equity and housing. This shows that not only do risky and safe rates follow different trends, but they are more generally disconnected. Column 8 shows that realised risky and safe returns are somewhat more correlated than expected returns, but this correlation remains low and realised risk premia and safe rates are strongly negatively correlated. Appendix Table 2.A.3 shows that the correlation between realised returns and safe rates remains weak under alternative regression specifications, and separately for equity and housing.

Figure 2.12 examines how the correlation between risky returns, safe rates and risk premia has changed over time. The left-hand panel shows the 20-year rolling window correlation coefficients between expected risky returns and safe rates (black diamonds) alongside those between risk premia and safe rates (blue circles), together with the corresponding 90% confidence intervals. For example, the data point for 1880 shows the correlation between risky and safe returns in the pooled sample covering the time period 1871–1890. The right-hand panel shows the corresponding correlations for realised returns. The absence of expected return co-movement is remarkably stable over time: throughout the whole sample, the correlation between risky and safe returns is around zero and the correlation between risk premia and safe rates is close to -1. The risky-safe rate disconnect in Figure 2.11 and Table 2.4 is not driven by some distant historical time period: if anything, risky and safe rates were somewhat less disconnected in late 19th century than today. The realised risky and safe return co-movement also remains low through time, but displays larger variation. The two world wars saw high co-movement with low returns on both risky and safe assets during this period, while the recent decades have seen an increasing divergence between the two.

Existing literature has highlighted the low and time-varying nature of the correlation between realised equity and bond returns in the US, but the sources behind this lack of co-movement and its variation over time remain difficult to pin down (Shiller and Beltratti, 1992; Baele, Bekaert, and Inghelbrecht, 2010; Campbell, Pflueger, and Viceira, 2020). We confirm that these patterns of realised return co-movement extend to broader cross-country data including housing as well as equity. Our findings also help shed some light on the likely drivers of the realised return co-movement. The overall lack of co-movement found in the literature is likely to reflect the general disconnect between expected risky and safe returns, and is an indicator that changes in the price of risk are an important driver of the returns on these two asset classes. The time-varying nature of realised return co-movement is, however, likely driven by unexpected shocks affecting both asset classes.

Expected risky returns are disconnected from safe rates. This suggests that variation in risk and risk premia – either through time-varying price of risk Λ or quantity of risk β in equation (2.8) – is a key driver of the expected risky return. Moreover, since the price of risk can affect the safe rate through the precautionary saving motive, changes in risk can also be a key driver of the safe rate. In fact, as equation (2.8) shows, increases in the price of risk drive the safe rate down at the same time as driving up risk premia, meaning that overall effect on expected returns is muted. These partly offsetting movements help explain both the near-zero safe-risky rate correlation and the strongly negative risk premium – safe rate correlation in the data. They also help reconcile the substantial secular safe rate and risk premium variation with the relative stability of the expected risky return (Figure 2.9). The next section maps out the trends in the price and quantity of risk and investigates their contribution to long-run changes in expected risky returns, safe rates and risk premia.

2.4.2 Expected returns and risk

We have shown that risky and safe rates follow different trends and are disconnected at shorter time horizons. This suggests that changes in the *price of risk*, which drive up the risk premium and drive down the safe rate, are a key driver of expected risky and safe returns. As shown in equation (2.8), the underlying driver of the price of risk in standard macro-finance theory is consumption volatility. High volatility of consumption means that investors should be willing to pay a high price for hedging consumption movements, which drives up the risk premium. At the same time, a greater desire to hedge against large consumption drops increases the precautionary saving motive and reduces the safe rate.

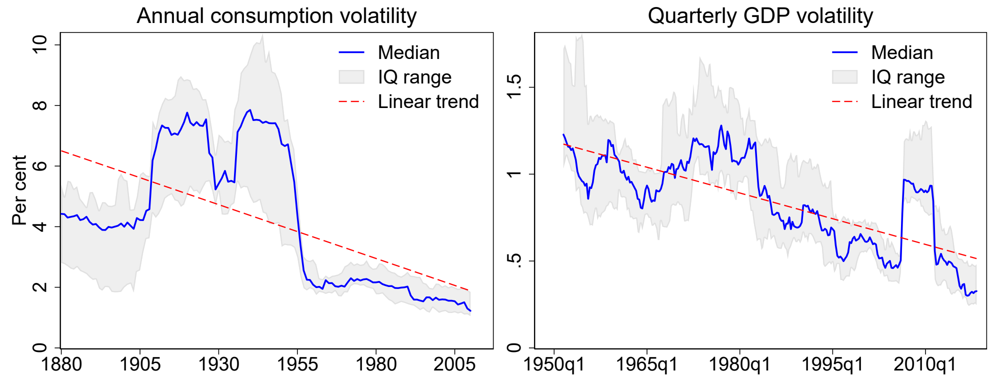
Figure 2.13 shows how consumption volatility in the 17 countries in our sample has evolved over the long run. The left-hand panel plots the standard deviation of real consumption growth in annual data stretching back to 1870 using 20-year centered rolling windows, and the right-hand panel shows the post-1950 quarterly real GDP growth volatility over 5-year rolling windows. Both the annual and the more recent quarterly data paint the picture of a long-run decline in consumption volatility, with standard deviation of annual real consumption growth falling from more than 4 ppts per year in the late 19th century to 1–2 ppts per year today. Equation (2.8) tells us that the risk premium should move proportionally with the variance of consumption growth:

$$\Delta RP \approx \Delta Var(g^c)\gamma\beta_{R,g^c} \tag{2.9}$$

This means that, all other things being equal, a halving in consumption growth variance should bring about a halving in the risk premium. Columns 1–3 of Table 2.5 show three snapshots of the levels of the risk premium and variance of annual consumption growth: 1880, corresponding to high historical risk premium levels; 1990, corresponding to the trough in the long-run risk premium trend; and 2010 corresponding to the recent uptick. The risk premium is for the specific year and consumption growth variance is for the 20-year centered window around that year. Columns 4 and 5 show the changes in these variables between these selected years (with -50% a halving and +100% a doubling). Table 2.5 shows that between years 1880 and 1990, the risk premium more than halved, falling from 5.4 to 2.5 ppts. The variance of consumption growth fell by even more – roughly four-fifths. This means that, within the framework of standard asset pricing theory, the entirety of the risk premium decline during the first 100 years of our data can be accounted for by the decrease in consumption volatility. This lower consumption volatility should have also increased the safe rate as investors became less willing to hedge against consumption drops, thereby muting the overall impact on the expected return.

Standard macro-finance theories struggle to match the average level of the equity premium (Mehra and Prescott, 1985), and short-run variation in the risk premium is difficult to reconcile with market expectations of future volatility (Dew-Becker, Giglio, Le, and Rodriguez, 2017). But when it comes to long-run trends, our results show that standard theory actually does a good job of explaining the observed risk premium movements in the data. The implications of our findings are consistent with those of Lettau, Ludvigson, and Wachter (2008), who use an asset pricing model with long-run risk to show that the decline in US consumption volatil-

Figure 2.13. Macroeconomic volatility over the long run



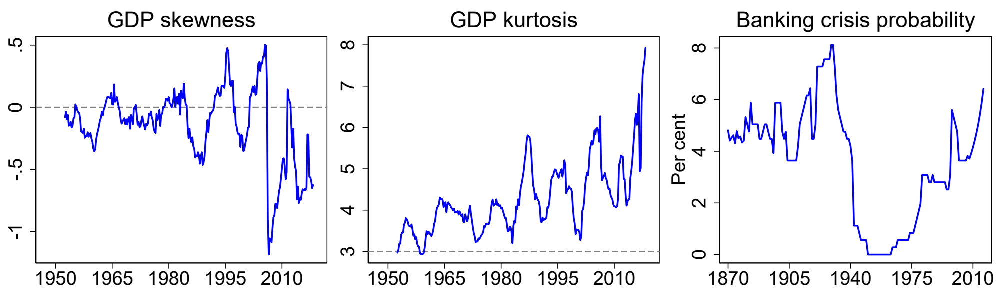
Notes: Data for 17 countries. Left-hand panel: centered rolling 20-year windows; right-hand panel: centered 5-year windows. Underlying data are winsorised at the 0.5% level.

Table 2.5. Long-run changes in the risk premium and macroeconomic risk

	(1)	(2)	(3)	(4)	(5)
		Level		Relative change	
	1880	1990	2010	1880–1990	1990–2010
Ex ante risk premium, %	5.40	2.52	3.78	-53%	+50%
Consumption variance, $\%^2$	20.59	3.76	2.07	-82%	-45%

Notes: Annual data for 17 countries. Ex ante risk premium is the cross-country average risk premium level in that year. Consumption variance is the square of average country-level volatility in the 20-year window around that year. Levels in percentage points and percentage points squared, relative changes in percent.

Figure 2.14. Macroeconomic tail risks



Notes: Data for 17 countries. Left-hand panel: annual unconditional crisis probability calculated using centered 20-year windows. Middle- and right-hand panels: skewness and kurtosis of quarterly GDP growth based on centered 5-year windows. Underlying GDP growth data are winsorised at the 0.5% level. Quarterly GDP data are sourced from Monnet and Puy (2019).

ity during the 1990s can explain the fall in the equity premium during this time period.

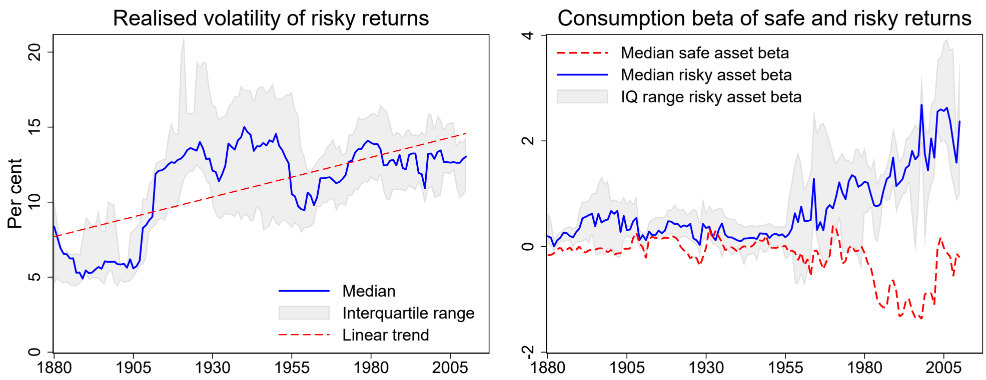
Even though declining consumption volatility can explain the long-run decline in the risk premium, it does not explain the recent uptick. Column 5 of Table 2.5 shows that between 1990 and 2010, risk premia increased by half while the variance of annual consumption growth rates declined further. One force which may be driving the risk premium up over recent decades is an increase in macroeconomic tail risk. Even though consumption is less volatile now than historically, large negative changes in consumption growth are now much more likely than, say, 50 years ago. This can be seen from Figure 2.5, which shows estimates for the kurtosis and skewness of quarterly GDP growth rates. The GDP growth distribution was close to normal at the beginning of the sample (kurtosis of around 3 and skewness of zero), but over recent decades the skew has become more negative and the tails have widened. In line with long-run risk and disaster risk models (Bansal and Yaron, 2004; Gabaix, 2012), these increases in tail risk should have increased the risk premium throughout recent decades.

Recent decades also saw systemic financial risks reappear after being more or less absent during the middle of the 20th century. The left-hand panel of Figure 2.5 shows the average systemic banking crisis probability calculated using twenty year windows, using the narrative-based crisis definition of Schularick and Taylor (2012). A value of, for example, 4% in 1990 means that crisis observations comprise 4% of total country-year observations in years 1980–2000.[10] Systemic risk was high between 1870 and 1940 with banking crises happening about once every 20 years (5 percent probability), but seemingly disappeared during the Bretton Woods era. Recent decades have seen the return of systemic risks, first in individual countries (Japanese and Scandinavian banking crises) and then more generally with the global financial crisis. Since crises are typically followed by low GDP growth, this re-emergence has contributed to the increasing macroeconomic tail risks in the left-hand panels of Figure 2.14. A higher crisis probability can also increase the risk premium level directly. Muir (2017) shows that financial crises tend to be associated with risk premium increases above and beyond any drops in GDP, a fact that can be explained by crises impairing the risk bearing capacity of intermediaries which price financial assets.

In addition, our evidence suggests that the riskiness of the assets themselves, represented by β_{R,g^c} in equation (2.8), increased in recent decades. Figure 2.15 shows the long-run trends in two proxies for the quantity of risk β: simple unconditional return volatility and consumption beta, again calculated over rolling 20-year periods. The unconditional return volatility is near its historical high, but shows little change over the past three decades. Consumption beta of risky assets has, however,

10. Note that we only count the first year of the crisis as a crisis observation, so the figure is more exactly interpreted as the probability of the emergence of a new systemic banking crisis.

Figure 2.15. Trends in the quantity of risk



Notes: Data for 17 countries. Centered rolling 20-year windows. Underlying data are the average of real equity and housing return. Realised volatility is the rolling standard deviation of annual real returns within the 20-year window. Consumption beta is the covariance of real returns with consumption growth, scaled by the variance of consumption growth: $\beta_{R,g^c} = Cov(R, g^c)/Var(g^c)$. The solid line and the shaded area are, respectively, the mean and interquartile range of the individual country data in each year. The dashed line represents the linear trend.

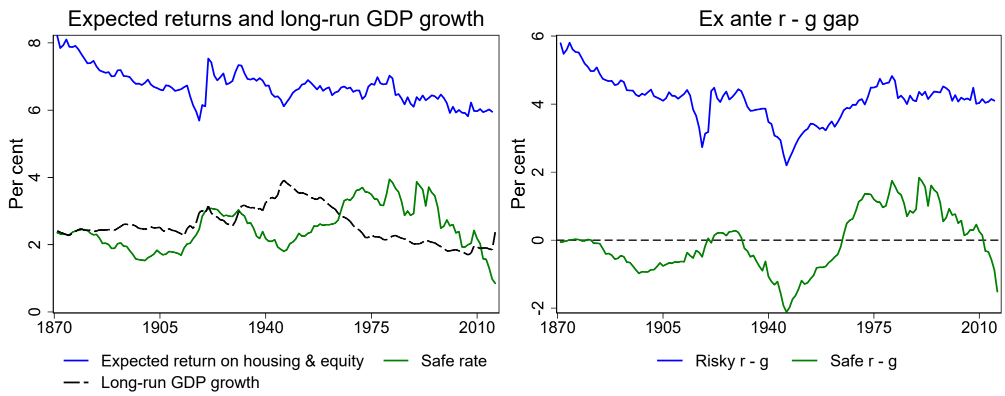
increased markedly from around 1 in 1980 to close to 2 today. Further to this, the government bond beta (dashed red line in Figure 2.15 right-hand panel) has fallen sharply and actually turned negative, suggesting that safe assets have become a better hedge for macroeconomic risk. The increasing risky and declining safe asset beta help explain the low levels of the safe rate and the high risk premium we observe today.

Other factors are, of course, likely to also be at play. Kopecky and Taylor (2020) argue that the recent risk premium increase and safe rate decline can be explained by population ageing with older households preferring to save in safer assets. Turning to supply rather than demand for assets, Caballero, Farhi, and Gourinchas (2017b) argue that a shortage of safe assets may have reduced the safe rate and increased the risk premium over the past decade. A safe asset shortage should also have increased the convenience premium on safe government debt, further driving down the safe rate without a corresponding reduction in the risky rate (Krishnamurthy and Vissing-Jorgensen, 2012; Del Negro, Giannone, Giannoni, and Tambalotti, 2019). Even aside from these other influences, however, trends in macroeconomic volatility and asset riskiness can together account for the long-run changes in prices of risky and safe assets as stipulated by standard asset pricing theory.

2.4.3 Expected returns and growth

A growing literature on the secular stagnation hypothesis has linked the recent safe rate decline to a slowdown in growth and increased willingness to save (Baldwin and Teulings, 2014; Holston, Laubach, and Williams, 2017; Summers and Rachel, 2019). But if risk is an important driver of long-run risky and safe rate trends, we

Figure 2.16. Expected returns and GDP growth



Notes: Unweighted averages of 17 countries. Expected risky return is the expected yield plus expected real cashflow growth, averaged across equity and housing. Trend real safe rate is estimated using the method of Del Negro, Giannone, Giannoni, and Tambalotti (2019). Long-run GDP growth is the annuity value of real GDP growth from the next year onwards, with post-2015 growth computed using the 2060 GDP forecast from the *OECD Economic Outlook*.

would expect to see some divergence between expected risky and safe returns on one hand, and the rate of economic growth on the other. Put differently, the *ex ante* risky and safe $r-g$ gaps may be time varying. This time variation, in turn, carries important implications for the dynamics of wealth inequality (Piketty, 2014), the costs and benefits of issuing public debt (Blanchard, 2019) and the economy's dynamic efficiency (Barro, 2020).

When examining these issues, existing research has mostly focussed on the *ex post* $r-g$ gaps between realised returns and growth rates (Piketty, 2014; Jordà, Knoll, Kuvshinov, Schularick, and Taylor, 2019). While these realised gaps are informative about the past evolution of inequality, public debt and returns on capital, they tell us relatively little about how these variables are likely to evolve going forward, and whether these past changes are driven by ex ante risk and saving preferences or unexpected shocks. We therefore estimate *ex ante* $r-g$ *gaps* by combining our expected return estimates with an estimate of the long-run growth in GDP. To do this, we compute the annuity value of GDP growth from year $t+1$ onwards in the same way we computed long-run cashflow growth $\tilde{g}$ in Section 2.2, by summing future realised growth rates at exponentially decaying weights (see footnote 2). We use realised GDP growth for years up to 2015, and compute expected growth after 2015 using the *OECD Economic Outlook* forecast for GDP in 2060.

The left-hand panel of Figure 2.16 shows the expected return and trend real safe rate estimates from Section 2.3 alongside the annuity value of real GDP growth, and the right-hand panel shows the corresponding ex ante $r-g$ gaps. Table 2.6 additionally shows the levels of the gap at selected points in time as well as the corresponding changes. Neither the risky nor the safe rate trend show a strong correspondence with trend long-run GDP growth. This means that both risky and safe $r-g$ gaps vary sub-

Table 2.6. Ex ante $r - g$ gaps through time

	(1)	(2)	(3)	(4)	(5)
	Level			Absolute change	
	1880	1990	2015	1880–1990	1990–2015
Safe $r - g$ gap	-0.04	1.81	-0.92	+1.85	-2.73
Risky $r - g$ gap	5.36	4.29	4.01	-1.07	-.28
equity $r - g$	4.30	4.43	4.59	+.14	+.15
housing $r - g$	6.43	4.15	3.43	-2.28	-.71

Notes: Averages of 16 countries, percentage points. Columns 1–3 show the levels of the variable in that year, and columns 4–5 show the absolute percentage point change between the cross-country averages in the respective years. Risky r is the average of expected returns on equity and housing, with expected returns calculated as the sum of expected yield and cashflow growth. Safe r is the long-term trend real safe rate computed using the method of Del Negro, Giannone, Giannoni, and Tambalotti (2019). g is the annuity value of economic growth from year $t + 1$ onwards, with post-2015 growth computed using the 2060 GDP forecast from the *OECD Economic Outlook*. To maintain sample consistency, the table excludes Canada, for which we have no housing data.

stantially over time. The safe $r - g$ gap (solid green line) displays no clear trend for most of the sample, but shows a sharp increase in the 1970s/80s and a pronounced decline afterwards. The current safe $r - g$ gap level is around -0.9%, close to the historical low observed during World War 2 and below its historical average of around zero. In line with the declining expected return, the risky $r - g$ gap has also declined from 5.4% in 1880 to 3.0% in 2015 (Table 2.6). Despite this decline, however, it remains high and positive. Appendix Figure 2.A.8 shows the corresponding trends and gaps for realised risky and safe returns. These data confirm that the risky $r - g$ gap is highly positive, and the safe $r - g$ gap is close to zero on average, but as before, the large volatility of realised returns makes it difficult to infer the corresponding trends.

Table 2.7 shows shows the full-sample average $r - g$ gaps in individual countries alongside their levels at the end of our sample in 2015. The full-sample ex ante $r - g$ gaps are unambiguously positive for risky assets and around zero for safe assets. In around half the countries, the full-sample safe $r - g$ gap is negative whereas the risky $r - g$ gap is positive in every country. Risky $r - g$ gaps also display larger cross-country variation than safe $r - g$ gaps, ranging from 1.5% in France to 8% in Finland. The risky $r - g$ gap today is close to its historical average of around 4%. The safe $r - g$ gap is, however, close its historical row at around -0.9%, and as low as -2% in some countries.

The observed trends in ex ante $r - g$ gaps are difficult to square with a growth-centric view of secular stagnation and low interest rates. Low growth should simultaneously affect risky and safe rates. Yet, while the risky $r - g$ has fallen over the long-run, it has only mildly decreased in recent decades, a period of sharp safe $r - g$

Table 2.7. Ex ante $r - g$ gaps by country

Country	Full Sample		2015	
	$r^{\text{risky}} - g$	$r^{\text{safe}} - g$	$r^{\text{risky}} - g$	$r^{\text{safe}} - g$
Australia	3.10	-0.57	2.94	-1.54
Belgium	4.15	0.51	3.21	-1.16
Canada	4.42	-0.32	5.53	-0.22
Denmark	4.97	0.61	3.80	-1.35
Finland	7.69	-0.09	8.43	-1.10
France	1.75	-0.53	1.11	-0.69
Germany	6.52	0.74	6.12	-0.51
Italy	3.99	0.31	4.91	0.55
Japan	3.19	-0.50	4.73	-0.98
Netherlands	5.37	-0.06	5.00	-1.72
Norway	3.70	-0.39	3.47	-1.97
Portugal	2.12	0.26	2.82	0.84
Spain	3.35	-0.62	3.58	-1.00
Sweden	4.55	-0.14	3.32	-1.46
Switzerland	3.88	0.07	3.91	-1.37
UK	3.54	0.12	2.93	-0.56
USA	3.19	-0.62	3.88	-0.66
Average, unweighted	4.15	-0.08	4.10	-0.88
Average, weighted	3.84	-0.17	4.02	-0.69

Note: r^{risky} is the expected return on housing and equity. r^{safe} is the trend real safe rate. g is the annuity value of future economic growth. Canadian data are for equities only. The average, unweighted and average, weighted figures are respectively the unweighted and real-GDP-weighted arithmetic averages of individual country gaps. The averages are slightly different to those in Table 2.6 because data in Table 2.6 exclude Canada to ensure consistency across the housing and equity asset classes.

declines. The recent relative stability of the expected return on risky wealth is also mirrored in estimates of the marginal product of capital computed using national accounts data (Gomme, Ravikumar, and Rupert, 2015).

Taken together, these trends carry important implications for the dynamics of capital accumulation, public debt and wealth inequality. In a wide range of neoclassical growth models, a positive gap between the return on productive capital and the economy's growth rate means that investment and capital accumulation increase long-run consumption growth, and hence the economy is dynamically efficient (Ramsey, 1928; Diamond, 1965; Abel, Mankiw, Summers, and Zeckhauser, 1989). Barro (2020) extends the standard model to incorporate both risky and safe assets, and shows that the $r-g$ condition applies to the expected risky return, while dynamic efficiency is compatible with safe $r-g$ gaps being below zero. We show that despite the substantial increase in capital-to-income ratios during the final few decades of the 20th century (Piketty and Zucman, 2014), the increase in the ex ante risk premium has meant that the expected return on capital has remained high, the risky $r-g$ gap has remained positive, and advanced economies in our sample are far from dynamically inefficient.

Blanchard (2019) argues that while low safe $r-g$ gaps make government borrowing cheap to finance, a high risky $r-g$ gaps means that additional public debt would incur a high opportunity cost in terms of foregone private investment, since this foregone investment would yield a high rate of return r. We confirm Blanchard (2019)'s finding of a low safe $r-g$ gap, both now and historically. But the high risky $r-g$ gap means that even though public debt is cheap to finance, it carries a high opportunity cost. Turning to the dynamics of wealth inequality, a high risky $r-g$ gap means that wealth growths at a much higher rate than income and equilibrium wealth inequality will tend to be high (Piketty, 2014; Benhabib and Bisin, 2018). The trends in our risky $r-g$ gap do show some correspondence with the long-run evolution of wealth inequality, with high levels in the late 19th century followed by a fall up until the 1950s and a subsequent increase.

The impact of return differentials on wealth inequality is further exacerbated by the variation in $r-g$ gaps across individual asset classes. Existing research suggests that households in the lower part of the wealth distribution hold most of their wealth in safe assets such as deposits, the middle of the wealth distribution holds mostly housing wealth and the top part – mostly equity wealth (Kuhn, Schularick, and Steins, 2020; Martıúnez-Toledano, 2020; Garbinti, Goupille-Lebret, and Piketty, 2021). Table 2.6 shows that out of these three asset classes, the safe $r-g$ gap has declined the most, while the equity $r-g$ gap has changed very little. In the late 19th century, the housing $r-g$ gap was larger than that on equity, but the current equity $r-g$ gap of 4.6% is higher than the housing gap of 3.4%, and both are substantially higher than the safe $r-g$ gap of -0.9%. This means that not only is wealth more likely to grow faster than income on average, but asset holdings of the wealthy are likely to yield higher returns, further exacerbating existing inequality. Taken together, these facts imply that wealth of the rich is likely to continue growing substantially faster than income, and steady-state levels of wealth inequality are likely to remain high.

2.5 Conclusion

The expected return on risky assets has been declining for the past 145 years, falling from 8% in the 1870s to 6% today. This long-run decline means that past realised returns are likely to somewhat overstate future returns to potential investors. Still, expected returns remain high and risk premia – far above zero, so there is little sign of the equity and housing premium puzzles disappearing. This high expected return level also means that the ex ante risky $r-g$ gap remains high, meaning that despite recent increases in advanced-economy wealth-to-income ratios, capital accumulation is yet to run into sharply diminishing returns.

Even though safe rates have also declined over recent decades, their movements are in general disconnected from the risky rate. This means that changes in risk are a key determinant of both risky and safe rates, and in fact much of the historical expected return decline can be linked to a fall in the risk premium and a corresponding decline in consumption volatility. Other factors which affect the relative demand and supply of risky and safe assets – such as safe asset shortages (Caballero and Farhi, 2014) and changes in market access (Iachan, Nenov, and Simsek, 2020) – provide additional channels through which changes in the price of risk can drive risky and safe returns in opposite directions. A further investigation of these other risk-based channels and their role in shaping long-run risky and safe rate trends offers a fruitful avenue for future research.

Appendix 2.A Additional results

2.A.1 Summary statistics

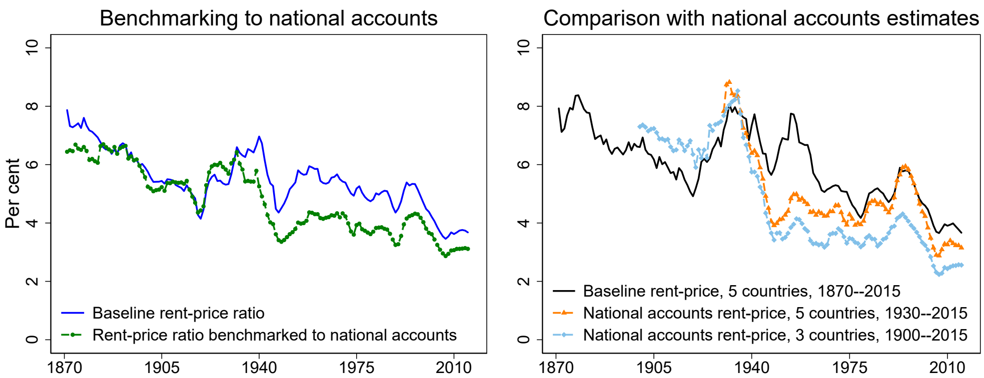
Table 2.A.1. Summary statistics

	Obs.	Mean	S.D.	Min	Max
Expected return on risky assets	1759	6.58	1.60	2.74	12.51
Expected return on equity	1759	6.62	2.05	1.10	14.47
Expected return on housing	1645	6.47	1.83	3.05	14.22
Ex ante risk premia on risky assets	1759	4.03	1.70	-2.35	9.69
Ex ante risk premia on equity	1759	4.06	2.11	-2.80	10.84
Ex ante risk premia on housing	1645	3.97	1.94	-2.79	10.55
Discount rate on risky assets	1759	4.54	1.43	0.56	11.74
Equity dividend yield	1759	3.96	1.65	0.07	14.19
Housing rental yield	1645	5.21	1.98	0.50	13.08
Long real safe rate (Del Negro et al., 2019)	1759	2.55	1.13	0.10	8.55
Short real safe rate (Del Negro et al., 2019)	1759	1.59	1.10	-1.16	7.20
Annuity value of real GDP growth	1759	1.99	0.67	0.45	4.71
Dividend growth rate (winsorized at 1% level)	1759	3.04	21.70	-49.76	95.95
Rent growth rate (winsorized at 1% level)	1645	1.55	6.85	-17.94	33.26

Notes: Annual data, 1870–2015, per cent.

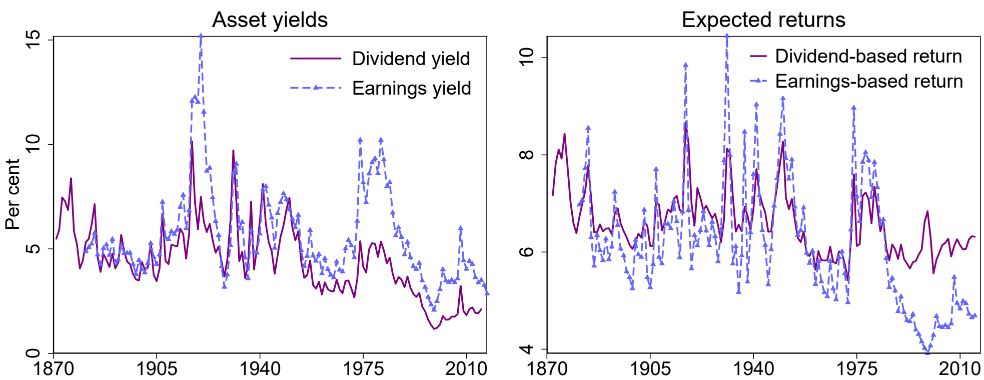
2.A.2 Trends in expected returns and risk premia: additional details

Figure 2.A.1. Comparison of rent-price ratios to national accounts data



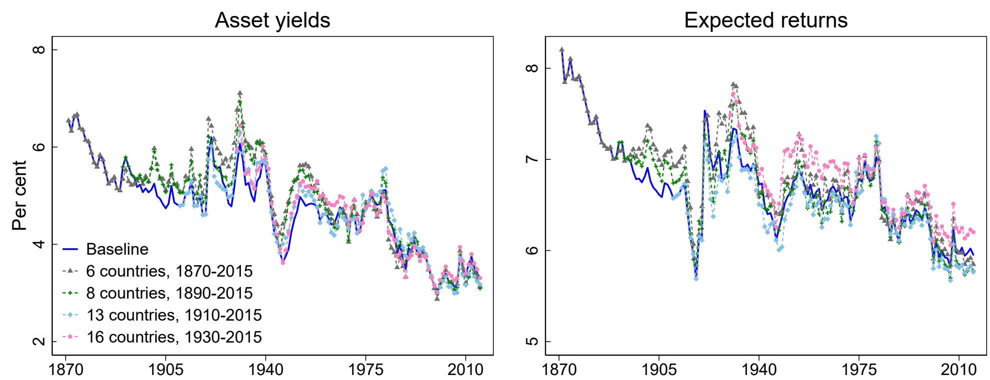
Notes: Left-hand panel: unweighted averages of 16 countries. The baseline uses the rent-price approach of Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) to construct historical rent-price ratios. The series benchmarked to national accounts uses the balance sheet approach estimates where possible, extrapolating using the growth in house prices and rents where these are not available. Right-hand panel: averages of countries for which we have long-run balance sheet approach data. The group of 5 countries includes Denmark, France, Germany, Sweden and USA. The group of 3 countries includes Denmark, France and Germany. The balance sheet approach yield is calculated as total rental income minus running costs (all non-tax housing expenditures and depreciation) as a share of housing wealth. Balance sheet approach measures in the right-hand panel do not rely on house price and rent growth extrapolation.

Figure 2.A.2. Comparison of dividend and earnings yields for the US



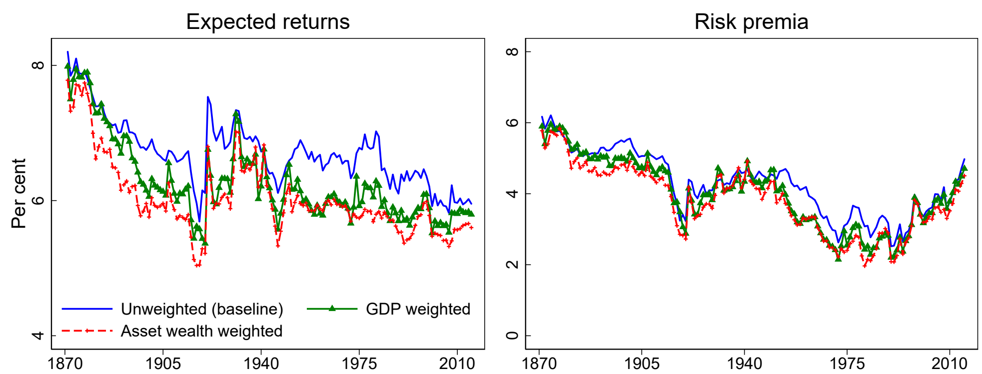
Notes: US data. The earnings yield is the the cyclically adjusted total return earnings-price ratio (inverse of P/E10 CAPE) from Shiller (2015), December values. The dividend-based expected return is our baseline estimate. The earnings-based expected return is the sum of the dividend-price ratio and expected earnings growth. To guard against the potential effects of share buybacks on total asset yields, we fix the growth of the dividend-price ratio to equal that of the earnings-price ratio from 1982 onwards.

Figure 2.A.3. Alternative sample groupings



Notes: Unweighted averages of groups of countries for which we have the data on both housing and equity yields and expected returns over the selected time period. Data for Canada use equities only. The 6 countries with long-run data going back to the 1870s are Canada, Denmark, France, Germany, Norway and Sweden. The 2 additional countries with data from 1890 onwards are Belgium and USA. The 5 additional countries with data from 1910 are Australia, Netherlands, Spain, Switzerland and the UK. The 3 additional countries with data from 1930 onwards are Finland, Italy and Japan. Portugal is only included in baseline estimates, from 1948 onwards.

Figure 2.A.4. Alternative weighting schemes



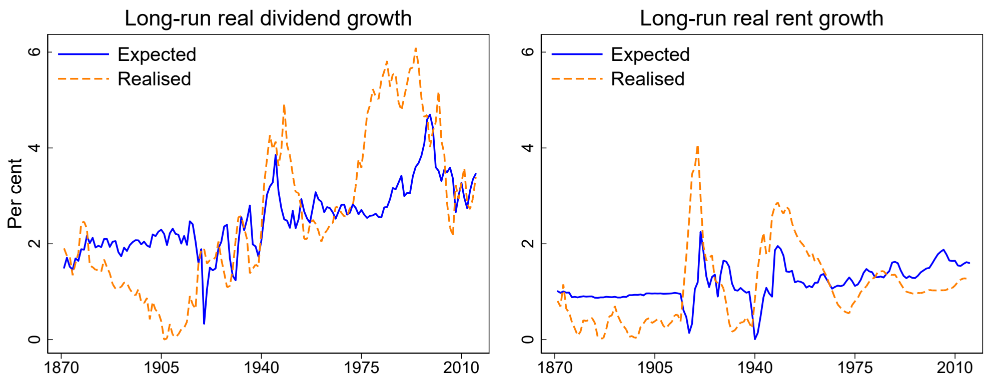
Notes: Baseline is the unweighted average of 17 countries. GDP-weighted average weights country-level observations by the respective country's real GDP level. Wealth-weighted average weights equity and housing returns within country by equity and housing market capitalization, and weights country-level returns by the level of risky wealth of the respective country.

Table 2.A.2. Return and cashflow predictability: full VAR results for selected time periods

	(1)	(2)	(3)	(4)	(5)	(6)
	Equity			Housing		
	r_t	dg_t	dp_t	r_t	dg_t	dp_t
VAR for years 1870–1910:						
r_{t+1}	0.09*	-0.01	0.06***	-0.23**	0.16	0.13***
	(0.05)	(0.02)	(0.02)	(0.10)	(0.15)	(0.02)
dg_{t+1}	0.23***	-0.19***	-0.09***	-0.05*	0.01	0.00
	(0.07)	(0.05)	(0.02)	(0.03)	(0.13)	(0.01)
dp_{t+1}	0.14**	-0.19***	0.88***	0.19*	-0.16	0.91***
	(0.06)	(0.06)	(0.02)	(0.10)	(0.11)	(0.02)
Observations	545	545	545	304	304	304
VAR for years 1910–1940:						
r_{t+1}	0.24***	0.00	0.05**	0.08	0.29***	0.07***
	(0.05)	(0.04)	(0.02)	(0.09)	(0.08)	(0.01)
dg_{t+1}	0.24***	-0.06	-0.14***	-0.02	0.53***	-0.03***
	(0.07)	(0.07)	(-0.02)	(0.04)	(0.06)	(0.01)
dp_{t+1}	-0.01	-0.07	0.84***	-0.10	0.25***	0.94***
	(0.06)	(0.06)	(0.03)	(0.08)	(0.06)	(0.02)
Observations	648	648	648	507	507	507
VAR for years 1940–1980:						
r_{t+1}	0.16***	-0.03	0.15***	0.17**	0.18***	0.07***
	(0.05)	(0.04)	(0.02)	(0.07)	(0.06)	(0.01)
dg_{t+1}	-0.02	0.11	-0.03	0.00	0.46***	-0.05***
	(0.07)	(0.08)	(0.04)	(0.06)	(0.07)	(0.01)
dp_{t+1}	-0.18**	0.15**	0.85***	-0.18**	0.29***	0.92***
	(0.07)	(0.07)	(0.04)	(0.08)	(0.08)	(0.01)
Observations	667	667	667	577	577	577
VAR for years 1970–2010:						
r_{t+1}	0.04	-0.00	0.05***	0.56***	-0.01	0.05***
	(0.04)	(0.04)	(0.02)	(0.07)	(0.07)	(0.01)
dg_{t+1}	-0.01	-0.20***	-0.10***	-0.00	0.31***	-0.02***
	(0.05)	(0.07)	(0.02)	(0.02)	(0.09)	(0.01)
dp_{t+1}	-0.05	-0.20***	0.88***	-0.60***	0.33***	0.98***
	(0.05)	(0.07)	(0.03)	(-0.07)	(0.10)	(0.01)
Observations	693	693	693	649	649	649

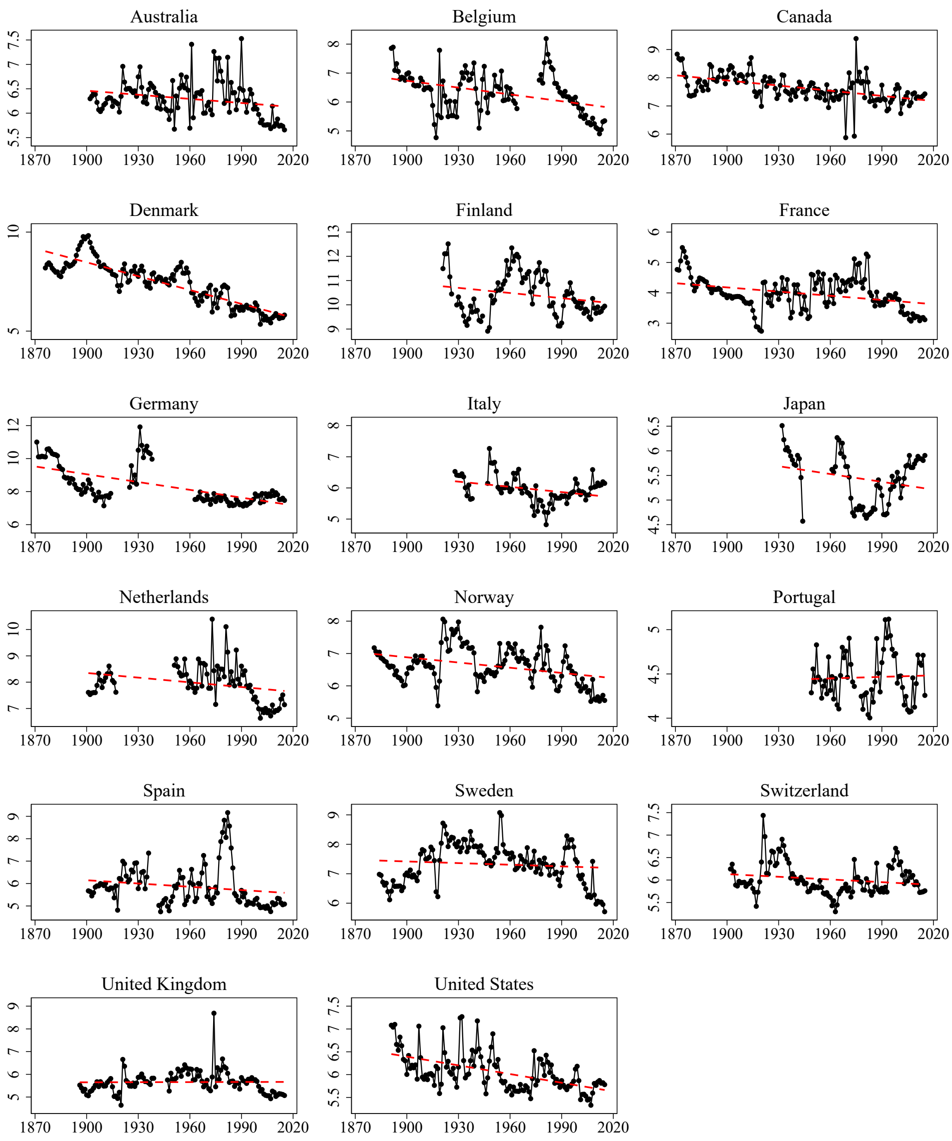
Note: Rolling window VAR estimates using GMM subject to present value moment constraints, accounting for cross-sectional and time dependence in standard errors. Variables are log real total equity or housing return *r*, log real dividend or rent growth *dg*, and log of dividend-price or rent-price ratio *dp*, demeaned at country level. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

Figure 2.A.5. Expected and realised cashflow growth



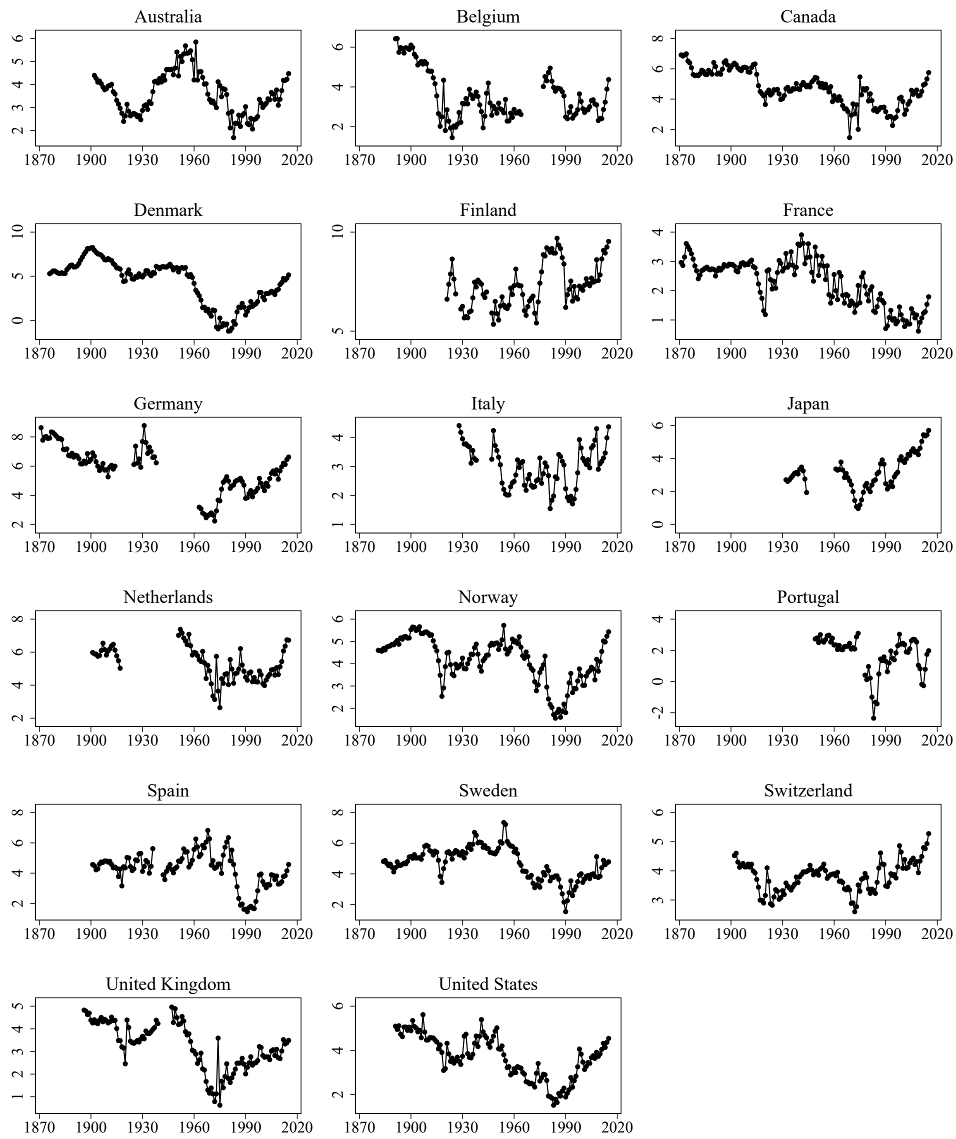
Notes: Expected cashflow growth is the VAR forecast of the present value of future cashflow growth using today's dividend- or rent-price ratio, returns and cashflows, $\left[(1-\rho_i)\hat{\mathbb{E}}\sum_{s=0}^{\infty}\rho_i^s dg_{i,t+1+s}\right]+\overline{DG_{i,j}}-1$, with the VAR estimated over a 40-year rolling window from $t-40$ to t (apart from the first 40 years, where the VAR for 1870–1910 is used to make the forecast for e.g. 1890). Realised cashflow growth is the annuity value of future cashflow growth from $t+1$ onwards, discounted at the asset-specific sample average rate of return r. Realised cashflow growth data are winsorized at 1% level.

Figure 2.A.6. Expected returns in individual countries



Notes: The expected return is the average of housing and equity, measured as the yield plus expected cashflow growth. Dashed lines show the country-specific linear trends. All data are in percent. Data for Canada are for equities only.

Figure 2.A.7. Ex ante risk premia in individual countries



Notes: The risk premium is the average of housing and equity, measured as the expected return minus the real ex ante safe rate. Dashed lines show the country-specific linear trends. All data are in percent. Data for Canada are for equities only.

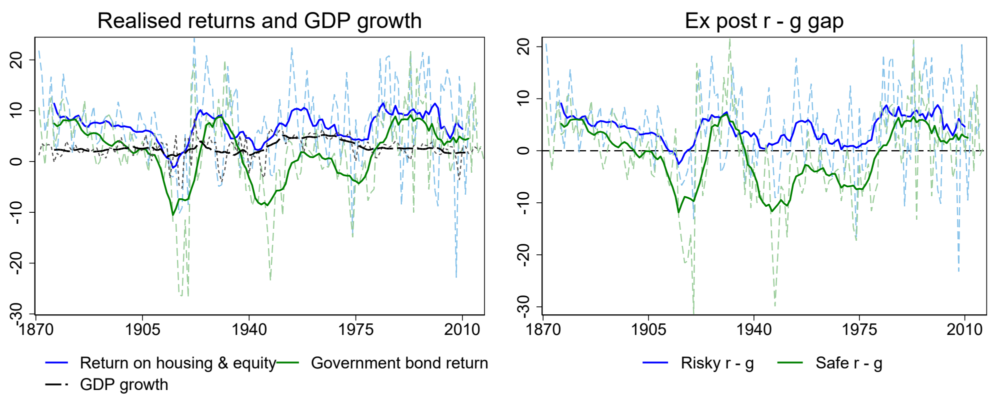
2.A.3 Drivers of expected returns: additional details

Table 2.A.3. Co-movement of realised returns, safe rates and risk premia

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Realised real return on risky assets					
	Baseline	Year effects	Equity	Housing	3-year MA	10-year MA
Safe rates	0.38***	0.38***	0.55***	0.17***	0.43***	0.39***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
Country fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects		✓				
R^2	0.11	0.47	0.10	0.04	0.18	0.30
Observations	1890	1890	2275	1818	1832	1631
	Dependent variable: Ex post risk premia on risky assets					
	Baseline	Year effects	Equity	Housing	3-year MA	10-year MA
Safe rates	-0.57***	-0.54***	-0.34***	-0.81***	-0.52***	-0.55***
	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
Country fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects		✓				
R^2	0.20	0.51	0.04	0.45	0.23	0.42
Observations	1890	1890	2274	1818	1832	1631

Notes: Regressions of realised risky returns and risk premia on the safe return. Realised returns are the sum of capital gain and yield, averaged across equities and housing and net of inflation. Safe return is the real total government bond return. Baseline specification in column 1 has country fixed effects only, column 2 adds year fixed effects, columns 3 and 4 consider equity and housing separately and columns 5 and 6 use 3-year and 10-year moving averages of both risky and safe returns. Standard errors in parentheses are clustered by country and year. *: $p < 0.1$ **: $p < 0.05$ ***: $p < 0.01$.

Figure 2.A.8. Realised returns and GDP growth



Notes: Unweighted averages of 17 countries. Dashed lines are annual data and solid lines are centered 10-year moving averages. Realised risky return is the average of total real holding period returns on equity and housing, realised safe return is the real return on long-term government bonds.

References

Abel, Andrew B., Gregory N. Mankiw, Lawrence H. Summers, and Richard J. Zeckhauser. 1989. "Assessing Dynamic Efficiency: Theory and Evidence." *Review of Economic Studies* 56 (1): 1–19.

Baele, Lieven, Geert Bekaert, and Koen Inghelbrecht. 2010. "The Determinants of Stock and Bond Return Comovements." *Review of Financial Studies* 23 (6): 2374–2428.

Baldwin, Richard, and Coen Teulings. 2014. *Secular Stagnation: Facts, Causes and Cures*. London: Centre for Economic Policy Research.

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.

Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75 (5): 2421–2463.

Barro, Robert J. 2020. "r minus g." NBER Working Paper 28002.

Benhabib, Jess, and Alberto Bisin. 2018. "Skewed Wealth Distributions: Theory and Empirics." *Journal of Economic Literature* 56 (4): 1261–1291.

Bianchi, Francesco, Martin Lettau, and Sydney C. Ludvigson. 2016. "Monetary Policy and Asset Valuation." NBER Working Paper 22572.

Blanchard, Olivier J. 1993. "Movements in the Equity Premium." *Brookings Papers on Economic Activity* 1993 (2): 75–138.

Blanchard, Olivier J. 2019. "Public Debt and Low Interest Rates." *American Economic Review* 109 (4): 1197–1229.

Caballero, Ricardo J., and Emmanuel Farhi. 2014. "On the Role of Safe Asset Shortages in Secular Stagnation." In *Secular Stagnation: Facts, Causes and Cures*. Edited by Coen Teulings and Richard Baldwin. London: CEPR Press. Chapter 9, 111–122.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2017a. "Rents, Technical Change, and Risk Premia Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares." *American Economic Review: Papers & Proceedings* 107 (5): 614–620.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2017b. "The Safe Assets Shortage Conundrum." *Journal of Economic Perspectives* 31 (3): 29–46.

Caballero, Ricardo J., and Alp Simsek. 2020. "A Risk-centric Model of Demand Recessions and Macroprudential Policy." *Quarterly Journal of Economics* 135 (3): 1493–1566.

Campbell, John Y., Carolin Pflueger, and Luis M. Viceira. 2020. "Macroeconomic Drivers of Bond and Equity Risks." *Journal of Political Economy* 128 (8): 3148–3185.

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.

Campbell, John Y., Adi Sunderam, and Luis M. Viceira. 2017. "Inflation Bets or Deflation Hedges? The Changing Risks of Nominal Bonds." *Critical Finance Review* 6: 263–301.

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.

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. 2009. *Asset Pricing*. Princeton, N.J.: Princeton University Press.

David, Alexander, and Pietro Veronesi. 2013. "What Ties Return Volatilities to Price Valuations and Fundamentals?" *Journal of Political Economy* 121 (4): 682–746.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2019. "Global Trends in Interest Rates." *Journal of International Economics* 118: 248–262.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez. 2017. "The Price of Variance Risk." *Journal of Financial Economics* 123 (2): 225–250.

Diamond, Peter A. 1965. "National Debt in a Neoclassical Growth Model." *American Economic Review* 55 (5): 1126–1150.

Elton, Edwin J. 1999. "Presidential Address: Expected Return, Realized Return, and Asset Pricing Tests." *Journal of Finance* 54 (4): 1199–1220.

Fama, Eugene F., and Kenneth R. French. 2002. "The Equity Premium." *Journal of Finance* 57 (2): 637–659.

Farhi, Emmanuel, and François Gourio. 2018. "Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia." *Brookings Papers on Economic Activity* 2018 (2): 147–250.

Fernald, John G. 2015. "Productivity and Potential Output before, during, and after the Great Recession." *NBER Macroeconomics Annual* 29: 1–51.

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. 2021. "Accounting for Wealth-Inequality Dynamics: Methods, Estimates, and Simulations for France." *Journal of the European Economic Association* 19 (1): 620–663.

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.

Golez, Benjamin, and Peter Koudijs. 2017. "Four Centuries of Return Predictability." NBER Working Paper 20814.

Golez, Benjamin, and Peter Koudijs. 2018. "Four Centuries of Return Predictability." *Journal of Financial Economics* 127 (2): 248–263.

Gomme, Paul, B Ravikumar, and Peter Rupert. 2015. "Secular Stagnation and Returns on Capital." *St Louis Federal Reserve Economic Synopses* 19: 1–3.

Gordon, Myron J. 1962. *The Investment, Financing, and Valuation of the Corporation*. Homewood, Ill.: RD Irwin.

Greenwald, Daniel L., Martin Lettau, and Sydney C. Ludvigson. 2019. "How the Wealth Was Won: Factors Shares as Market Fundamentals." NBER Working Paper 25769.

Grullon, Gustavo, and Roni Michaely. 2002. "Dividends, Share Repurchases, and the Substitution Hypothesis." *Journal of Finance* 57 (4): 1649–1684.

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.

Iachan, Felipe S., Plamen T. Nenov, and Alp Simsek. 2020. "The Choice Channel of Financial Innovation." *American Economic Journal: Macroeconomics*, Forthcoming.

Jagannathan, Ravi, Ellen R. McGrattan, and Anna Scherbina. 2000. "The Declining US Equity Premium." *Federal Reserve Bank of Minneapolis Quarterly Review* 24 (4): 3–19.

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* 134 (3): 1225–1298.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual* 31 (1): 213–263.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.

Kopecky, Joseph, and Alan M. Taylor. 2020. "The Murder-Suicide of the Rentier: Population Aging and the Risk Premium." NBER Working Paper 26943.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2012. "The Aggregate Demand for Treasury Debt." *Journal of Political Economy* 120 (2): 233–267.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2020. "Income and Wealth Inequality in America, 1949–2016." *Journal of Political Economy* 128 (9): 3469–3519.

Kuvshinov, Dmitry. 2020. "The Co-Movement Puzzle." Working Paper.

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2020. "The Big Bang: Stock Market Capitalization in the Long Run." CEPR Discussion Paper 14468.

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.

Martıúnez-Toledano, Clara. 2020. "House Price Cycles, Wealth Inequality and Portfolio Reshuffling." Working Paper.

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

Monnet, Eric, and Mr Damien Puy. 2019. "One Ring to Rule them All? New Evidence on World Cycles." IMF Working Paper 19/202.

Muir, Tyler. 2017. "Financial Crises and Risk Premia." *Quarterly Journal of Economics* 132 (2): 765–809.

Pflueger, Carolin, Emil Siriwardane, and Adi Sunderam. 2020. "Financial Market Risk Perceptions and the Macroeconomy." *Quarterly Journal of Economics* 135 (3): 1443–1491.

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.

Ramsey, F. P. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38 (152): 543–559.

Rognlie, Matthew. 2015. "Deciphering the Fall and Rise in the Net Capital Share." *Brookings Papers on Economic Activity* 46 (1): 1–69.

Schmelzing, Paul. 2020. "Eight Centuries of Global Real Interest Rates, R-G, and the 'Suprasecular' Decline, 1311–2018." Bank of England Staff Working Paper 845.

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): 1029–1061.

Shiller, Robert J. 2015. *Irrational Exuberance*. Princeton, N.J.: Princeton University Press.

Shiller, Robert J., and Andrea E. Beltratti. 1992. "Stock Prices and Bond Yields: Can their Comovements be Explained in Terms of Present Value Models?" *Journal of Monetary Economics* 30 (1): 25–46.

Song, Dongho. 2017. "Bond Market Exposures to Macroeconomic and Monetary Policy Risks." *Review of Financial Studies* 30 (8): 2761–2817.

Summers, Lawrence H., and L. Rachel. 2019. "On Falling Neutral Real Rates, Fiscal Policy and the Risk of Secular Stagnation." *Brookings Papers on Economic Activity* 1: 1–66.

Chapter 3

The Profit-Credit Cycle*

Joint with Björn Richter

3.1 Introduction

The credit cycle takes center stage in the scholarly analysis of the 2007/2008 crisis. The financial turmoil was preceded by a boom in private credit in many countries, just as so many other crises episodes before (Schularick and Taylor, 2012). More generally, the credit cycle also predicts medium-term output growth, but economic forecasters often fail to account for this relationship (Mian, Sufi, and Verner, 2017). Asset return data suggest they are not alone: capital markets often neglect the treacherous link between credit expansions and downside risk (Baron and Xiong, 2017; Fahlenbrach, Prilmeier, and Stulz, 2017; Krishnamurthy and Muir, 2017). In response to the output risks associated with credit expansions, policy-makers today monitor credit aggregates closely and apply a widening range of macro-prudential tools, once they detect overheating. While these policies are often effective in dampening credit growth (Akinci and Olmstead-Rumsey, 2018), they are rather a treatment of symptoms than causes. This is no surprise as the understanding of the ultimate sources of credit supply expansions and how they turn into a crisis is still limited (Mian and Sufi, 2018).

* This work is part of a larger project kindly supported by a research grant from the Bundesministerium für Bildung und Forschung (BMBF). Special thanks to participants at the Frontiers in Financial History Workshop 2018 in Rotterdam, the EDP Jamboree 2018 in Florence, the CESifo Workshop on Banking and Institutions 2019 in Munich, and the 19th FDIC-JFSR Fall Banking Research Conference, as well as seminar participants at the Banque de France, Board of Governors of the Federal Reserve, BI Norwegian Business School, Bundesbank, Erasmus School of Economics Rotterdam, ESADE, IESE, Norges Bank, Sveriges Riksbank, Universitat Pompeu Fabra, University of Bonn, University of Luxembourg and University of Mainz. We are grateful to Christian Bayer, Pedro Bordalo, Christian Eufinger, Roger Farmer, Andrej Gill, Florian Hett, Dmitry Kuvshinov, Yueran Ma, Enrico Perotti, José-Luis Peydró, Moritz Schularick, Kasper Roszbach, Daniele Verdini, Gian-Luca Violante for helpful comments and to Matthew Baron for sharing data.

In this paper, we revisit the origins and turning points of the credit cycle. It is well documented that firms and managers overpredict future earnings when profits are high and that this has consequences for investment (Greenwood and Hanson, 2015; Gennaioli, Ma, and Shleifer, 2016). We show that there is a similar pattern underlying bank lending. What we observe in the data is in fact a "profit-credit cycle". An increase in bank profitability predicts a credit expansion in the next years, but also elevated crisis risk down the road. Banking panics often occur after profits start declining during such boom episodes. These findings connect well with older ideas of "displacements" in the credit market triggering waves of optimism followed by a "Minsky moment" (Minsky, 1977; Kindleberger, 1978) and mesh nicely with new modeling approaches to the credit cycle based on extrapolative expectations (Bordalo, Gennaioli, and Shleifer, 2018; Greenwood, Hanson, and Jin, 2019).

To study the profit-credit relationship, we introduce a new dataset on bank profitability in 17 advanced economies over the last 145 years. The data come from banking sector profit and loss accounts and allow us to systematically assess the relationship between bank profits, the credit cycle, and financial instability in modern financial history. One advantage of our accounting data is that profits are, by definition, backward looking and enable us to study the link between realized outcomes and subsequent lending. Credit spreads and stock returns on the other hand, which have been studied in the previous literature (Baron and Xiong, 2017; Krishnamurthy and Muir, 2017; Baron, Verner, and Xiong, 2020) do not necessarily reflect realized outcomes and incorporate expectations about the future. Furthermore, our accounting data cover a large share of the overall banking system, while equity and bond prices are often only available for a small subset of the banking sector. For a subsample of countries and episodes we were furthermore able to decompose bottom line bank profitability into its sources – revenue, costs and loan losses – and its uses – funds paid out to shareholders and funds retained as equity in the balance sheet. Our new dataset is complemented by the data of the Macrohistory Database (Jordà, Schularick, and Taylor, 2017), which provides us with credit aggregates, a chronology of banking crises and a large number of control variables for our investigation.

We find that bank profitability leads the credit cycle. High bank profits are followed by credit expansions. We measure profitability as return on equity (*RoE*) or return on assets (*RoA*) and proxy a sequence of increasing or decreasing profits with the three-year change in this return measure ($\Delta_3 RoE$ and $\Delta_3 RoA$). Our results imply that a one standard deviation higher $\Delta_3 RoE$ predicts a 0.2 standard deviation higher change in credit-to-GDP over the subsequent three years. This is equivalent to an increase of the credit to GDP ratio by 4.9% instead of the sample average of 3.4% over a three year window. This relationship remains robust when we include additional controls, time effects and analyze subsamples. It holds for alternative measures of profitability or credit growth, during and outside of financial crises and on a country-by-country level. We show in the appendix that these results also carry over to the bank level using panel data from Federal Reserve call reports.

Policy-makers may be predominantly concerned about large credit expansions as these are often associated with the left tail of macroeconomic growth outcomes. To study these boom episodes separately, we rely on an indicator variable for the start of a large credit boom, defined by the three-year change in credit-to-GDP being elevated by more than one country-specific standard deviation. In line with our previous results, we find that the start of a credit boom episode can be forecasted with increasing profitability. During the ensuing boom *RoE* and *RoA* return to their mean values within a few years. The ratio of total bank profits to GDP remains elevated throughout the boom due to increasing quantities of intermediated funds. This evidence suggests that there would be ample room for the banking sector to increase capital through retained earnings during credit booms to shield the economy from harm during the bust (Jordà, Richter, Schularick, and Taylor, 2021).

Which mechanisms can explain the strong association between profits and the credit cycle? We first show that the relationship cannot be explained by credit demand. An outward shift in credit demand should be associated with high interest rate spreads during the credit expansion. We find the opposite: the price of credit – a corporate bond spread – is negatively associated with recent improvements in profitability. Focusing on expansions in credit supply, we distinguish between financial constraints and time-varying beliefs as explanations for the profit-credit cycle. High profits, if not paid out completely to shareholders, increase net worth in the banking sector and thereby relax borrowing constraints (Bernanke and Gertler, 1989; Holmstrom and Tirole, 1997; Kiyotaki and Moore, 1997). In our long-run data, we find evidence consistent with such a net worth channel. Bank capital ratios and retained earnings, as measures of the level and change in bank net worth, predict credit expansions.

However, we provide several findings that cast doubt on whether the "profit-credit cycle" can be explained exclusively by financial constraints. First, the relationship between profits and future lending growth remains stable and significant when we introduce direct controls for net worth. We find the same strong link between profits and credit growth in specifications that include the capital ratio and changes in banking sector capital. Second, when decomposing bank profitability into loan losses, revenues and costs, we find that decreasing loan losses are associated with expanding credit, while lower costs or increasing revenues are not. If the relationship between profitability and credit expansion would simply be due to the effect of profits on net worth, we would expect that the source of profits is largely irrelevant and results for all sources should be similar.

Third, we rely on the idea that dividends paid to shareholders do not relax financial constraints in the banking sector. Decomposing profits into dividends and retained earnings, we find a significant effect of dividend payments on future credit expansion, while controlling for retained earnings. Finally, when we include the current level and recent changes in profitability in one specification, the coefficients for both variables are positive and significant. This suggests that not only additional net worth (measured by *RoE* levels), but also the change of *RoE* relative to previous periods matters for credit expansion. These findings are consistent with expectations-based credit cycle models (Bordalo, Gennaioli, and Shleifer, 2018; Greenwood, Hanson, and Jin, 2019). In these models, positive news – displacements in the language of Minsky (1977) – are extrapolated into the future and thereby trigger a wave of optimism. It is during these episodes that investors willingly supply credit, to be systematically disappointed in the following years.

To study the expectation formation process in further detail, we use data from a survey among bank CFOs in the United States. We find that recent changes in profitability are strongly associated with measures of optimism and expected profits. The link between realized and past profitability is weaker, and as a result, bank CFOs make predictable forecast errors. When current profits are high, bank CFOs are optimistic, but realized future earnings are lower than expected. We then show that survey expectations and optimism are reflected in the aggregate credit cycle. Higher optimism today is associated with higher lending volumes over the next 12 months. This creates a link between forecast errors and lending, implying that extrapolation could be associated with a misallocation of credit.

Motivated by the Minsky (1977)-narrative, we study the link between bank profitability and the incidence of banking crises in the second part of the paper. We find that increases in profitability predict financial instability over the medium term. A one standard deviation higher increase in profitability between years $t-3$ and t is associated with a one percentage point higher crisis likelihood in $t+3$. This corresponds to more than a 25 percent increase relative to a baseline crisis frequency of 3.1%. Looking at the transition from a boom into a crisis, we find that crises are associated with a decline in the growth of profitability shortly before their onset. Unsurprisingly, profitability then drops further in the year of a banking crisis and remains low for several years thereafter.

Using data on crisis characteristics collected by Baron, Verner, and Xiong (2020), we then assess whether this pattern is specific to certain types of banking crises in order to gain insights on the underlying mechanisms. We find that the reversal of fundamentals – improving profitability followed by a decline shortly before the crisis – is a peculiarity of banking panics. While panics are predicted by increasing profitability, non-panic banking crises are preceded by declining profitability already three years prior. For a panic to occur, creditors' expectations about bank asset returns must be extremely negative, such that bank capital would not be able to absorb losses and creditors' claims are at stake. Our evidence hence suggests that expectations of creditors turn extremely negative when they are surprised by a sudden reversal in bank fundamentals following a boom. Panics then occur as a sudden end to a boom, while non-panic crises occur after prolonged periods of weak fundamentals.

Taken together, the credit-cycle patterns we document are consistent with the model of Bordalo, Gennaioli, and Shleifer (2018) where good news create overoptimism, and subsequently incoming disappointing news lead to sharp reversals in expectations and banking panics.[1] The relationship between bank profitability, especially loan losses, and crisis also lines up well with previous studies on the behavior of market-based risk metrics before crises: credit spreads (Krishnamurthy and Muir, 2017) and stock market volatility (Danielsson, Valenzuela, and Zer, 2018) have both been found to be low in the prelude to a crisis. In a similar vein, Meiselman, Nagel, and Purnanandam (2018) show that elevated bank profits are measuring risk in the cross-section of banks. High *RoE* levels during the boom are linked to a worse performance of banks during the crisis.

Our paper contributes to three strands of research. One strand discusses patterns of the credit cycle (Aikman, Haldane, and Nelson, 2015; Dell'Ariccia, Igan, Laeven, and Tong, 2016) and identifies markers that help to tell different kinds of credit booms apart (Gorton and Ordonez, 2020; Kirti, 2020; Richter, Schularick, and Wachtel, 2020). A rapidly growing literature surveyed in Mian and Sufi (2018) studies the interplay of credit and business cycles with a focus on credit supply based explanations. Our results support the view that credit supply plays an important role in shaping the credit cycle and shows that credit booms start when banking sector profitability has been increasing.

Second, our paper extends the behavioral credit cycle literature. Evidence for overextrapolation of recent shocks or trends is pervasive. Greenwood and Shleifer (2014) show that survey-based investor expectations are extrapolative and hard to reconcile with rational expectations models. Similar results have been obtained analyzing macroeconomic expectations of professional forecasters (Bordalo, Gennaioli, Ma, and Shleifer, 2020), households' house price expectations (Kuchler and Zafar, 2019; De Stefani, 2020) and expectations in laboratory experiments (Landier, Ma, and Thesmar, 2018). Recent research relates the extrapolation bias to fluctuations in real investment (Gennaioli, Ma, and Shleifer, 2016) and incorporates extrapolative biases in models of the credit cycle (Bordalo, Gennaioli, and Shleifer, 2018; Greenwood, Hanson, and Jin, 2019).

It is important to note, that our data on bank profitability allow us to show that such a relationship holds for the bank credit cycle, while most previous studies focused on cyclical developments in the bond market (Greenwood and Hanson, 2013), or linked expansions in bank credit with data on prices and defaults from the bond market (Krishnamurthy and Muir, 2017; Greenwood, Hanson, and Jin, 2019). Linking bank profitability to bank credit is important, as the underlying theory of extrapolation most likely applies within a specific asset-class. Kuvshinov (2018) shows that measures of asset market sentiment are not necessarily correlated across asset classes, so that extrapolation seems to be domain-specific.

Third, our paper is related to a literature that studies the relationship between net worth and credit intermediation in models with financial frictions (Bernanke

1. Note that changes in expectations are based on fundamentals, consistent with models of fundamentals-based bank runs (Goldstein and Pauzner, 2005; He and Xiong, 2012).

and Gertler, 1989; Holmstrom and Tirole, 1997; Kiyotaki and Moore, 1997). A vast literature builds on these early contributions, studying alternative frictions and amplification mechanisms and integrating the mechanisms into richer macroeconomic models (e.g. Brunnermeier and Sannikov, 2014). The profit-credit cycle is consistent with these mechanisms, but our results suggest that this channel alone cannot account for it. In that regard, recent attempts to integrate belief-driven cycles into models that feature amplification through intermediaries seem in the light of our results particularly promising (Bordalo, Gennaioli, Shleifer, and Terry, 2019; Kaplan, Mitman, and Violante, 2020; Krishnamurthy and Li, 2020).

3.2 A new dataset on bank profitability

This paper is built around a novel long-run dataset on bank profitability across countries and time. We construct new return on equity and return on asset series for 17 countries from 1870 to today using banking sector income statements. So far, research with long-run historical data on credit cycles and systemic banking crises heavily relied on banking sector balance sheet information (Schularick and Taylor, 2012; Jordà, Richter, Schularick, and Taylor, 2021). A second strand of the literature recently started to incorporate market prices for debt and equity into the analysis (Baron and Xiong, 2017; Krishnamurthy and Muir, 2017; Baron, Verner, and Xiong, 2020). Banking sector income – in particular realized banking sector profitability – has been largely ignored. Adding data from the income statement creates a link between balance sheet data and market prices. The new dataset therefore complements these existing data. Our main profitability series – return on equity (*RoE*) – is computed by dividing total profits of the banking system by book equity:

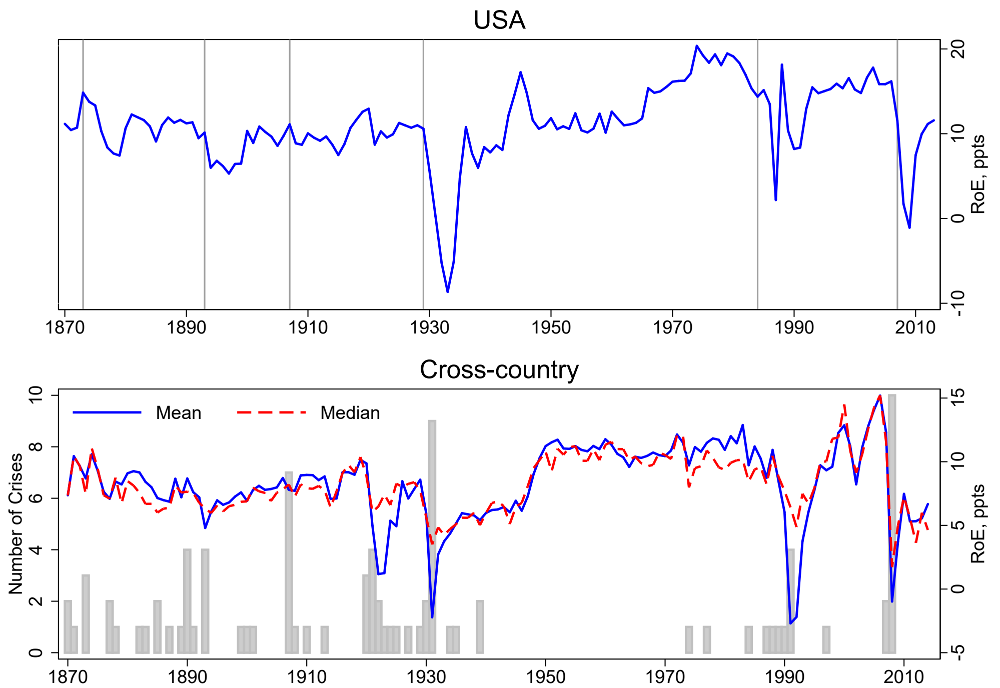
$$RoE = \frac{Net\ profits\ after\ Tax}{Book\ Equity} \tag{3.1}$$

The numerator of the equation measures accounting income of the banking system after the deduction of all relevant expenditures and corporate taxes. The denominator includes paid-in capital, reserves and retained earnings. The equity items also include profits carried forward and the issuance premium gained by selling stocks above their nominal value. Aside from the return on equity series, we also construct a return on asset series by dividing profits by total assets instead of total equity.[2]

The data come from a wide range of sources including publications of the OECD, central banks, banking supervisory institutions, work of banking historians and individual bank reports. The new series includes on average more than 125 years of

2. Return on equity and return on assets are connected through the leverage ratio of the underlying financial institutions. Due to sampling and coverage differences, the implicit leverage ratio of the return on equity and return on asset series in some cases differs slightly from the leverage ratio of Jordà, Richter, Schularick, and Taylor (2021).

Figure 3.1. Long-run evolution of RoE in the United States and across sample countries



Notes: This figure displays the evolution of *RoE* in % between 1870 and today for the USA and for a cross-country mean (median). Vertical bars indicate starting years of systemic financial crises in the USA and the number of countries experiencing the start of a financial crisis respectively (see appendix for dates).

data for each country in our sample. The paper is complemented by a detailed Internet Appendix describing sources and data construction. Summary statistics of the profitability measures can be found in Table 3.A.1.[3]

Figure 3.1 illustrates the data. It shows the *RoE* series for the United States and yearly sample averages. The vertical lines in the upper graph indicate systemic banking crises in the US and grey bars in the cross-country graph indicate the number of countries with systemic banking crises in a given year. We rely on the narrative chronology by Jordà, Schularick, and Taylor (2017) to identify systemic banking crises events. Several features stand out: Bank profitability, measured by *RoE*, was relatively stable over the last 145 years. *RoE* fluctuated around 8 percent in most

3. A large share of the dataset is based on aggregate banking statistics. In some countries, we gathered data of the largest commercial banks to extend the data back into the 19th century. Relying on data of a few banks might generate excess volatility compared to the aggregate banking sector statistics and add bank idiosyncrasies to the final series. However, in most cases the deviations are likely small, as the respective banking systems were dominated by a small number of banks (e.g. Canada) with a large market share. Another issue is related to the use of accounting data. We treat this data at face value. The sophistication of accounting standards and practice however varied significantly historically. As a consequence, the data might be distorted by profit smoothing and hidden reserves in bank balance sheets. In our empirical analysis, we will therefore focus especially on changes in profitability and the resulting estimates are most likely downward biased by profit or dividend smoothing.

countries (see also the summary statistics in Table 3.A.1). In some countries – such as the United States – there is a gradual upward trend in *RoE* in the second half of the 20th century. Major deviations from the trend follow or coincide with systemic banking crises. These crises often drive bank profitability into negative territory. For example, the *RoE* series for the United States shows three major negative shocks with *RoE* around or below zero: the Great Depression, the S&L crisis and the Global Financial Crisis. The defining feature of the cross-country averages are the extraordinarily low profits during clustered crisis events. Comparing profitability in crisis and non-crisis episodes reveals that *RoE* in a crisis-year is around 7% lower than the non-crisis average. However, not all systemic banking crises are characterized by pronounced negative profitability. While some crises nearly wiped out the entire banking sector capital, others are difficult to eyeball in the profitability series (e.g. the crisis of 1907 in the United States).

Our new dataset also allows us to decompose banking sector profits into sources and uses. Drawing from additional banking sector accounting information, we separate *RoE* by the use of funds into a dividend and a retained earnings component. We gathered data on dividends directly and back out retained earnings as the difference between profits and dividends. Furthermore, we were able to obtain information on the sources of bank profitability. We decompose profits into revenues (net interest plus net fee income), operating costs and loan losses. We then compute 3-year changes in all profitability variables (e.g. *RoE*) as a proxy for medium-term changes

$$\Delta_3 RoE_{i,t} = RoE_{i,t} - RoE_{i,t-3}. \tag{3.2}$$

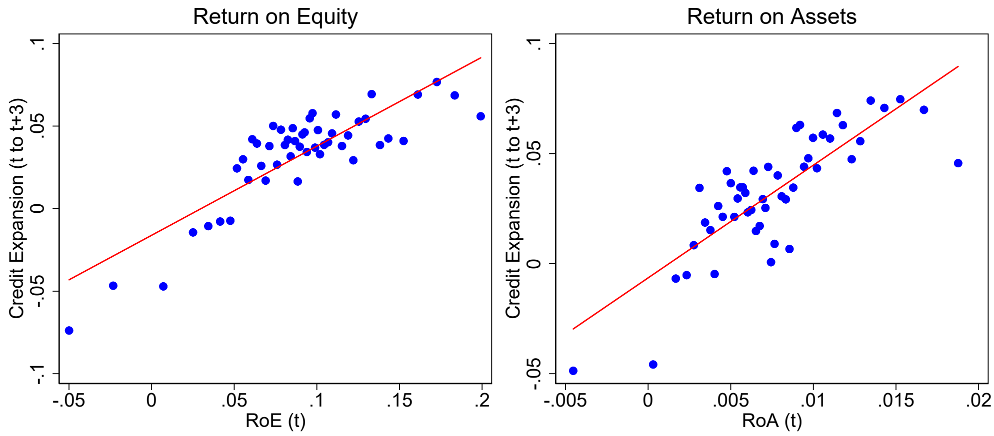
As there are no clear trends in *RoE* over our sample period, this variable is on average close to zero. The bank profitability data is in some countries dominated by extreme loss events during crises (see Figure 3.B.1). We therefore winsorize all profitability measures at the 2.5% level to ensure that empirical results will not be driven by extreme outliers in profitability. The main results of the paper also hold in the raw data with the same significance level and similar point estimates. Our main dependent variable to analyze the relationship between profitability and credit cycles will be the change in the credit-to-GDP ratio over a three-year interval (similar to Mian, Sufi, and Verner, 2017):

$$\Delta_3 y_{i,t+3} = (Credit/GDP)_{i,t+3} - (Credit/GDP)_{i,t} \tag{3.3}$$

Credit here refers to bank credit extended to the domestic private non-financial sector. It includes loans to households as well as loans to non-financial firms. In contrast to profitability measures, there has been an upward trend in the ratio of credit to GDP over the past 150 years and $\Delta_3 y_{i,t}$ is around 3.4% on average.

3.3 Bank profitability and the credit cycle

Figure 3.2. Binned scatterplot for the relationship between profitability and credit-to-GDP changes



Notes: The figure links bank profitability and subsequent three-year changes in credit to GDP. Observations are collapsed into 50 equal sized bins according to profitability (RoE or RoA). Each point represents the group specific means of profitability and credit expansion after controlling for country fixed effects and a time trend. Fitted regression lines illustrate the correlation between bank profitability and subsequent credit expansion.

What is the relationship between bank profitability and credit growth? This section establishes that increasing bank profitability is a significant and robust predictor of subsequent credit expansions.

Figure 3.2 graphically illustrates the relationship between current *RoE* or *RoA* and credit expansion over the following years. In both panels, the data are collapsed into 50 equal-sized bins according to profitability measure and the graph displays the mean profitability for observations in each of these bins. In addition, on the y-axis, the mean of three-year credit-to-GDP changes for each of the 50 groups is presented. The graph shows the relationship of residuals after controlling for country fixed effects and including a time trend to account for the long-term decline in *RoA*. Both panels display a strong positive correlation between profitability and credit expansion.

We will now assess the relationship between profits and credit more formally. Since there is a strong time trend in *RoA*, we will focus on the two previously defined measures of medium-term variation in profitability, $\Delta_3 RoE$ and $\Delta_3 RoA$. We assess their relationship with the credit cycle using three year changes in the credit-to-GDP ratio as the dependent variable. Similar to the approach in Mian, Sufi, and Verner (2017) we estimate variants of equation

$$\Delta_3 y_{i,t+3} = \alpha_i + \beta^{\Delta RoE} \Delta_3 RoE_{i,t} + \sum_{\tau=0}^{2} \gamma_\tau \Delta y_{i,t-\tau} + \eta X_{i,t} + \theta Z_{i,t} + u_{i,t+3}, \tag{3.4}$$

where we include changes in profitability, $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$ as well as a distributed lag of the dependent variable ($\sum_{\tau=0}^{2} \Delta y_{i,t-\tau}$). $X_{i,t}$ is a vector of macro-controls including the three most recent values of real GDP growth, short-term interest rates, long-term interest rates, inflation, the current account-to-GDP ratio as well as log real GDP per capita to account for the state of development in our long run sample. As a second set of controls ($Z_{i,t}$), and as a first step towards disentangling possible channels, we add two proxies that account for financial constraints in the banking sector: the capital ratio of the banking sector as a measure of leverage constraints and three-year changes in bank capital relative to GDP as a measure of net worth in the banking sector (Adrian, Moench, and Shin, 2019). We exclude data from the two world wars to avoid measuring the effects of wartime government intervention in the banking sector.

Table 3.1 column (1) shows that an increase in profitability over the past three years ($\Delta_3 RoE_{i,t}$) predicts significantly higher credit expansion over the following three years ($\Delta_3 y_{i,t+3}$). A similar result emerges when we include changes in RoA, $\Delta_3 RoA_{i,t}$, in column (4). Banks extend more credit when measures of realized profitability start looking better over time. Adding macroeconomic controls in (2) and (5) reduces the coefficients slightly, but the results remain highly significant. Consistent with a role of intermediary leverage, we find that a high capital ratio is associated with increases in the credit-to-GDP ratio over the following years – relaxed funding constraints are associated with increased lending. Increases in the ratio of capital to GDP as a measure of aggregate net worth do not predict credit expansion. Importantly, including both measure of financial constraints does not affect the results for the profitability measures.

How sizable is the effect of profits on lending? Increasing $\Delta_3 RoE_{i,t}$ ($\Delta_3 RoA_{i,t}$) by one standard deviation is associated with a 1.53% (1.48%) higher increase in credit-to-GDP over a three-year window. The mean of three-year changes in credit-to-GDP in our non-war estimation sample is 3.4%. Our estimates hence imply that this rate of change increases by almost 50 percent when realized profitability growth is elevated by one standard deviation. Our long run evidence shows that booms in profitability are an important vector to understand credit expansions.

3.3.1 Robustness

In the following subsections, we show that this relationship is a robust feature of the data, the relationship is mainly driven by loan losses, and that profitability also helps to predict large credit booms. We first discuss the robustness of the relationship between bank profitability and credit expansion identified before.

Table 3.1. Multivariate models for changes in credit-to-GDP, baseline specification

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.42***	0.34***	0.33***			
	(0.06)	(0.04)	(0.04)			
$\Delta_3 RoA_{i,t}$				4.68***	3.88***	3.89***
				(0.85)	(0.69)	(0.69)
$Capital\ Ratio_{i,t}$			0.23***			0.23***
			(0.07)			(0.07)
$\Delta_3(Capital/GDP)_{i,t}$			-0.06			-0.17
			(0.24)			(0.23)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
R^2	0.12	0.19	0.20	0.11	0.19	0.20
Observations	1636	1492	1491	1642	1498	1491

Notes: This table reports regressions of credit-to-GDP changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$. All specifications control for three lags of credit-to-GDP changes. Columns (2), (3), (5) and (6) add a vector of macroeconomic control variables described in the text. Columns (3) and (6) additionally control for financial constraint proxies. All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Subsamples:. In Table 3.2 we look at subsamples of the data. All specifications include the full set of control variables. In a first step we restrict the sample to the post Bretton-Woods era to understand whether the strong relationship can also be observed in the current international monetary framework. We find that the results are robust to restricting the analysis to this time period. The same is true in a subsample of pre-2000 data, which we analyse to ensure that the relationship was not only a feature of the credit cycle that found a sudden end in the 2007/2008 crisis. In column (3), we use non-overlapping windows of observations in the dependent variable to deal with autocorrelation introduced through overlapping data and results remain highly significant. In column (4), we address possible cross-country correlation of variables and include year-fixed effects. The year fixed effects increase the R^2 to more than 0.3 in both cases, indicating that there is a high degree of cross-country correlation in credit expansion, as identified in other studies (Rey, 2016; Jordà, Schularick, Taylor, and Ward, 2019). The coefficients on profitability measures remain however highly significant.

Finally, we include in (5) a dummy for a banking crises occurring in the last three years (i.e. between $t-2$ and t) and its interaction with profitability measures. This exercise is addressing two concerns. First, the relationship could be entirely driven by low credit growth following high losses in a banking crisis. As the first

Table 3.2. Multivariate models for changes in credit-to-GDP, subsamples and time effects

	Dependent variable: $\Delta_3 y_{i,t+3}$				
	(1) Post-1973	(2) Pre-2000	(3) No overlap	(4) Year effects	(5) Crisis
$\Delta_3 RoE_{i,t}$	0.25***	0.27***	0.29***	0.20***	0.21***
	(0.05)	(0.07)	(0.05)	(0.04)	(0.03)
$Crisis_{[t-2,t]}$					-0.04**
					(0.02)
$Crisis_{[t-2,t]} \times \Delta_3 RoE_{i,t}$					0.27
					(0.20)
Country fixed effects	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓
Year effects				✓	
R^2	0.35	0.17	0.18	0.36	0.23
Observations	640	1304	496	1491	1483
	Dependent variable: $\Delta_3 y_{i,t+3}$				
	(1) Post-1973	(2) Pre-2000	(3) No overlap	(4) Year effects	(5) Crisis
$\Delta_3 RoA_{i,t}$	4.15***	3.06***	3.55***	2.61***	2.42***
	(0.93)	(0.95)	(0.59)	(0.70)	(0.52)
$Crisis_{[t-2,t]}$					-0.04**
					(0.02)
$Crisis_{[t-2,t]} \times \Delta_3 RoA_{i,t}$					1.68
					(2.38)
Country fixed effects	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓
Year effects				✓	
R^2	0.35	0.17	0.18	0.36	0.22
Observations	640	1304	496	1491	1491

Notes: This table reports regressions of credit-to-GDP changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ *and*$\Delta_3 RoA_{i,t}$. All specifications control for three lags of credit-to-GDP changes and a vector of financial constraint proxies and macroeconomic control variables (see text in section 3.3). Column (1) uses only post-1973 data. Column (2) uses only pre-2000 data. Column (3) restricts the data to non-overlapping observations only. Column (4) includes year-fixed effects. Column (5) includes a banking crisis dummy and its interaction with profitability measures. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

coefficient in column (5) shows, this is not the case. Profitability changes remain a robust predictor of credit expansion even when we account for crisis events. The second row shows that credit expansion is significantly dampened following crisis episodes. The second issue we seek to address is whether the relationship between profitability and credit expansion is stronger during a crisis episode. If the relationship was primarily driven by financial constraints in the banking sector, we would expect that the effect is stronger in crisis periods when these constraints are more likely to be binding. In both specifications, the interaction is positive, but insignificant. Another way to address this question is to focus on the relationship in a crisis only. In Table 3.A.2 we include only the three years after the start of systemic banking crises into the regression sample. We find coefficients similar to our baseline estimates: large losses during crisis episodes translate into significantly lower credit expansion, but they do so in the same way as profits do in non-crisis episodes.

Alternative credit measures:. The appendix presents further robustness tests with respect to variable definitions. In a first step, we vary the dependent variable. So far, $\Delta y_{i,t+3}$ referred to the three-year change in the credit-to-GDP ratio. In Table 3.A.3 we replace credit-to-GDP with logged real private credit per capita to rule out the possibility that the effect is driven by the denominator. The results are in line with our previous findings. In Table 3.A.4 we move away from credit variables and look at the bank-assets-to-GDP ratio. The findings are similar to those for credit variables. In Table 3.A.5 we ask whether the relationship is similar for non-credit assets. Here, we find weaker results, so the mechanism seems to be more relevant for credit expansion than for other bank assets.

Alternative profit measures:. Furthermore, the appendix also shows results for different definitions of the explanatory variables. In Table 3.A.6 we vary the denominator and normalize net income by GDP and CPI. In Table 3.A.7 we include levels in profitability measures instead of changes. In all these specifications, profitability robustly predicts credit expansion.

Timing:. We explore the dynamic relationship between profits and credit growth by shifting the dependent variable over time in appendix section 3.A.5. The response of the credit-to-GDP ratio to variation in profitability measures is strongest over the subsequent three years – our baseline evidence – and slowly dissipates at longer horizons. We also find that profit changes and credit growth are contemporaneously negatively correlated. This evidence indicates that high bank profitability and the credit cycle are unlikely to be linked through correctly anticipated improvements of economic fundamentals. The expansion is associated with decreasing rather than improving bank profitability, in line with the observed decline in output (Mian, Sufi, and Verner, 2017).

Country level evidence:. In Figure 3.A.1 we plot the coefficients at the country level. We run a time series regression of $\Delta_3 y_{i,t+3}$ on profitability measures for all our sample countries one by one. The graphs show that the coefficients are all positive and significant in a majority of countries, so that the strong association between profitability and credit expansion seems to be a common feature across our sample countries.

Table 3.3. Multivariate models for changes in credit-to-GDP, profit components

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1) $\frac{Revenue}{Equity}$	(2) $\frac{Costs}{Equity}$	(3) $\frac{LoanLosses}{Equity}$	(4) $\frac{Revenue}{Assets}$	(5) $\frac{Costs}{Assets}$	(6) $\frac{LoanLoss}{Assets}$
$\Delta_3 Change_{i,t}$	-0.01	-0.10	-0.27***	0.40	-1.14	-2.52***
	(0.05)	(0.09)	(0.06)	(0.65)	(2.57)	(0.53)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓
R^2	0.22	0.22	0.24	0.22	0.22	0.24
Observations	855	855	855	855	855	855

Notes: This table reports regressions of credit-to-GDP changes from t to $t+3$ on levels and three-year changes in banking sector revenue (net interest + net fee income), costs (administrative expenses) and loan losses. All specifications control for three lags of credit-to-GDP changes and a vector of balance sheet constraint and macroeconomic control variables (see text in section 3.3). All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Bank level evidence:. The channels that link profits and subsequent credit growth should also be operative at the bank level. Indeed, the literature already provides evidence that higher bank capital is associated with more lending (Jiménez, Ongena, Peydró, and Saurina, 2017) and that loan officers extrapolate from their recent experiences (Carvalho, Gao, and Ma, 2020). In appendix section 3.A.6 we study whether the relationships we are describing here also hold in US bank level panel data. This allows us to control for financial constraints at the bank level which could matter if aggregate leverage ratios are hiding changes in the distribution of leverage ratios. The data also allows to rule out that aggregate credit demand is driving results by including time fixed effects. We find that the relationship between profits and credit expansion remains highly significant.

3.3.2 Decomposing profitability

So far, our analysis was based on bottom-line measures of profitability, *RoE* and *RoA*, which are both based on net income of the banking sector. We now re-estimate the profit-credit relationship for three major constituents of bank profits: revenue, operating costs and loan losses. This decomposition will help us to gain further insights into the mechanisms underlying the profit-credit cycle.

We define six new variables, expressing each of the separate profit components relative to equity and assets to maintain comparability to the baseline estimates. We then run regressions of the following form

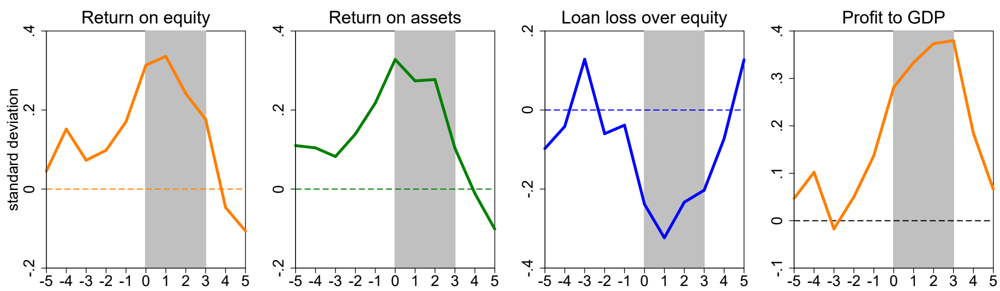
$$\Delta_3 y_{i,t+3} \;=\; \alpha_i + \beta\,\Delta_3 (Revenue/Equity)_{i,t} + \eta X_{i,t} + u_{i,t+3}\,, \qquad (3.5)$$

where we replace $\Delta_3(Revenue/Equity)_{i,t}$ with costs and loan losses and vary the denominator between *Equity* and *Assets* across specifications. The results are shown in Table 3.3. The results for loan losses are highly significant. A decrease in loan losses is associated with subsequent credit expansion. Revenues and cost variables only display a weak relationship with subsequent credit expansion. Table 3.A.15 presents corresponding evidence at the bank level for our sample of US banks. Again, loan losses are highly significant.[4]

We will look at the channels that link profits and credit expansion in more detail below, but this finding is already at odds with models exclusively based on financial constraints. Financial constraints most likely do not depend on the sources of income. Hence, we should observe similar coefficients for all three profitability components. However, we find that there is something particular about loan losses and the source of income contains information over and above the raw change in leverage or net worth. While not necessarily unique to these theories, the pattern is consistent with models that feature updating of beliefs based on recent credit market outcomes. In Bordalo, Gennaioli, and Shleifer (2018) agents' expectations overweight states of the world that have become more likely in the light of new data. Applied to our setting, news about low or decreasing loan losses could lead agents to assign an inflated probability to future states of the world with low defaults. These low expected losses enter the lending decisions of banks and thereby create an incentive to expand lending in line with the empirical results presented above.

3.3.3 Credit Booms

Figure 3.3. Event study of profitability around credit boom dates, standardized



Notes: These figures display the evolution of profit variables around credit booms. All variables are detrended and standardized with mean zero and standard deviation one by country. Observations are classified as boom years when $\Delta_3 Loans/GDP_{i,t}$ exceeds one standard deviation. 0 refers to a year in which a credit boom starts. The grey area marks the three-year window used to define the credit boom. Solid lines display means of variables around the start of a boom. See text.

4. In the bank level data, changes in non-interest expenses turn out to be a significant predictor of credit growth as well.

The previous subsections have shown that there is a tight link between bank profitability and subsequent credit growth. While this relationship is interesting per se, policymakers naturally care more about large credit booms which are often followed by costly crises (Schularick and Taylor, 2012). How then is profitability related to these large credit booms?

The notion of a large credit boom implies a strong deviation from normal circumstances. To be consistent with the other exercises, we will identify large credit booms from the three-year change in the ratio of credit to GDP. As normal circumstances may differ across countries, we first standardize the three-year change in the ratio of credit to GDP at the country level. We then define year t as the start of a credit boom if the change of the credit to GDP ratio between year t and year $t+3$ exceeds one standard deviation.[5] With this definition, the probability to experience the start of a credit boom is 4.9% per year (95 booms), which is roughly similar to the frequency of banking crises. When we look at the starting years of the booms, we see that many well-known historical boom episodes are reflected by this definition. For example, we detect the start of a credit boom in 10 of our sample countries between 2000 and 2005.

In Figure 3.3 we show the evolution of profitability variables around these credit boom episodes. The graphs are centered around the start of a credit boom as defined above. Starting with the evolution of *RoE*, we see that unusually large three-year increases in credit-to-GDP are preceded by an increase in *RoE*. *RoE* is close to the sample average 5 years prior to the start of the boom and increases on average by a third of a country-specific standard deviation until the credit boom starts in $t=0$. *RoE* peaks at the start of the credit boom and starts falling as the boom continues.[6] 4 years into the boom, *RoE* is back at the sample mean. The patterns are almost the same when we look at *RoA* or when we constrain the sample to post 1945 credit booms in Figure 3.A.2. In the third panel, we see that loan losses are a major driver of these developments. Loan losses are decreasing before the credit boom starts. Once the credit boom is underway, loan losses start increasing again and they are back to the sample mean 5 years after the credit boom started. The fourth panel shows the evolution of bank profits relative to GDP around credit boom dates. The pattern here differs slightly from the previous graphs. While profitability ratios, profits per unit of equity or assets, are reversing quickly during the boom, profits relative to GDP remain elevated for some more years. Increasing quantities of intermediated funds balance decreasing profits per unit during the boom.

5. If there are subsequent observations fulfilling this condition, we group all these observations into one credit boom episode and define the first year of this boom episode as the starting year.

6. Based on our definition the boom lasts at least for three years, these three years are marked in grey in the graph. Note that many of the booms we detect last actually longer and credit is elevated for a few more years.

Given the strong association between credit expansion and crisis likelihood, and taking into account the beneficial effects of bank capital during a crisis (Jordà, Richter, Schularick, and Taylor, 2021), it may be optimal if banks increase capital buffers during credit expansions. Our results suggest that there exists additional wiggle room for banks to increase capital buffers during booms, strengthening the case for recent regulatory efforts to implement countercyclical capital requirements and to limit dividend payouts.[7]

Turning to a formal econometric model to study the link between profitability and large credit booms, we estimate probit regressions with the indicator variable $B_{i,t}$ for the start of a boom as the dependent variable. We assume, as is standard in the literature, that the probability of a boom start conditional on observables $X_{i,t}$ can be represented in terms of the normal cumulative distribution function,

$$Pr[B_{i,t} = 1 | \alpha_i, X_{i,t}] = \Phi(\alpha_i + \boldsymbol{\beta} X_{i,t}). \tag{3.6}$$

Here α_i is a country fixed effect and $X_{i,t}$ includes three-year changes in profitability and a vector of macroeconomic control variables. The results are shown in Table 3.4. The odd-numbered columns show estimates with profit changes and country fixed effects only while the even-numbered columns also include the full set of control variables. Three-year changes in *RoE*, *RoA* and *LoE* (loan losses relative to equity) are significantly related to the start of a large credit expansion. A one standard deviation higher $\Delta_3 RoA$ is associated with an increase of one percentage point in the probability of experiencing the start of a credit boom. Given the average boom frequency of 4.9%, this is a quite sizable effect.

3.4 Channels

This section studies the mechanism that links bank profitability and credit growth in further detail. First, we present evidence in favor of the hypothesis that bank profitability is associated with expansions in credit supply. We then distinguish between different credit supply channels and find that the relationship cannot be fully explained by financial constraints of intermediaries and is instead consistent with mechanisms featuring time-varying beliefs and extrapolation.

3.4.1 Credit demand and supply

Credit expansions follow improvements in bank profitability. This relationship could be due to an increase in the supply of credit or due to higher demand for credit. In

7. Figure 3.A.3 shows that capital ratios remain constant or increase slightly during the boom. For a subsample of booms our data allows us to distinguish between profits that are paid out and profits that are retained as equity. Figure 3.A.3 shows a significant share of profits is paid out to shareholders at later stages of the boom.

Table 3.4. Multivariate probit models for boom prediction

	$\Delta_3 RoE_{i,t}$		$\Delta_3 RoA_{i,t}$		$\Delta_3 LoE_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
See column header	0.22**	0.24***	3.64***	3.02***	-0.35***	-0.33***
	(0.11)	(0.09)	(0.80)	(1.07)	(0.06)	(0.08)
Country fixed effects	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓
AUC	0.65	0.75	0.66	0.75	0.66	0.75
Observations	1658	1491	1669	1491	944	889

Notes: The table shows probit classification models where the dependent variable is a an indicator that is one at the start of a credit boom and zero else. Coefficients are marginal effects. Controls includes the three most recent values of short and long term interest rates, GDP growth, inflation and the current account as well as three-year changes in credit-to-GDP between $t-3$ and t. Country clustered standard errors in parentheses. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

our data, a simple test can help to distinguish between demand and supply-based explanations. More specifically, the two yield conflicting predictions regarding the price of credit during a credit expansion. The price of credit should be high if credit expansions after high bank profitability are due to an outward shift in credit demand. On the other hand, the price of credit should be low if high profitability is associated with increased supply of credit by the banking sector. We use data on bond spreads from Kuvshinov (2018) as a measure of the price of credit to test these competing hypotheses.[8] We analyse the relationship between spreads and three-year changes in profitability:

$$Bond\ spread_{i,t+1} = \alpha_i + \beta \Delta_3 RoE_{i,t} + \gamma X_{i,t} + u_{i,t}. \tag{3.7}$$

The results are presented in column (1) of Table 3.5. The price of credit in the next period is negatively associated with recent changes in profitability. In combination with our baseline result, namely an expansion of credit following improvements in bank profitability, this suggests that credit supply explanations better capture the dynamics than demand side explanations. This result is robust to adding financial constraint proxies and macroeconomic controls as can be seen in columns (2) and (3).The price of credit is low when banking sector profitability increases. This finding corroborates earlier work on supply driven credit cycles (Krishnamurthy and

8. Hence, like usually done in this literature, we implicitly assume that lending standards in bond and bank credit markets are correlated. The advantage of bond spreads compared to data on lending rates is that they are forward-looking and immediately reflect lending conditions for new credit and greater data availability.

Table 3.5. Multivariate models for credit spreads

	Dependent variable: *Bond Spread*$_{i,t+1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	-1.04***	-1.07**	-0.84**			
	(0.35)	(0.48)	(0.43)			
$\Delta_3 RoA_{i,t}$				-20.93***	-19.92***	-20.23***
				(6.75)	(5.68)	(5.15)
Country fixed effects	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.00	0.09	0.11	0.01	0.09	0.11
Observations	1279	1279	1279	1279	1279	1279

Notes: This table reports regressions of credit spreads in $t+1$ on three-year changes in *RoE*. Column (2) adds the vector of macroeconomic control variables, column (3) additionally includes financial constraint proxies (see text in section 3.3). All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Muir, 2017; Mian, Sufi, and Verner, 2017) adding an explanation for expansions in credit supply.

3.4.2 Disentangling supply based explanations

In a next step, we want to distinguish between two possible credit supply explanations of the profit-credit cycle. Here, we collect two additional pieces of evidence which suggest that financial constraints alone cannot explain the profit-credit relationship. First, we decompose return on equity into retained earnings over equity (*REToE*) and dividends over equity (*DoE*) for a subset of countries and years. The underlying idea of this exercise is simple. Dividends paid out to shareholders are not available in the banking sector to relax financial constraints. We can hence use *DoE* as a measure of profitability that is unrelated to changes in net worth. Applying this logic, the results in columns (1) and (2) of Table 3.6 confirm that the link between profits and credit expansion goes beyond a pure financial constraints channel. Column (1) shows that the growth in *DoE* over the previous three years is a predictor of credit expansion over the next three years. Retained earnings are robustly linked to subsequent credit expansion in column (2), but their addition does not affect the relationship between dividends and subsequent credit expansion.[9]

As a second piece of evidence, we include the 3-year change in profitability together with the level of additional net worth gained over the three-year window in

9. Since retained earnings directly measure changes in net worth, we do not include indirect controls of financial constraints here.

Table 3.6. Multivariate models for changes in credit-to-GDP

	Dependent variable: $\Delta_3 y_{i,t+3}$			
	Uses of profits		Profit path	
			RoE	RoA
	(1)	(2)	(3)	(4)
Δ_3*Dividends over Equity*$_{i,t}$	0.85***	0.78***		
	(0.24)	(0.25)		
Δ_3*Retained earnings over Equity*$_{i,t}$		0.22***		
		(0.07)		
$3-$ *year Accumulated Profits*$_{i,t}$			0.10***	1.02***
			(0.03)	(0.27)
Δ_3*Change*$_{i,t}$			0.28***	3.37***
			(0.04)	(0.66)
R^2	0.186	0.200	0.224	0.221
Country fixed effects	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓
Control variables	✓	✓	✓	✓
Observations	949	949	1485	1486

Notes: This table reports regressions of credit-to-GDP changes for t to $t+3$. Columns (1) and (2) focus on the uses of profits by decomposing $\Delta_3 RoE_{i,t}$ into changes in dividends over equity ($\Delta_3 DoE_{i,t}$) and retained earnings over equity ($\Delta_3 REToE_{i,t}$). Columns (3) and (4) study the profit path and include both, the level of accumulated profits ($RoE_{i,t}$) and the change ($\Delta_3 RoE_{i,t}$) in the same regression. All specifications control for the three most recent values of credit-to-GDP changes and macroeconomic control variables (see text in section 3.3). All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

one specification. We compute this measure as the sum of profits over the three year window, scaled by pre-existing capital or assets. Controlling for the cumulative profits over the three-year window, $\Delta_3 RoE$ is a proxy for the path the banking sector took to arrive at a certain level of profitability. When agents update expectations overweighing recent information, this path will affect expectations. Under diagnostic expectations for example, for the same level of three-year accumulated profits, agents may be more optimistic when profits over this period have been increasing as opposed to recent decreases. Columns (3) and (4) show that three-year changes in profitability predict a credit expansion over the next years, even when controlling for the level of additional net worth gained over this period. We repeated these two exercises using bank level data from the United States. The results in Table 3.A.16 confirm the aggregate findings presented here.

3.4.3 Survey expectations and credit expansions

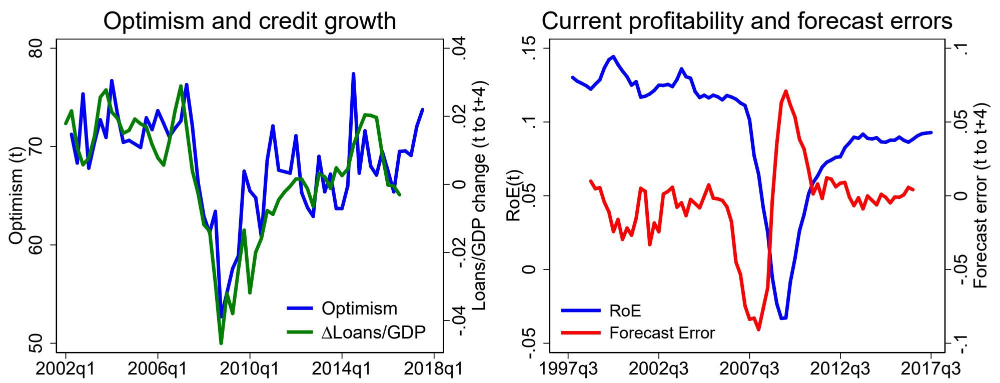
The long run data shows that profitability in the banking sector predicts credit expansion. As we have argued before, this relationship is consistent with models of credit cycles that feature extrapolative expectations. A quickly growing literature is using survey responses to understand how economic agents are actually forming their expectations (e.g. Bordalo, Gennaioli, Ma, and Shleifer, 2020). Survey-based information about bankers' expectations is however scarce, especially in a long run cross-country setting. We therefore complement our approach with an analysis of recent survey data from the United States. Based on responses of bank CFOs (from the Duke CFO Global Business Outlook, 2018), we ask whether optimism and expectations about future profitability are related to recent changes in profitability and to subsequent changes in bank credit.

The Duke CFO Global Business Outlook (2018) asks respondents to rate their optimism about the financial prospects of their own company on a scale from 0-100, with 0 being the least optimistic and 100 being the most optimistic. CFOs are further asked about their expectations of changes in earnings over the next twelve months. For both questions, we have quarterly data on the mean response of CFOs from the banking and finance industry (starting in 2002 and 1998 respectively). We combine these measures with quarterly accounting information on realized profitability and credit growth for the US banking sector.[10] The baseline relationships between profitability measures and subsequent credit growth in this sample mirror the correlations in the long-run cross-country data (see Figure 3.A.7 in the appendix). When we look at the consistency of the two survey measures, we find that reported CFO optimism and earnings growth expectations are indeed highly correlated (see Figure 3.A.8).

To study the links between profitability, optimism and credit growth, the left panel of Figure 3.4 shows that optimism at time t and changes in the credit-to-GDP ratio between t and $t+4$ (i.e. over one year) track each other closely. The banking sector extends more credit over the following year, when CFO optimism is elevated today. Optimism is an appealing measure for credit market sentiment, but it is important to note that optimism could be justified by subsequent developments in profitability. We therefore rely on additional survey responses about expected changes in earnings in the next 12 months to compare realized and expected profitability. We first calculate the time t expectation of RoE_{t+4} multiplying actual earnings over the past twelve months at time t with expected earnings changes over the next twelve months scaled with time t equity capital.

10. Quarterly balance sheet and income information are based on FDIC statistics. We use aggregated data from quarterly banking profile spreadsheets, in particular "Assets and Liabilities of FDIC-Insured Commercial Banks and Savings Institutions" and "Quarterly Income and Expense of FDIC-Insured Commercial Banks and Savings Institutions". The data can be accessed here https://www.fdic.gov/bank/analytical/qbp/.

Figure 3.4. CFO expectations and the profit-credit cycle



Notes: The left panel presents the evolution of bank CFO optimism and subsequent 4-quarter changes in the ratio of net loans and leases to GDP (between t and $t+4$) in the United States. The right-hand panel displays the evolution of bank RoE_t and time t return on equity forecast errors ($RoE_{t+4} - E_t[RoE_{t+4}]$) of bank CFOs in the United States between 1997 and 2017. See text.

$$E_t[RoE_{t+4}] = \frac{ExpectedChange_{t\to t+4} \times \sum_{i=0}^{3} NetOperatingIncome_{t-i}}{EquityCapital_t} \tag{3.8}$$

We compare $E_t[RoE_{t+4}]$ to realized RoE_{t+4} computed as realized earnings over the following twelve months also scaled with time t equity capital. We refer to the difference between the two as the time t forecast error ($Error_t = RoE_{t+4} - E_t[RoE_{t+4}]$). The time series for this variable is visualized in the right-hand panel of Figure 3.4 together with realized profitability over the past twelve months. The negative relationship between the two measures suggests that CFOs are too optimistic (expected profitability is higher than realized profitability) when current *RoE* is high and vice versa.

Table 3.7 presents empirical tests of these relationships. In column (1), we find a positive and significant relationship between changes in optimism and changes in *RoE*. An increase in profitability is associated with a more optimistic outlook of the average CFO on the future financial prospects of the bank. Column (2) shows that this optimism is not justified in the data. There is in fact no association between changes in *RoE* today and the change over the next year. At the same time, in line with the optimism measure, expectations of profitability over the following year are elevated if *RoE* increases (column (3)). As a result, expectations are systematically biased. The difference between realized and expected earnings, the forecast error, is negatively related with changes in *RoE*. Put differently, an increase in *RoE* is associated with an increase in expected profitability relative to realized profitability over the following year. In column (5), we study the implications for credit supply conditions. The dependent variable here is the change in the net percentage of banks

Table 3.7. Relationship between profitability, expectations about future profitability and credit supply conditions

	$\Delta Optimism$	ΔRoE_{t+4}	$\Delta E_t(RoE_{t+4})$	$\Delta Error$	$\Delta\%Tightening$
	(1)	(2)	(3)	(4)	(5)
ΔRoE_t	1.70***	0.06	0.73***	-0.66***	-7.14***
	(0.52)	(0.14)	(0.19)	(0.23)	(0.99)
R^2	0.08	0.00	0.17	0.10	0.18
Observations	57	78	73	69	82

Notes: This table reports estimates for univariate regressions of expectation measures on the change in *RoE*. In column (1), the dependent variable is the quarterly change in optimism from the bank CFO survey, in column (2) the quarterly change in realized earnings between t and t+4 normalized with equity capital at time t, in column (3) the quarterly change in expected earnings between t and t+4 normalized with equity capital at time t, in column (4) the quarterly change in the difference between realized and expected earnings between t and t+4, and in column (5) the change in the net percentage of domestic banks tightening standards for C&I loans to large and middle-market firms. Newey-West standard errors in parentheses are computed using the automatic bandwidth selection procedure in Newey and West (1994). *, **, ***: Significant at 10%, 5% and 1% levels respectively.

tightening standards for loans to large and middle-market firms from the Federal Reserve's senior loan officer opinion survey. The negative coefficient implies that a significant fraction of banks loosens credit standards when *RoE* increases. Appendix Table 3.A.10 and Table 3.A.11 show very similar results for RoA as an alternative profitability measure and consistent but weaker results when excluding the years 2007-2009.[11]

In a second step, we link these variables to the credit cycle. Gennaioli, Ma, and Shleifer (2016) shows that firm investment is explained by earnings expectations of CFOs. We now ask whether this pattern also holds for banks, interpreting credit growth as banks' investments. In the quarterly US data, we measure credit growth as the change in the ratio of net loans and leases to GDP between t and $t+4$. Column (1) in Table 3.8 confirms that 4-quarter changes in credit are predicted by optimism, where lagged credit growth, a crisis and a recession dummy, as well as GDP growth, interest rates and bank capital ratios are included as control variables. In column (2) we include realized profitability over the past year. Columns (3) and (4) analyze the relationship between profit forecasts and credit growth. The profit forecast itself (column 3) is positively related to subsequent credit growth. When expected profits are high, credit grows rapidly. The forecast error is negatively related to credit growth: credit growth is low when bank CFOs are excessively pessimistic

11. It is clear from the graph that most of the variation is during and shortly after the 2007/2008 financial crisis. Figure 3.4 shows that the crisis was a surprise to bank CFOs. The positive forecast error after the crisis also suggests that bank CFOs were excessively pessimistic (compared to realized profits one year later) when profits were lowest during the crisis.

Table 3.8. Multivariate models for changes in credit-to-GDP, profitability and expectations

	Dependent variable: 4-quarter change in credit/GDP				
	(1)	(2)	(3)	(4)	(5)
	Optimism	RoE_t	$E_t(RoE_{t+4})$	*Error*	*%Tightening*
RHS variable (see column header)	0.13***	0.48***	0.36***	-0.19***	-0.02**
	(0.04)	(0.02)	(0.03)	(0.03)	(0.01)
R^2	0.79	0.88	0.85	0.71	0.66
Controls	✓	✓	✓	✓	✓
Observations	56	75	71	71	75

Notes: The dependent variable is the change in the ratio of net loans and leases to GDP between t and t+4. In column (1) this change is regressed on optimism from the bank CFO survey, in column (2) on realized earnings between t-4 and t normalized with equity capital at time t, in column (3) on expected earnings between t and t+4 normalized with equity capital at time t, in column (4) on the difference between realized and expected earnings between time t and t+4 normalized with equity capital at time t, in column (5) on the net percentage of domestic banks tightening standards for C&I loans to large and middle-market firms. Newey-West standard errors in parentheses are computed using the automatic bandwidth selection procedure in Newey and West (1994). *, **, ***: Significant at 10%, 5% and 1% levels respectively.

and it is high when they are excessively optimistic. Finally, column (5) illuminates one possible channel and shows that a tightening (loosening) in the standards at which banks supply credit is associated with lower (higher) credit growth over the following year.

Overall, the findings are consistent with the idea that bankers' expectations rely excessively on recent performance. Furthermore, survey-based measures of expectations are linked to credit growth, and expectational errors are reflected in the growth rate of credit.

3.4.4 Discussion of results

We find that bank profits predict credit growth in general, and that increases in profitability forecast large credit booms. The results in this section suggest that these credit expansions are driven by credit supply rather than demand. Disentangling different supply-based explanations, the results are consistent with the predictions of recent behavioral credit cycle models that incorporate time-varying beliefs due to overweighting of recent experience. The analysis of recent survey data on bank CFO expectations supports this interpretation.

The relationship between recent fundamentals, in particular loan losses, and subsequent credit expansion corroborates earlier work by Greenwood and Hanson (2013) who find that bond issuance increases after periods of low defaults in a US time series. They note that one possible explanation for this result are extrapolative expectations. Our new long run data on bank profitability and the decomposition into sources and uses of funds allows us to carefully separate time-varying expectations from financial constraints, and we arrive at a similar conclusion. Furthermore, the long run bank profitability data also allows us to jointly study the bank credit cycle and recent bank performance, instead of linking bond market developments and bank credit (Kirti, 2020).

Similar to our results, Baron, Verner, and Xiong (2020) document a positive relationship between shareholder returns and subsequent credit expansion. However, the accounting data have two important advantages relative to market based data. Shareholder returns largely reflect expectations about future profitability or time-varying discount rates and not current fundamentals (Shiller, 1981; Cochrane, 2017). Hence, a relationship between shareholder returns and subsequent outcomes could simply mean that shareholders correctly anticipate future developments: low returns would then forecast low future GDP growth, which may be associated with little profitable lending opportunities for banks. The accounting data on past profitability allows us to circumvent this problem. Furthermore, as noted by Meiselman, Nagel, and Purnanandam (2018), due to the concavity of bank asset returns equity prices are most likely informative during a crash, but not very informative during good times. This is reflected in the differences between Baron, Verner, and Xiong (2020) and our results. While they find stock returns to be particularly informative about future credit growth during stock market crashes, we find that also large credit booms can be forecasted with increases in profitability.

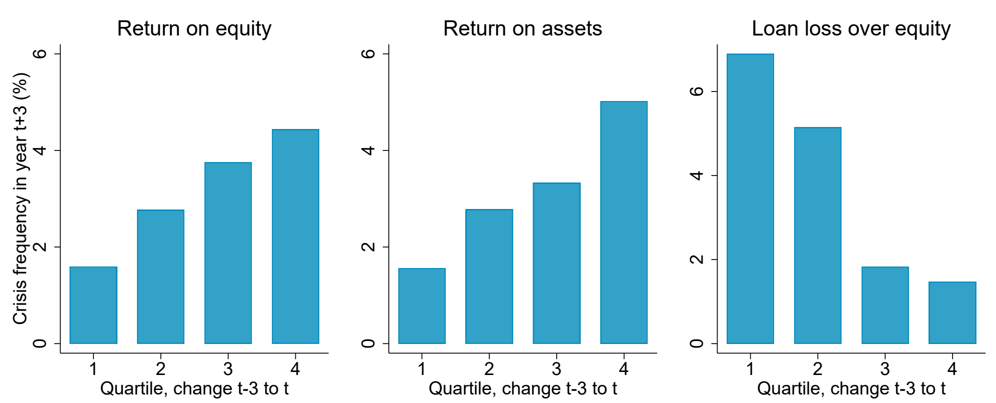
3.5 Profits and Crisis

Minsky (1977) and recent formalizations thereof such as Bordalo, Gennaioli, and Shleifer (2018) and Greenwood, Hanson, and Jin (2019) suggest that increases in profitability should be associated with optimism and credit expansion first, and with predictable crises a few years later when optimism wanes. We have seen that credit booms can be forecasted with profitability measures. We now show that banking crises a few years ahead can also be predicted with increases in bank profitability. Furthermore, we will show that bank profitability allows us to characterize the transition from credit boom to crisis and to distinguish between panic and non-panic crises.

3.5.1 Predicting Crises

Is there a systematic relationship between increases in profitability and banking crisis risks? As a simple way to study this relationship, we sort observations into four equal-sized bins based on the change in profitability (RoE and RoA) between $t-3$ and t. Figure 3.5 shows the frequency of the start of banking crises in year $t+3$ (as a measure of medium term crisis risk) for each bin. We rely on the narrative chronology by Jordà, Schularick, and Taylor (2017) to identify crises events. The

Figure 3.5. Crisis probability in t+3 by change in profitability



Notes: This figure shows the relationship between changes in RoE (RoA and LoE) between $t-3$ and t and financial crisis frequencies for the year $t+3$. Observations are sorted into four equal-sized bins according to the increase in RoE (RoA) between $t-3$ and t. Vertical bars indicate the frequency of financial crises in year $t+3$ for each of the bins.

yearly banking crisis start frequency in our sample is 3.1%. Focusing on RoE in the left panel, we see that banking crisis frequencies in $t+3$ are below 2% in the bin with the lowest changes in RoE over a three year window. This is in stark contrast to observations in the highest quartile of increases in RoE where the crisis frequency in $t+3$ exceeds 4%. A similar pattern is observable looking at crisis frequencies when binning observations based on $\Delta_3 RoA$. Here, the frequency of a banking crisis in the year $t+3$ rises up to about 5% if the three-year change in RoA is in the top quartile. Finally, the right panel looks at changes in loan losses and the frequency of future financial crises is highest when loan losses declined the most, as shown by the crisis frequency of more than 6% in the lowest quartile of loan loss changes. Figure 3.A.4 and Figure 3.A.5 show that similar patterns can be observed when restricting the sample to post 1945 data or when using a chronology of banking panics from Baron, Verner, and Xiong (2020).

In a second step, we ask whether different crisis frequencies in quartiles of $\Delta_3 RoE$ and $\Delta_3 RoA$ are due to the relationships between profits and credit only, or whether profits contain additional information. In Table 3.9, we look again at the frequency of crises for the different quartiles of profitability increases. The mean values reported in the bottom row of the table correspond closely (with small sample differences) to the probabilities displayed in Figure 3.5. Here, we additionally divide each quartile in the profitability distribution ($\Delta_3 RoE$ and $\Delta_3 RoA$) into three bins based on changes in the ratio of credit-to-GDP. In the right column, we report the crisis frequency for low, medium and high credit growth observations. As expected, the frequency of crises in $t+3$ is increasing in credit growth. Focusing on the results in Table 3.9, we see that crises frequencies are generally increasing from

Table 3.9. Crisis frequency in t+3 by credit growth level and profit change quartile

	$\Delta_3 RoE_{i,t}$ quartile				$\Delta_3 RoA_{i,t}$ quartile				
Credit growth	1	2	3	4	1	2	3	4	Mean
Low	0.00	1.24	1.90	2.58	1.26	1.27	1.89	2.63	1.91
Medium	1.88	2.47	4.29	2.52	0.61	3.14	4.29	2.48	2.80
High	2.45	3.77	6.21	8.23	1.84	4.40	4.35	9.55	4.67
Mean	1.46	2.49	4.15	4.45	1.24	2.94	3.52	4.89	3.13

left to right, and from the top to the bottom. Both dimensions, profitability increases and credit expansion are associated with crisis incidence. No crisis occurred in the last 150 years when $\Delta_3 RoE_{i,t}$ was in the lowest quartile (column 1) of observations and credit growth was also low. When credit growth was higher (third row), the frequency of crises increased (2.45%), but was still well below the sample average (3.13%). Focusing on high credit growth (third row), the table shows that crisis incidence three years ahead increases to more than 8 percent, when we move to the highest quartile of profitability increases. These patterns are also reflected in the right panel that uses $\Delta_3 RoA$ to bin the data.

We will now explore these relationships econometrically using prediction models that relate changes in bank profitability to the likelihood of experiencing a financial crisis. Specifically, we estimate a probit model for a financial crisis starting in country i in year $t+3$, denoted by the indicator variable $C_{i,t+3}$,

$$Pr[C_{i,t+3} = 1|\alpha_i, X_{i,t}] = \Phi(\alpha_i + \boldsymbol{\beta} X_{i,t}). \tag{3.9}$$

$X_{i,t}$ includes three-year changes in profitability as well as a vector of control variables. The control vector contains 3-year changes in credit to GDP to proxy for the well-known relationship between credit and financial crises (Schularick and Taylor, 2012). $\boldsymbol{\beta}$ denotes the vector of coefficients of interest for the various specifications. We follow the literature and include country fixed effects to account for cross-country heterogeneity in the risk of experiencing a financial crisis.

Table 3.10 reports marginal effects for the relationship between changes in profitability measures and crisis likelihood in year $t+3$. The odd-numbered columns show results only including the country fixed effects, even-numbered columns also include 3-year changes in credit to GDP. The results confirm the visual impression from the binned scatterplots. Column (1) shows that an increase in RoE over three years by 5 percentage points (about 1 standard deviation) is associated with a 1 percentage point higher crisis probability in year $t+3$. This result is unaffected by the inclusion of credit as a control variable in column (2). The relationship between increases in profitability and crisis risk goes above and beyond the mere effect of profitability on credit expansion and the associated increase in crisis risk. We ob-

Table 3.10. Multivariate probit models for systemic financial crisis prediction

	$\Delta_3 RoE_{i,t}$		$\Delta_3 RoA_{i,t}$		$\Delta_3 LoE_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Profitability (see column header)	0.21***	0.23***	2.70***	3.05***	-0.31***	-0.40***
	(0.07)	(0.06)	(0.62)	(0.65)	(0.08)	(0.07)
$\Delta_3 Loans/GDP_{i,t}$		0.18***		0.18***		0.26***
		(0.03)		(0.03)		(0.04)
AUC	0.67	0.72	0.67	0.72	0.67	0.74
Observations	1700	1641	1721	1647	916	914

Notes: The table shows probit classification models where the dependent variable is an indicator that is one if the country experiences the start of a financial crisis in year $t + 3$ and zero else. Coefficients are marginal effects. Regressors are described in the column header. All models include country fixed effects. Country clustered standard errors in parentheses. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

tain similar results with $\Delta_3 RoA$ and $\Delta_3 LoE$. Table 3.A.8 shows that these patterns are robust to using panic banking crisis chronology from Baron, Verner, and Xiong (2020).

3.5.2 Transition into crisis

As Mian and Sufi (2018) argue, we are still missing a good understanding of the transition from credit booms into crises. Credit cycle theories featuring diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018) predict a sharp reversal when, after a period of good news, fundamentals turn out to be disappointing, i.e. they decrease or grow at a lower pace. We take this idea to the data and operationalize it using the yearly change δ in our main profitability measures. Since $\Delta_3 RoE$ is already a change (over a three-year horizon), the first difference in this variable, $\delta\Delta_3 RoE$ measures whether profits are growing at an increasing or decreasing pace. We then run a sequence of probit regressions of the form

$$Pr[C_{i,t+h} = 1|\alpha_i, X_{i,t}] = \Phi(\alpha_i + \boldsymbol{\beta} X_{i,t}), \tag{3.10}$$

for $h = 0, 1, ..., 5$. In this procedure, we keep the RHS of the equation fixed and move the prediction horizon for a banking crisis into the future. $X_{i,t}$ now includes $\Delta_3 RoE$ and its change $\delta\Delta_3 RoE$ as well as three-year changes in credit-to-GDP ratios. The results for $h = 3$ in column (3) correspond to the specification in Table 3.10, only that we additionally include $\delta\Delta_3 RoE$. We first see that the results for $\Delta_3 RoE$ do not depend on the exact timing and $\Delta_3 RoE$ captures crisis risk over the medium term ($t + 2$ to $t + 4$). On the other hand, the coefficient for $\delta\Delta_3 RoE$ in the second row is negative and significant for short horizons of 1 or 2 years: while three-year increases

Table 3.11. Multivariate probit models for systemic financial crisis prediction – crisis transition

	Dependent variable: Crisis at time...					
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+4	(6) t+5
$\Delta_3 RoE_{i,t}$	-0.30***	0.04	0.26**	0.29***	0.19**	0.02
	(0.04)	(0.08)	(0.10)	(0.08)	(0.09)	(0.11)
$\delta\Delta_3 RoE_{i,t}$	-0.11*	-0.18*	-0.16**	-0.09	0.00	0.06
	(0.06)	(0.11)	(0.06)	(0.07)	(0.09)	(0.08)
$\Delta_3 Loans/GDP_{i,t}$	0.10**	0.18***	0.20***	0.16***	0.10***	-0.01
	(0.04)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)
AUC	0.87	0.73	0.72	0.73	0.69	0.62
Observations	1667	1650	1633	1616	1599	1582
	Dependent variable: Crisis at time...					
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+4	(6) t+5
$\Delta_3 RoA_{i,t}$	-3.34***	0.94	3.60***	3.61***	2.21*	0.96
	(0.44)	(1.04)	(0.74)	(0.74)	(1.17)	(1.49)
$\delta\Delta_3 RoA_{i,t}$	-1.78**	-2.12**	-2.10***	-0.94	-0.09	-0.13
	(0.71)	(0.84)	(0.63)	(0.73)	(1.28)	(1.04)
$\Delta_3 Loans/GDP_{i,t}$	0.10***	0.19***	0.20***	0.17***	0.10***	-0.01
	(0.03)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)
AUC	0.86	0.72	0.72	0.72	0.69	0.62
Observations	1675	1658	1641	1624	1607	1590

Notes: The table shows probit classification models where the dependent variable is an indicator that is one if there is a crisis in $t + h$ years, specified in the column header. Coefficients are marginal effects. All specifications include country-fixed effects. Country clustered standard errors in parentheses. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

in profitability are associated with a higher probability of crisis in the medium term, the short-term risk of a crisis increases sharply once the profitability spurt is wearing off. For completeness, column (1) shows the contemporaneous correlation between profitability measures and crisis indicators. Both measures are significantly negative, indicating that crisis are associated with strong decreases in *RoE*. The second panel shows that these patterns are similar for *RoA* and Appendix Table 3.A.9 shows that these results do not depend on the crisis chronology employed.

3.5.3 Panic crises vs. non-panic crises

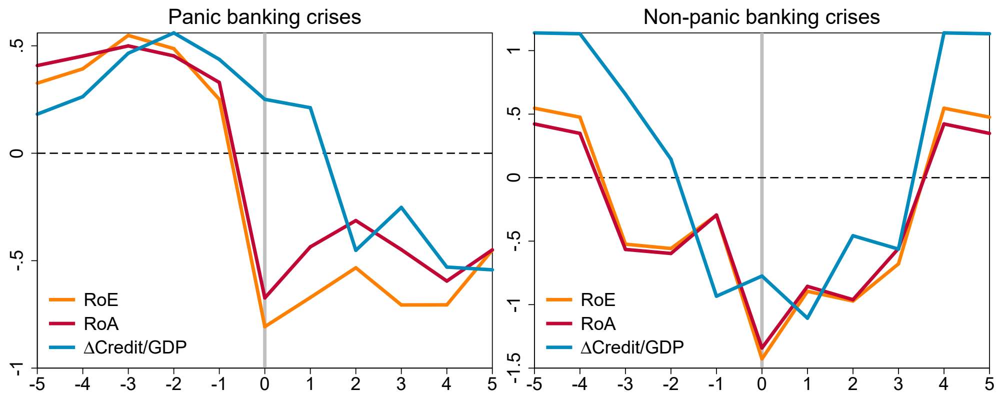
Banking crises are often associated with bank runs and panics. In fact many banking crisis chronologies rely on the very occurrence of panics to define a crisis. In a recent contribution, Baron, Verner, and Xiong (2020) distinguish between crises with and without panics and define panic crises as “severe and sudden withdrawals of funding by bank creditors from a significant part of the banking system”. Non-panic crises on the other hand are defined as large declines in bank net worth based on bank index returns. As a result their set of panic crises closely overlaps with the definition of a crisis in other chronologies (as the (Jordà, Schularick, and Taylor, 2017) chronology used in the previous subsections), while the non-panic crises are often less well-known crisis events.

We will now exploit the difference between the two types of crises in the chronology of Baron, Verner, and Xiong (2020) to better understand the role of profitablity in the credit-crisis nexus. For holders of short-term debt to panic and withdraw their funds, expectations about bank fundamentals have to be sufficiently negative such that they perceive their stakes to be at risk. Panics can therefore be interpreted as a signal for expectations turning extremely negative. On the other hand, a non-panic crisis indicates that net worth in the financial sector is lost, but there is not a strong reversal of expectations. We will now study how these different crisis types are associated with bank profitability.

As a first pass of the data, Figure 3.6 displays the mean evolution of standardized credit and profit variables in the years around crises events with year 0 indicating the start of a banking crisis. Blue lines correspond to the yearly change in the ratio of credit to GDP. The left panel shows that changes in credit-to-GDP are above average in the years prior to panic crisis events and the ratio of credit to GDP starts declining two years after the panic started. The orange and red line display the evolution of standardized *RoE* and *RoA* around panic crisis observations. The patterns for both are very similar. Banking sector profitability is high and rising until two years before the crisis. In the two years prior to a crisis, there is a reversal with *RoE* and *RoA* being elevated but declining, before they fall below the sample mean once the crisis starts. The right panel in Figure 3.6 presents the same relationship for non-panic crisis. As in the left panel the crisis year coincides with a profitability trough. But the patterns of the decline are strikingly different. Profitability starts declining several years ahead of a crisis. In the same way, the change in credit to GDP is initially high but starts decreasing in the years prior to the crisis. Appendix Figure 3.A.6 shows that these patterns also hold for banking crises after 1945.

This finding is reflected in Table 3.12 where we repeat the specification of Table 3.10 with panic crisis (in $t+3$) as the dependent variable in odd-numbered columns and non-panic equity crises (in $t+3$) in even-numbered columns. While the probability of experiencing a panic in $t+3$ is increasing in profitability, the probability of experiencing a non-panic crisis is, if anything, decreasing in profitability. The results suggest that panics are more likely when profitability booms created room for optimism and subsequent disappointment, while creditors are less surprised by weak performance when the banking sector performance has been weak for some time.

Figure 3.6. Event study of profitability and credit variables around financial crisis dates



Notes: These figures display the evolution of credit and profit variables around a banking crisis, i.e. 0 refers to a year in which a crisis starts. Crises are panic crises in the left panel and non-panic crises in the right panel. Blue lines display the mean of changes in credit/GDP around crises. The orange (red) line displays *RoE* (*RoA*) around crises. All variables have been standardized at the country level.

Table 3.12. Multivariate probit models for systemic financial crisis prediction – panic and non-panic crises

	$\Delta_3 RoE_{i,t}$		$\Delta_3 RoA_{i,t}$		$\Delta_3 LoE_{i,t}$	
	(1) Panic	(2) Non-panic	(3) Panic	(4) Non-panic	(5) Panic	(6) Non-panic
See column header	0.26***	-0.14**	3.07***	-1.50	-0.33***	0.16**
	(0.05)	(0.05)	(0.59)	(1.08)	(0.05)	(0.07)
$\Delta_3 Loans/GDP_{i,t}$	0.18***	0.12***	0.18***	0.13***	0.25***	0.06**
	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)
Country fixed effects	✓	✓	✓	✓	✓	✓
AUC	0.74	0.80	0.73	0.77	0.77	0.93
Observations	1641	668	1647	668	914	354

Notes: The table shows probit classification models. In columns (1), (3) and (5) the dependent variable is an indicator that is one if there is a panic crisis in year $t+3$ and zero else. In columns (2), (4) and (6) the dependent variable is an indicator that is one if there is a non-panic crisis in year $t+3$ and zero else. Coefficients are marginal effects and regressors are described in the column header. All models include country fixed effects. Country clustered standard errors in parentheses. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

3.5.4 Discussion of results

Increases in profitability are associated with subsequent credit expansions. The credit booms that eventually end in banking crises are no exception and are preceded by increases in profitability and low loan losses. This finding mirrors previous evidence in the empirical macro-finance literature. Krishnamurthy and Muir (2017) argue that credit spreads are too low prior to financial crises and Danielsson, Valenzuela, and Zer (2018) show that equity volatility is low. Greenwood, Hanson, and Jin (2019) argue that credit markets often appear to be "calm before the storm". We find similar evidence in measures of bank profitability.

The evolution of profitability around panic crisis events shares some of the key characteristics of behavioral credit cycle models (Bordalo, Gennaioli, and Shleifer, 2018). Panics, associated with sudden changes in expectations, are preceeded by increases in profitability that create room for optimistic beliefs. The cycle turns when profitability starts declining after a series of good realizations. Non-panic crisis on the other hand only mark the final stage of slow-moving profitability declines. They occur when bank fundamentals in previous years left little room for the buildup of optimism and bank net worth has already been depleted.

3.6 Conclusion

The Minsky (1977)-cycle starts with a positive displacement. Positive news breed optimism, and lead to a boom in credit markets, but also to elevated crisis risk down the road. In this paper, we set out to study the origins of this boom, to make sense of the bust.

We establish a new robust fact: bank profitability leads the credit cycle. Credit expands following increases in profitability. Decomposing profitability, we find that loan losses play an important role for this relationship between profits and credit aggregates. Our results are consistent with a recent theoretical literature on the role of expectational biases in shaping the credit cycle. When loan losses are low, economic agents seem to extrapolate these conditions into the future, increasing aggregate leverage in the economy. Similarly, when loan losses are high, banks become more pessimistic and the availability of credit is reduced. We show that reported expectations of bank CFOs from survey data are consistent with such a channel. A caveat of the approach taken in this paper is that we cannot causally identify this link. The long-run evidence presented here should therefore be considered in combination with a growing body of micro-level evidence linking individual experiences to expectation formation and credit market conditions (Landier, Ma, and Thesmar, 2018; Carvalho, Gao, and Ma, 2020). The empirical relationship between profits and credit expansion is also consistent with a financial constraints channel that links profitability and credit expansion. However, we have presented several findings that are inconsistent with this channel being the main explanation for our finding.

The relationship between profits and credit also helps to understand the transition from boom to bust. Bank profitability increases for a few years and peaks two years prior to a crisis. The following reversal in profits and loan losses marks the turning point of the credit cycle and is often associated with a banking panic. Banking crises without panics on the other hand are characterized by decreasing profitability and low credit growth in preceding years. These results suggest that sudden reversals in expectations may indeed be linked to bad profitability news after a sequence of good news. These findings on the differential paths of credit and profitability around panic and non-panic crises may also help to reconcile seemingly contradictory theories of financial crises: the patterns around panic crises seem consistent with the Minsky-view and recent formalizations thereof. Non-panic crises on the other hand are characterized by persistent bank losses, which may result in excessive risk-taking in the financial sector.

We have taken previous findings on firm investment (Greenwood and Hanson, 2015; Gennaioli, Ma, and Shleifer, 2016) to the study of the credit cycle. Is there anything special about credit as an instrument and banks as intermediaries? Simsek (2013) shows that overoptimism of lenders about downside states matters in particular. A similar reasoning leads us to believe that the biases at the bank level may be more important than at the borrower level. If corporate managers extrapolate and become excessively optimistic, but bankers rationally anticipate the growing risks from corporate optimism, then risk would still be priced. This reasoning is also mirrored in recent theoretical contributions stressing the importance of biased expectations of lenders for credit dynamics (Bordalo, Gennaioli, Shleifer, and Terry, 2019; Kaplan, Mitman, and Violante, 2020).

Appendix 3.A Additional results

3.A.1 Summary statistics

Table 3.A.1. Summary statistics

	Obs.	Mean	S.D.	Min	Max
Return on equity	1816	8.59	7.72	-125.36	40.57
Return on assets	1835	0.77	0.79	-7.71	5.27
Capital ratio	1906	10.14	7.86	0.85	46.86
Credit to GDP	1961	57.37	35.22	0.47	204.52
Δ_3 Credit to GDP	1878	3.42	7.90	-56.09	53.08
Δ_3 Capital to GDP	1836	0.45	1.59	-16.20	10.39
Winsorized income data (2.5% level)					
Return on equity	1816	8.93	5.01	-3.97	20.01
Return on assets	1835	0.78	0.61	-0.26	2.54
Dividends over equity	1164	5.54	2.35	1.32	12.38
Retained earnings over equity	1162	3.01	4.53	-10.24	12.90
Revenue over equity	1151	50.53	28.99	8.73	119.01
Cost over equity	1151	31.99	22.08	2.34	85.63
Loan loss over equity	1032	5.93	6.13	0.24	27.79
Δ_3 Return on equity	1751	-0.23	4.63	-13.83	11.28
Δ_3 Return on assets	1772	-0.03	0.38	-1.21	1.00
Δ_3 Dividends over equity	1096	0.01	1.49	-3.96	3.88
Δ_3 Retained earnings over equity	1092	-0.15	4.68	-13.70	12.09
Δ_3 Revenue over equity	1086	-1.13	9.88	-28.07	21.62
Δ_3 Cost over equity	1086	-0.81	6.75	-20.07	14.81
Δ_3 Loan loss over equity	979	0.10	5.45	-15.80	15.19

Notes: All variables in percentage points. World war periods are excluded.

3.A.2 Robustness: main results

Table 3.A.2. Models for changes in credit-to-GDP, subsample of crisis observations

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.56***	0.37**	0.34**			
	(0.15)	(0.14)	(0.14)			
$\Delta_3 RoA_{i,t}$				5.69***	4.54***	3.96**
				(1.81)	(1.52)	(1.73)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.20	0.37	0.42	0.17	0.38	0.42
Observations	176	160	160	176	160	160

Notes: This table reports regressions of credit-to-GDP changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$, where we restrict the sample to up to three observations per financial crisis episode (crisis in $[t-2, t]$). Columns (2), (3), (5) and (6) add a vector of macroeconomic control variables (see text in section 3.3). Columns (3) and (6) additionally control for financial constraints proxies. All specifications include country fixed effects. Standard errors in parentheses are clustered at the country level. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.3. Alternative dependent variable – real private credit per capita

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.79***	0.66***	0.67***			
	(0.23)	(0.18)	(0.18)			
$\Delta_3 RoA_{i,t}$				7.69***	6.73***	7.47***
				(2.31)	(2.01)	(2.00)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.09	0.18	0.19	0.08	0.18	0.18
Observations	1644	1493	1491	1650	1499	1491

Notes: This table reports regressions of real private credit per capita changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$. All specifications control for three lags of real private credit per capita. Columns (2), (3), (5) and (6) add a vector of macroeconomic control variables (see text in section 3.3). Columns (3) and (6) additionally control for financial constraints proxies. All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.4. Alternative dependent variable – bank assets/GDP

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.57**	0.62***	0.61***			
	(0.22)	(0.21)	(0.19)			
$\Delta_3 RoA_{i,t}$				6.42***	5.91***	5.74***
				(2.00)	(1.88)	(2.00)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.10	0.15	0.15	0.10	0.14	0.14
Observations	1650	1504	1504	1651	1505	1504

Notes: This table reports regressions of bank assets/GDP changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$. All specifications control for three lags of bank assets-to-GDP. Columns (2), (3), (5) and (6) add a vector of macroeconomic control variables (see text in section 3.3). Columns (3) and (6) additionally control for financial constraints proxies. All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.5. Alternative dependent variable – non-loan bank assets/GDP

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.18	0.33*	0.33*			
	(0.19)	(0.20)	(0.18)			
$\Delta_3 RoA_{i,t}$				1.96	2.81*	2.77*
				(1.48)	(1.63)	(1.65)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.07	0.12	0.12	0.07	0.12	0.12
Observations	1617	1473	1473	1618	1474	1473

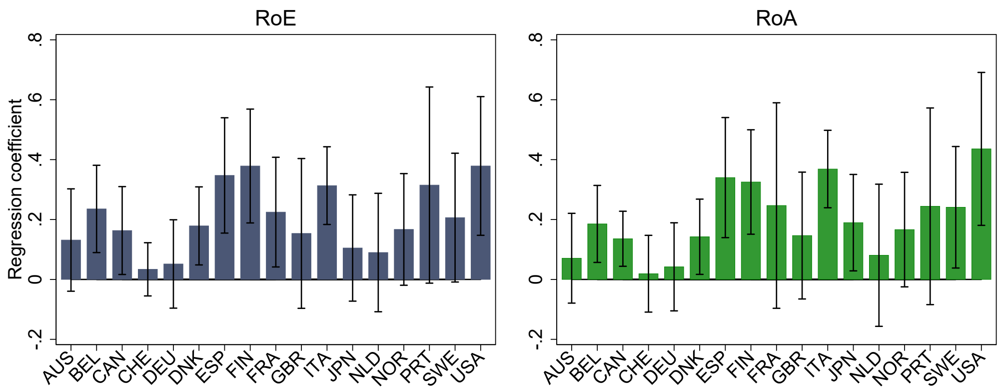
Notes: This table reports regressions of non-loan bank assets/GDP changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$. All specifications control for three lags of non-loan bank assets/GDP. Columns (2), (3), (5) and (6) add a vector of macroeconomic control variables (see text in section 3.3). Columns (3) and (6) additionally control for financial constraints proxies. All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.6. Alternative profitability measures – profits/GDP and log real profits per capita

	Dependent variable: $\Delta_3 y_{i,t+3}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ_3*Profits to GDP*$_{i,t}$	2.08***	1.64***	1.61***			
	(0.71)	(0.58)	(0.57)			
Δ_3*Log*(*profits*)$_{i,t}$				0.03***	0.03***	0.03***
				(0.00)	(0.00)	(0.00)
Country fixed effects	✓	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓	✓
Macrocontrols		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.10	0.19	0.19	0.13	0.20	0.21
Observations	1635	1491	1491	1512	1372	1372

Notes: This table reports regressions of credit-to-GDP changes from t to $t+3$ on Δ_3*Profits to GDP*$_{i,t}$ and Δ_3*log*(*profits*)$_{i,t}$. *Log*(*profits*) is the logarithm of real profits per capita. All specifications control for three lags of credit-to-GDP changes. Columns (2) and (5) add a vector of macroeconomic control variables (see text in section 3.3). Columns (3) and (6) additionally control for financial constraints proxies. All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Figure 3.A.1. Country-level regression coefficients



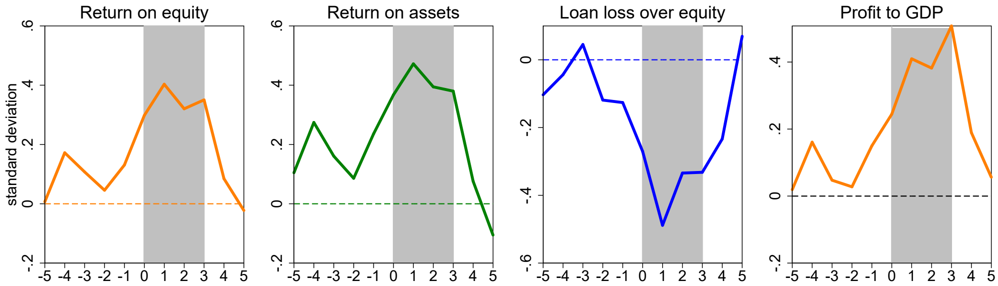
Notes: This figure reports regression coefficients and 90% confidence intervals from individual country regressions of credit-to-GDP changes from t to $t+3$ on $\Delta_3 RoE_{i,t}$ and $\Delta_3 RoA_{i,t}$. The specifications $\Delta_3 y_{t+3} = \alpha + \beta\Delta_3 RoE_t + u_{t+3}$ and $\Delta_3 y_{t+3} = \alpha + \beta\Delta_3 RoA_t + u_{t+3}$ are estimated on individual country subsamples. Variables have been standardized by country for comparability of coefficients.

Table 3.A.7. Alternative profitability measure – level variables

	Dependent variable: $\Delta_3 y_{i,t+3}$				
	(1)	(2)	(3)	(4)	(5)
$RoE_{i,t}$	0.50***				
	(0.08)				
$RoA_{i,t}$		5.44***			
		(1.13)			
$Profits\ to\ GDP_{i,t}$			2.51***		
			(0.65)		
$Log(profits)_{i,t}$				0.02***	
				(0.00)	
$LoanLoss/Equity_{i,t}$					-0.52***
					(0.09)
Country fixed effects	✓	✓	✓	✓	✓
Distributed lag in Δy	✓	✓	✓	✓	✓
Macrocontrols	✓	✓	✓	✓	✓
Financial constraints	✓	✓	✓	✓	✓
R^2	0.24	0.23	0.21	0.21	0.28
Observations	1516	1516	1516	1444	935

Notes: This table reports regressions of credit-to-GDP changes from t to $t+3$ on levels of profitability. All specifications control for three lags of credit-to-GDP changes, macroeconomic control variables (see text in section 3.3) and financial constraints proxies. All specifications include country fixed effects. Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

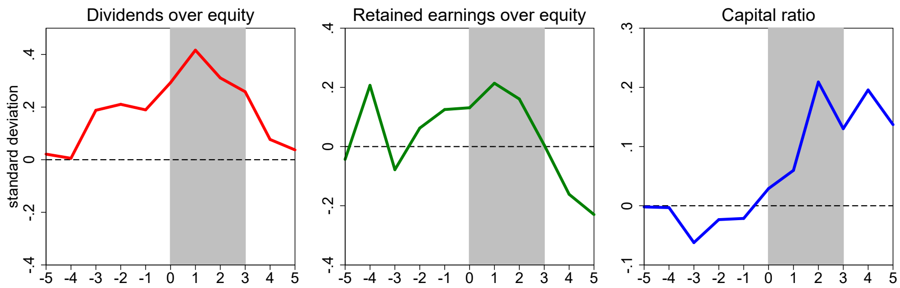
Figure 3.A.2. Event study of profitability around credit boom dates after 1945



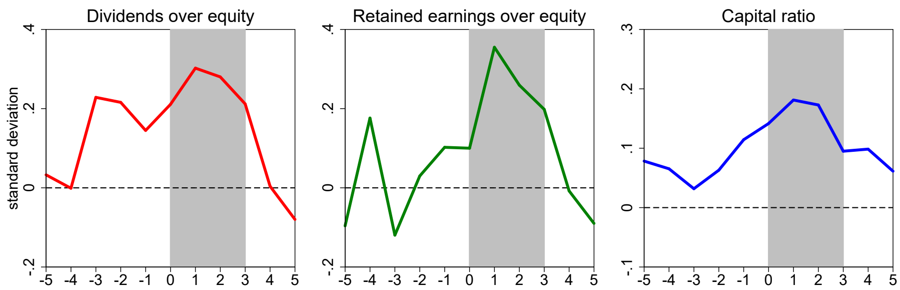
Notes: These figures display the evolution of profit variables around credit booms. All variables are detrended and standardized with mean zero and standard deviation one by country. Observations are classified as boom years when $\Delta_3 Loans/GDP_{i,t}$ exceeds one standard deviation. 0 refers to a year in which a credit boom starts. The grey area marks the three-year window of the credit boom. Solid lines display means of variables in the header around booms.

Figure 3.A.3. Event study of profitability around credit boom dates – additional variables

(a) Full sample



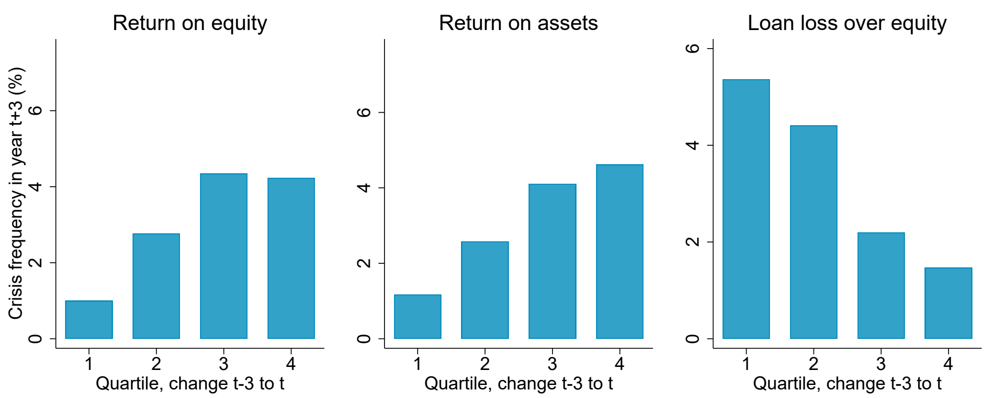
(b) After 1945



Notes: These figures display the evolution of profit variables around credit booms. All variables are detrended and standardized with mean zero and standard deviation one by country. Observations are classified as boom years when $\Delta_3 Loans/GDP_{i,t}$ exceeds one standard deviation. 0 refers to a year in which a credit boom starts. The grey area marks the three-year window of the credit boom. Solid lines display means around booms.

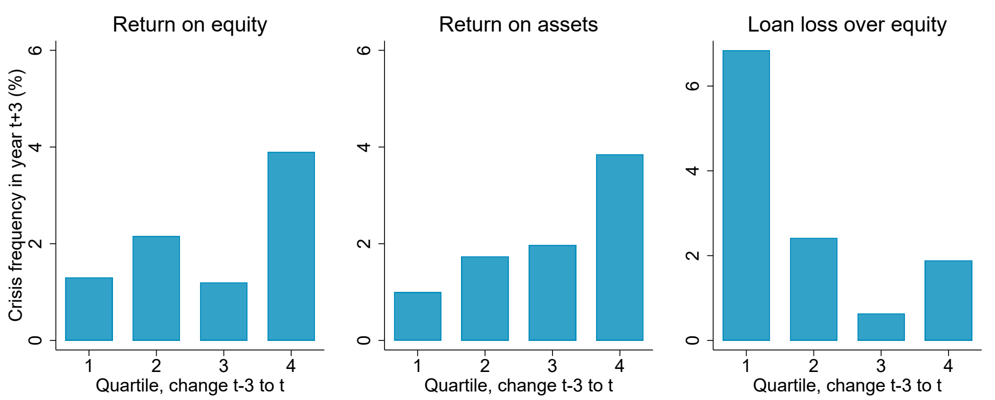
3.A.3 Robustness: profitability around financial crises

Figure 3.A.4. Crisis probability in t+3 by change in profitability – BVX panic banking crisis dates



Notes: This figure shows the relationship between changes in RoE (RoA) between $t-3$ and t and banking panic (Baron, Verner, and Xiong, 2020) frequencies for the year $t+3$. Observations are sorted into four equal-sized bins according to the increase in RoE (RoA) between $t-3$ and t. Vertical bars indicate the frequency of financial crises in year $t+3$ for each of the bins.

Figure 3.A.5. Crisis probability in t+3 by change in profitability – post WW2 sample



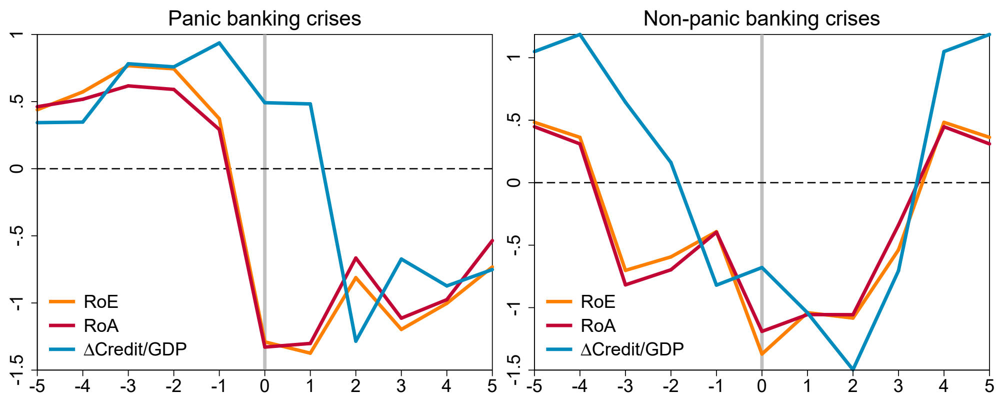
Notes: This figure shows the relationship between changes in RoE (RoA) between $t-3$ and t and financial crisis frequencies (Jordà, Schularick, and Taylor (2017)-chronology) for the year $t+3$. Observations are sorted into four equal-sized bins according to the increase in RoE (RoA) between $t-3$ and t. Vertical bars indicate the frequency of financial crises in year $t+3$ for each of the bins.

Table 3.A.8. Multivariate probit models for systemic financial crisis prediction – BVX panic banking crises

	$\Delta_3 RoE_{i,t}$		$\Delta_3 RoA_{i,t}$		$\Delta_3 LoE_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
Profitability (see column header)	0.25***	0.26***	2.92***	3.07***	-0.25***	-0.33***
	(0.05)	(0.05)	(0.68)	(0.59)	(0.07)	(0.05)
$\Delta_3 Loans/GDP_{i,t}$		0.18***		0.18***		0.25***
		(0.03)		(0.03)		(0.03)
AUC	0.67	0.74	0.67	0.73	0.66	0.77
Observations	1700	1641	1721	1647	916	914

Notes: The table shows probit classification models where the dependent variable is an indicator that is one if the country experiences a banking panic (Baron, Verner, and Xiong, 2020) in year $t + 3$ and zero else. Coefficients are marginal effects. Regressors are described in the column header. All models include country fixed effects. Country clustered standard errors in parentheses. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Figure 3.A.6. Event study of profitability and credit variables around financial crisis dates – post 1945



Notes: These figures display the evolution of credit and profit variables around a banking crisis after 1945, i.e. 0 refers to a year in which a crisis starts. Crises are panic crises in the left panel and non-panic crises in the right panel. Blue lines display the mean of changes in credit/GDP around crises. The orange (red) line displays RoE (RoA) around crises. All variables have been standardized at the country level.

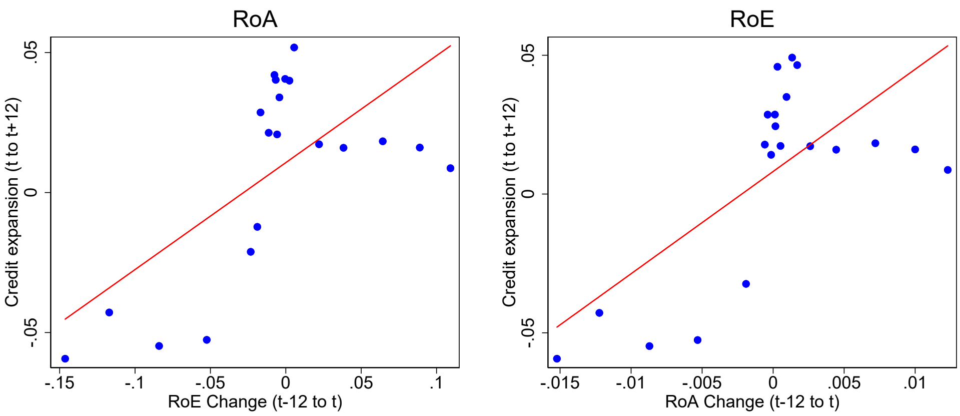
Table 3.A.9. Multivariate probit models for systemic financial crisis prediction – BVX panic banking crisis dates

	Dependent variable: Crisis at time...					
	(1)	(2)	(3)	(4)	(5)	(6)
	t	t+1	t+2	t+3	t+4	t+5
$\Delta_3 RoE_{i,t}$	-0.23***	0.10	0.25***	0.31***	0.20***	-0.00
	(0.05)	(0.08)	(0.08)	(0.06)	(0.07)	(0.09)
$\delta\Delta_3 RoE_{i,t}$	-0.16**	-0.12	-0.20***	-0.07	0.03	0.10
	(0.06)	(0.10)	(0.05)	(0.06)	(0.08)	(0.07)
$\Delta_3 Loans/GDP_{i,t}$	0.11***	0.21***	0.19***	0.17***	0.11***	0.04
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
AUC	0.86	0.75	0.75	0.75	0.68	0.63
Observations	1667	1650	1633	1616	1599	1582
	Dependent variable: Crisis at time...					
	(1)	(2)	(3)	(4)	(5)	(6)
	t	t+1	t+2	t+3	t+4	t+5
$\Delta_3 RoA_{i,t}$	-2.66***	1.24	3.24***	3.65***	2.30**	1.00
	(0.52)	(0.88)	(0.70)	(0.74)	(0.92)	(1.11)
$\delta\Delta_3 RoA_{i,t}$	-2.08***	-1.51	-2.00***	-1.03	0.32	-0.81
	(0.66)	(1.05)	(0.67)	(0.84)	(0.95)	(0.86)
$\Delta_3 Loans/GDP_{i,t}$	0.12***	0.21***	0.19***	0.17***	0.11***	0.04
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
AUC	0.85	0.75	0.75	0.74	0.68	0.63
Observations	1675	1658	1641	1624	1607	1590

Notes: The table shows probit classification models where the dependent variable is an indicator that is one if there is a banking panic in $t + h$ years, specified in the column header. Coefficients are marginal effects. All specifications include country-fixed effects. Country clustered standard errors in parentheses. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

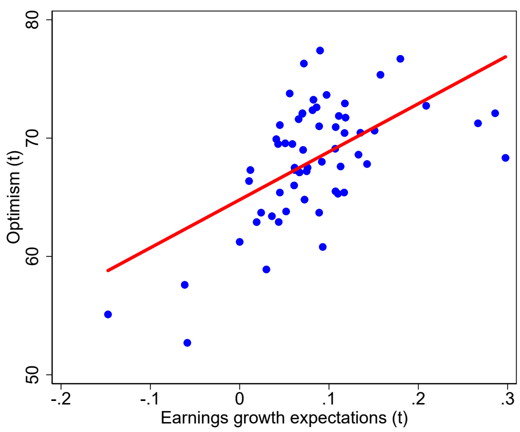
3.A.4 Robustness: survey on earnings expectations

Figure 3.A.7. Confirmation of main result: the profit-credit cycle in quarterly US data



Notes: The figure relates bank profitability and subsequent three-year changes in credit to GDP. Observations are collapsed into 20 equal sized bins according to their profitability (or changes therein). Each point represents the group specific means of profitability and credit expansion. Fitted regression lines illustrate the correlation between bank profitability and subsequent credit growth.

Figure 3.A.8. Earnings growth expectations and bank CFO optimism



Notes: The figure shows the relationship between bank CFO optimism and bank CFO earnings growth expectations. Fitted regression lines illustrate the correlation between the two variables.

Table 3.A.10. Relationship between profitability, expectations about future profitability and credit supply conditions

	$\Delta Optimism$	ΔRoA_{t+4}	$\Delta E_t(RoA_{t+4})$	$\Delta Error$	$\Delta\%Tightening$
	(1)	(2)	(3)	(4)	(5)
ΔRoA_t	16.95***	0.13	0.79***	-0.65***	-72.68***
	(4.15)	(0.14)	(0.19)	(0.22)	(10.34)
R^2	0.08	0.01	0.18	0.10	0.17
Observations	57	78	73	69	82

Notes: This table reports estimates for univariate regressions of expectation measures on the change in *RoA*. In column (1), the dependent variable is the quarterly change in optimism from the bank CFO survey, in column (2) the quarterly change in realized earnings between t and t+4 normalized with assets at time t, in column (3) the quarterly change in expected earnings between t and t+4 normalized with assets at time t, in column (4) the quarterly change in the difference between realized and expected earnings between t and t+4, and in column (5) the change in the net percentage of domestic banks tightening standards for C&I loans to large and middle-market firms. Newey-West standard errors in parentheses are computed using the automatic bandwidth selection procedure in Newey and West (1994). *, **, ***: Significant at 10%, 5% and 1% levels respectively.

Table 3.A.11. Relationship between profitability, expectations about future profitability and credit supply conditions, excluding the years 2007 – 2009

	$\Delta Optimism$	ΔRoE_{t+4}	$\Delta E_t(RoE_{t+4})$	$\Delta Error$	$\Delta\%Tightening$
	(1)	(2)	(3)	(4)	(5)
ΔRoE_t	-0.79	0.02	0.15***	-0.12***	-8.32*
	(1.82)	(0.02)	(0.02)	(0.04)	(4.37)
R^2	0.00	0.02	0.28	0.18	0.08
Observations	45	66	61	57	70

Notes: This table reports estimates for univariate regressions of expectation measures on the change in *RoE*. In column (1), the dependent variable is the quarterly change in optimism from the bank CFO survey, in column (2) the quarterly change in realized earnings between t and t+4 normalized with assets at time t, in column (3) the quarterly change in expected earnings between t and t+4 normalized with assets at time t, in column (4) the quarterly change in the difference between realized and expected earnings between t and t+4, and in column (5) the change in the net percentage of domestic banks tightening standards for C&I loans to large and middle-market firms. Newey-West standard errors in parentheses are computed using the automatic bandwidth selection procedure in Newey and West (1994). *, **, ***: Significant at 10%, 5% and 1% levels respectively.

3.A.5 Timing

This section extends the baseline setup and describes the dynamic relationship between profitability measures and changes in credit-to-GDP over varying 3-year windows (similar to Mian, Sufi, and Verner, 2017). In the following equation, the RHS of the equation is held constant, while we shift the dependent variable $\Delta_3 y_{i,t+k}$ in time:

$$\Delta_3 y_{i,t+k} \ = \ \alpha_i + \beta \Delta_3 RoE_{i,t} + \eta X_{i,t} + \theta Z_{i,t} + u_{i,t+k} \qquad (3.A.1)$$

where $k = 0, ..., 6$. The results are shown in Table 3.A.12. Column (1) ($k = 0$) assesses the contemporaneous relationship between changes in profitability from $t-3$ to t and the change in the credit-to-GDP ratio between $t-3$ and t. In subsequent columns we report the results for a shift of the dependent variable one year further into the future. Column (4) ($k = 3$) is therefore equivalent to our baseline specification. We include the full set of controls except for the three lags of $\Delta y_{i,t}$ (for $k = 0$ the dependent variable is a linear combination of these).

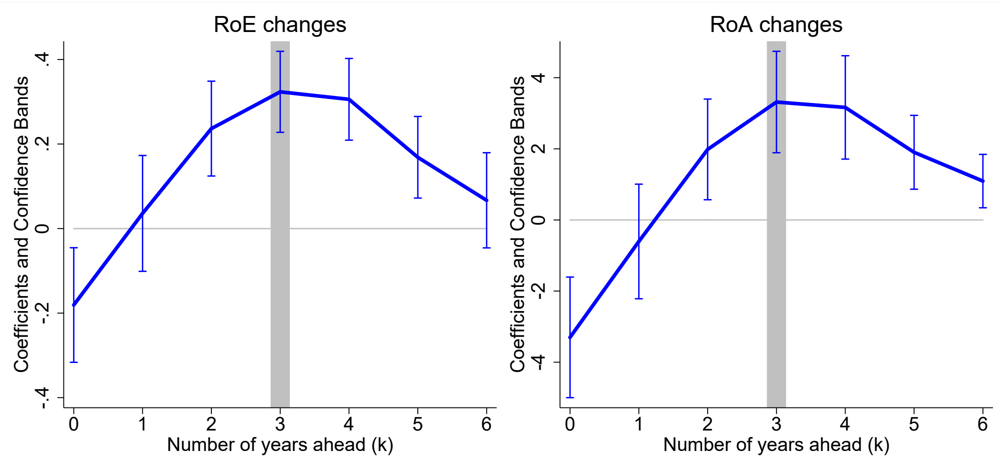
The results in column (1) show that changes in credit-to-GDP and *RoE* are contemporaneously negatively correlated. Importantly, the relationship is reversed in the medium run: in column (4) ($k = 3$) we see that changes in *RoE* between $t-3$ and t are positively associated with credit growth between t and $t+3$. The effect is strongest for $k = 3$ and $k = 4$ and the coefficients become smaller for larger k. The lower panel of Table 3.A.12 shows the equivalent relationship for $\Delta_3 RoA_{i,t}$. The size of the coefficient peaks at $k = 3$ and decays afterwards, much like the $\Delta_3 RoE$ results.

The dynamic relationship between profitability and credit displays a particular pattern: a "profit-credit cycle". This relationship is visualized in Figure 3.A.9 with changes in return on equity in the left panel and changes in return on assets in the right panel. Both figures show an inverted u-shaped relationship, that is, the response of the credit-to-GDP ratio to variation in profitability measures is strongest over the subsequent three years. This timing is difficult to square with with credit demand explanations. If credit demand was the driver of the relationship, we would have expected to observe increases in credit-to-GDP against good current and future prospects. In that case changes in profitability and credit growth should display a positive contemporaneous correlation or, if households and firms borrow against anticipated good future fundamentals, credit expansion should lead profitability. We find the opposite.

3.A.6 Bank level dataset

To supplement our long run aggregate evidence with bank level results, we employ bank call report data provided by the Federal Reserve. Banks are required to file these reports for regulatory purposes and the data contain detailed quarterly income and balance sheet statements for all US commercial banks. We use data between 1983 and 2012, when all balance sheet and income statement items for our analysis

Figure 3.A.9. Multivariate models for changes in credit-to-GDP, dynamic relationship



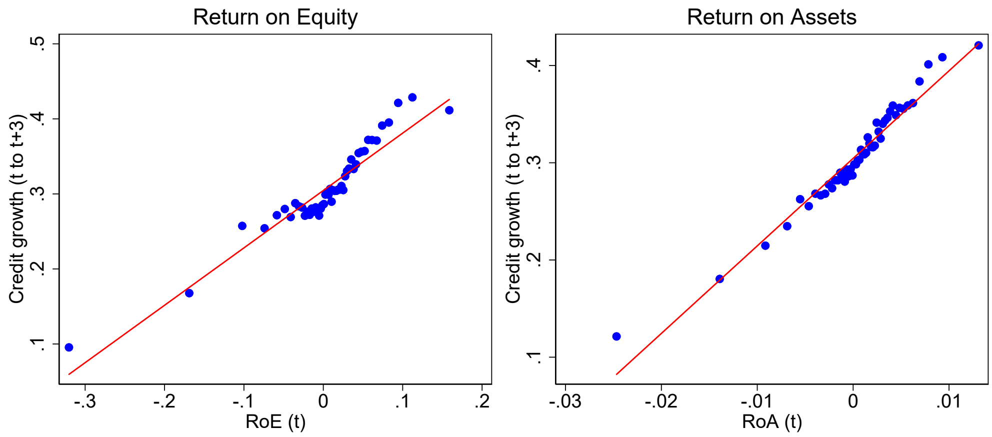
Notes: This figure displays coefficients from estimating Equation 3.A.1 for $k = 0, ..., 6$. See Table 3.A.12 for more information. Standard errors are dually clustered on country and year. Bars denote 95% confidence intervals around the coefficient estimates.

Table 3.A.12. Multivariate models for changes in credit-to-GDP, dynamic relationship

	Dependent variable: $\Delta_3 y_{i,t+k}$, $k = 0, ..., 5$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\Delta_3 y_{i,t}$	$\Delta_3 y_{i,t+1}$	$\Delta_3 y_{i,t+2}$	$\Delta_3 y_{i,t+3}$	$\Delta_3 y_{i,t+4}$	$\Delta_3 y_{i,t+5}$	$\Delta_3 y_{i,t+6}$
$\Delta_3 RoE_{i,t}$	-0.18***	0.04	0.24***	0.32***	0.31***	0.17***	0.07
	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓	✓
R^2	0.24	0.19	0.16	0.13	0.10	0.06	0.05
Observations	1526	1531	1518	1504	1490	1475	1458
	Dependent variable: $\Delta_3 y_{i,t+k}$, $k = 0, ..., 5$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$\Delta_3 y_{i,t}$	$\Delta_3 y_{i,t+1}$	$\Delta_3 y_{i,t+2}$	$\Delta_3 y_{i,t+3}$	$\Delta_3 y_{i,t+4}$	$\Delta_3 y_{i,t+5}$	$\Delta_3 y_{i,t+6}$
$\Delta_3 RoA_{i,t}$	-3.30***	-0.61	1.98***	3.32***	3.16***	1.90***	1.09***
	(0.87)	(0.82)	(0.72)	(0.73)	(0.74)	(0.53)	(0.38)
Country fixed effects	✓	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓	✓
R^2	0.26	0.19	0.15	0.12	0.09	0.06	0.05
Observations	1526	1531	1518	1504	1490	1475	1458

Notes: This table presents results from estimating Equation 3.A.1 for $k =, 0, ..., 6$. Each column gradually leads the left-hand-side variable by one year. All specifications control for a vector of net-worth and macroeconomic control variables (see text in section 3.3). Standard errors in parentheses are dually clustered on country and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively

Figure 3.A.10. Binned scatterplot for the relationship between profitability and credit growth, bank level data



Notes: The figure relates bank profitability and subsequent credit growth on a bank level. Bank level observations are collapsed into 50 equal sized bins according to the two profitability measures. Each point represents group specific profitability and credit growth means for our regression sample. Fitted regression lines illustrate the correlation between bank profitability and subsequent credit growth.

are available in the same format. We first transform quarterly call report data into annual observations, by summing income items over the four quarters of a given year. We then combine yearly income with end-of-year balance sheet values. We exclude bank-year observations with assets or loans being less than one million USD, or with negative equity, and we winsorize all variables at the 2.5% level.

The resulting panel dataset with bank-year observations allows us to run specifications mirroring closely the empirical exercises of aggregate setting. The dependent variable is defined as the change in *net loans and leases* of bank i between year t and year $t+3$. $RoE_{i,t}$ is defined as yearly net income scaled by end-of-year equity. As before, we also compute the three-year change in this variable $\Delta_3 RoE_{i,t} = RoE_{i,t} - RoE_{i,t-3}$.

Figure 3.A.10 shows scatterplots with the data collapsed into fifty bins, depending on profitability measures. There is a strong positive correlation between the profitability of individual banks ($RoE_{i,t}$ and $RoA_{i,t}$) and their subsequent credit growth. In order to test this relationship more formally, we run the following regression:

$$\Delta_3 y_{i,t+3} = \alpha_i + \alpha_t + \beta \Delta_3 RoE_{i,t} + \gamma X_{i,t} + u_{i,t+3}. \tag{3.A.2}$$

Crucially, this regression includes a year fixed effect α_t to absorb aggregate credit demand conditions at time t. α_i is a bank fixed effect that controls for bank specific time-invariant characteristics. β will be the coefficient of interest that refers to the three-year change in profitability ($\Delta_3 RoE_{i,t}$ or $\Delta_3 RoA_{i,t}$). Control variables $X_{i,t}$ are now at the bank level. We include past credit growth, and in addition lagged balance sheet shares of equity, loans, deposits, fed funds (liabilities) and bank size (natural log of assets). Three-year changes in capital proxy for the net worth channel. One advantage in this setup is that we can control for net-worth at the bank level and rule out balance sheet constraints more directly, accounting for the possibility that the distribution of net worth and leverage across banks matters.

The results are shown in Table 3.A.13. Columns (1) and (4) only include bank and year fixed effects, the subsequent columns add a rich set of bank level controls for bank asset and liability composition and changes in bank net-worth. Across specifications, credit growth over the following 3-year window is signiciantly higher when profitability has been increasing. In line with a net-worth channel, three-year changes in equity capital are associated with elevated subsequent loan growth. Table 3.A.14 shows that these results are robust when using non-overlapping observations only. Importantly, the bank level results are not affected by the inclusion of time fixed effects. The channel that links profits and subsequent credit growth is not contingent on or subsumed by aggregate credit demand.

Table 3.A.15 and Table 3.A.16 replicate two other key results from the aggregate analysis at the bank level. Table 3.A.15 shows regression evidence for the three major profit components revenue, operating expenses and loan losses mirroring the analysis in Table 3.3. Again, the profit-credit relationship is largely coming from the loan loss component of banking income. However, bank level operating expenses also show a significant, albeit weaker, association with subsequent credit growth. Table 3.A.16 replicates Table 3.6 at the bank level. Column (1) and (2) separate return on equity into a dividend over equity and a retained earnings over equity component and show that both components predict subsequent credit growth at the bank level. Column (3) and (4) include levels and changes of profitability. As argued before, controlling for the level of *RoE*, three-year changes proxy for the trajectory that led a bank to a certain level of profitability. In line with the expectations channel, changes in *RoE* are significantly related to subsequent credit growth.

Table 3.A.13. Multivariate models for credit growth, bank level data

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.48***	0.32***	0.29***			
	(0.04)	(0.03)	(0.03)			
$\Delta_3 RoA_{i,t}$				5.05***	3.39***	2.86***
				(0.49)	(0.37)	(0.34)
Bank fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Control variables		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.08	0.20	0.21	0.07	0.20	0.21
Observations	192579	192579	192579	192579	192579	192579

Notes: This table reports regression results from estimating variants of Equation 3.A.2 using US Call Report data. The dependent variable $\Delta_3 y_{i,t+3}$ is the three year growth of bank credit (net loans and leases). All variables are winsorized at the 2.5% level. All specifications control for the lagged three-year growth rate of net loans and leases, balance sheet ratios, bank size and financial constraints (see text). All specifications also include bank and year fixed effects. Standard errors in parentheses are dually clustered on bank and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.14. Multivariate models for credit growth, bank level data, non-overlapping observations

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta_3 RoE_{i,t}$	0.48***	0.33***	0.30***			
	(0.07)	(0.06)	(0.05)			
$\Delta_3 RoA_{i,t}$				4.96***	3.49***	3.05***
				(0.83)	(0.66)	(0.62)
Bank fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Control variables		✓	✓		✓	✓
Financial constraints			✓			✓
R^2	0.08	0.21	0.21	0.08	0.21	0.21
Observations	59043	59043	59043	59043	59043	59043

Notes: This table reports regression results from estimating variants of Equation 3.A.2 using US Call Report data. The dependent variable $\Delta_3 y_{i,t+3}$ is the three year growth of bank credit (net loans and leases). All variables are winsorized at the 2.5% level. All specifications control for the lagged three-year growth rate of net loans and leases, balance sheet ratios, bank size and financial constraints (see text). All specifications also include bank and year fixed effects. Standard errors in parentheses are dually clustered on bank and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.15. Multivariate models for credit growth, bank level data

	(1) $\frac{Revenue}{Equity}$	(2) $\frac{Costs}{Equity}$	(3) $\frac{LoanLosses}{Equity}$	(4) $\frac{Revenue}{Assets}$	(5) $\frac{Costs}{Assets}$	(6) $\frac{LoanLoss}{Assets}$
$\Delta_3 Change_{i,t}$	-0.01	-0.13***	-0.41***	-0.47	-1.69***	-5.40***
	(0.02)	(0.02)	(0.05)	(0.30)	(0.31)	(0.59)
Bank fixed effects	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓
R^2	0.21	0.21	0.21	0.21	0.21	0.21
Observations	179072	179072	179072	179072	179072	179072

Notes: The dependent variable $\Delta_3 y_{i,t+3}$ is the three year growth of bank credit (net loans and leases). All variables are winsorized at the 2.5% level. All specifications control for the lagged three-year growth rate of net loans and leases, balance sheet ratios, bank size and financial constraints (see text). All specifications also include bank and year fixed effects. Standard errors in parentheses are dually clustered on bank and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

Table 3.A.16. Multivariate models for credit growth, bank level data, dividend decomposition and path

	Uses of profits		Profit path	
			RoE	RoA
	(1)	(2)	(3)	(4)
Δ_3*Dividends over Equity*$_{i,t}$	0.12***	0.41***		
	(0.03)	(0.05)		
Δ_3*Retained earnings over Equity*$_{i,t}$		0.39***		
		(0.04)		
$3-$ *year Accumulated Profits*$_{i,t}$			0.01	0.28
			(0.01)	(0.18)
$\Delta_3 Change_{i,t}$			0.28***	2.75***
			(0.03)	(0.34)
Bank fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Control variables	✓	✓	✓	✓
R^2	0.24	0.25	0.21	0.21
Observations	75241	75241	192402	192402

Notes: The dependent variable $\Delta_3 y_{i,t+3}$ is the three year growth of bank credit (net loans and leases). All variables are winsorized at the 2.5% level. All specifications control for the lagged three-year growth rate of net loans and leases, balance sheet ratios, bank size and financial constraints (see text). All specifications also include bank and year fixed effects. Standard errors in parentheses are dually clustered on bank and year. *,**,*** indicates significance at the 0.1, 0.05, 0.01 level, respectively.

3.A.7 Systemic banking crises

Dates of systemic banking crises are based on Jordà, Schularick, and Taylor (2017).

AUS: 1893, 1989.
BEL: 1870, 1885, 1925, 1931, 1934, 1939, 2008.
CAN: 1907.
CHE: 1870, 1910, 1931, 1991, 2008.
DEU: 1873, 1891, 1901, 1907, 1931, 2008.
DNK: 1877, 1885, 1908, 1921, 1931, 1987, 2008.
ESP: 1883, 1890, 1913, 1920, 1924, 1931, 1978, 2008.
FIN: 1878, 1900, 1921, 1931, 1991.
FRA: 1882, 1889, 1930, 2008.
GBR: 1890, 1974, 1991, 2007.
ITA: 1873, 1887, 1893, 1907, 1921, 1930, 1935, 1990, 2008.
JPN: 1871, 1890, 1907, 1920, 1927, 1997.
NLD: 1893, 1907, 1921, 1939, 2008.
NOR: 1899, 1922, 1931, 1988.
PRT: 1890, 1920, 1923, 1931, 2008.
SWE: 1878, 1907, 1922, 1931, 1991, 2008.
USA: 1873, 1893, 1907, 1929, 1984, 2007.

Appendix 3.B Data appendix

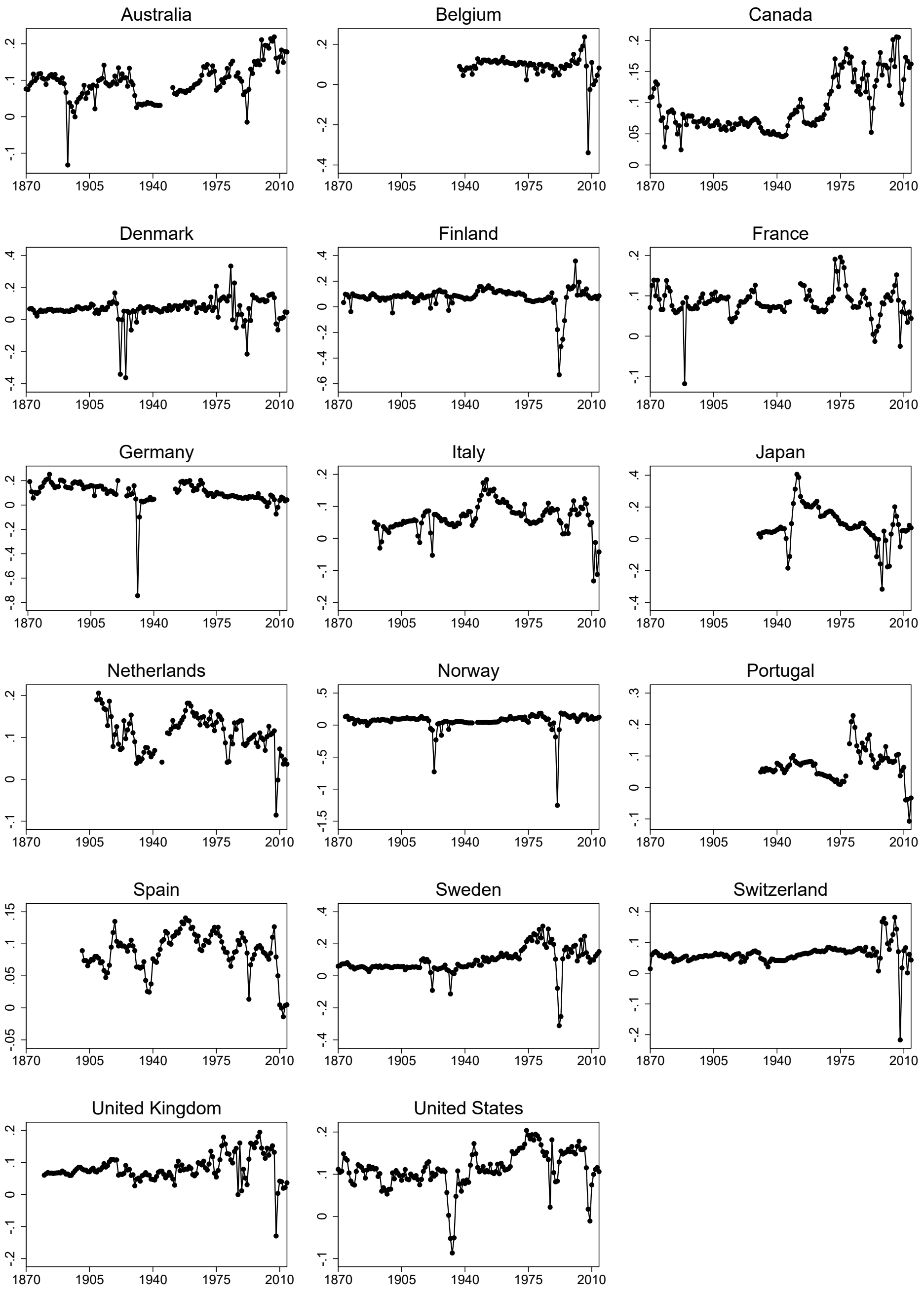
This appendix details the sources of our banking sector profitability estimates for each country. The data contains aggregate profitability series for the banking system and decomposes this profitability into its sources. It includes separate time series for bank return on assets and its main components - revenue (net interest income + net fee income), operating expenses and loan losses. All variables are constructed relative to total assets of the financial system. Items are then rescaled using leverage data from Jordà, Richter, Schularick, and Taylor (2021) (JRST henceforth). We use end of year total capital and total liabilities as denominators in the calculation.

Table 3.B.1. Variable definitions

Item	Description
Return on equity	After tax profitability of the banking system relative to end of year equity.
Return on assets	After tax profitability of the banking system relative to end of year assets.
Dividends	Total dividends of the banking system relative to end of year assets.
Costs	Operating expenses of the banking system relative to end of year assets.
Revenues	Total revenue (net interest and fee income) relative to end of year assets.
Loan losses	Loan loss item in the bank income statement relative to end of year assets (charge-offs or provisions for charge-offs).

Our primary goal in constructing the series is consistency across series and within country. We use growth rate splicing if there are significant inconsistencies across sources and coverage, but aim to keep original data levels as much as possible. Maintaining original levels has the advantage that it allows for an bias free construction of ratios and manipulations of the individual series (for example when considering the revenue to cost relationship). We sometimes use profit and loss accounts of individual banks to extend the aggregate series back in time. This data typically relies on the largest banks in a given country. Since we choose the banks based on their historic dominance and not based on their recent success or the survival until today, a potential survivorship bias is unlikely to be large. Finally, the sophistication of accounting standards and practice varied significantly historically. We adjust the data whenever we find the appropriate means to do so. For example, Capie and Billings (2001) provide us with an updated series of banking sector profitability in the United Kingdom that accounts for transactions that involved hidden reserves in the balance sheet. Figure 3.B.1 displays the main profitability series – return on equity – on a country by country basis.

Figure 3.B.1. Return on equity



Australia

Table 3.B.2. Data sources: Australia

Year	Data source
Bank profitability	
1870–1944	Butlin, Hall and White (1971). Australian banking and monetary statistics, 1817-1945. Reserve Bank of Australia Occasional Paper No. 4A.
1946–1970	White (1973). Australian banking and monetary statistics 1945-1970. Reserve Bank of Australia Occasional Paper No. 4B. Major trading banks.
1971–1980	Statistical Yearbook (various years). Data for joint stock banks.
1981–2001	OECD Banking Statistics. Income statement and balance sheet.
2002–2003	Annual Reports of the four major banks (various years): ANZ, NAB, Commonwealth Bank and Westpac.
2004–2015	Australian Prudential Regulation Authority (2016). Quarterly ADI performance statistics.
Bank P&L components	
1946–1970	White (1973). Australian banking and monetary statistics 1945-1970. Reserve Bank of Australia Occasional Paper No. 4B. Major trading banks.
1963–1974	Statistical Yearbook (various years). Data for joint stock banks.
1981–2001	OECD Banking Statistics. Income statement and balance sheet.
2004–2015	Australian Prudential Regulation Authority (2016). Quarterly ADI performance statistics.
Bank dividends	
1870–1944	Butlin, Hall and White (1971). Australian banking and monetary statistics, 1817-1945. Reserve Bank of Australia Occasional Paper No. 4A.
1946–1974	White (1973). Australian banking and monetary statistics 1945-1970. Reserve Bank of Australia Occasional Paper No. 4B. Major trading banks.
1981–2001	OECD Banking Statistics. Income statement and balance sheet.

Belgium

Table 3.B.3. Data sources: Belgium

Year	Data source
Bank profitability	
1937–1980	Rapport Annuel de la Commission Bancaire (various years). All banks for 1944 to 1980 and large banks for 1937 to 1943.
1983–1999	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2000–2017	National Bank of Belgium (various years). Financial Stability Report. All credit institutions.
Bank P&L components	
1981–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2017	National Bank of Belgium (various years). Financial Stability Report. All credit institutions.
Bank dividends	
1981–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.

Canada

Table 3.B.4. Data sources: Canada

Year	Data source
Bank profitability	
1870–1967	Annual Reports of major banks (various years): Bank of Montreal, Scotiabank, Canadian Bank of Commerce, Royal Bank of Canada, Bank of Toronto, Dominion Bank, Toronto Dominion Bank (after merger).
1968–1981	Bank of Canada Review (various years). Table A4 of the February or March issue.
1982–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2009–2015	Canadian Bankers Association. Database of Domestic Banks' Financial Results. Fiscal year-end 2006-2015, 8 banks.
Bank P&L components	
1929–1967	Historical Statistics of Canada. Link: https://www150.statcan.gc.ca/n1/pub/11-516-x/3000140-eng.htm. Tables J181-201 and J261-272.
1968–1981	Bank of Canada Review (various years). Table A4 of the February or March issue.
1982–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	Canadian Bankers Association. Database of Domestic Banks' Financial Results. Fiscal year-end 2006-2015, 8 banks.
Bank dividends	
1870–1963	Annual Reports of major banks (various years): Bank of Montreal, Scotiabank, Canadian Bank of Commerce, Royal Bank of Canada, Bank of Toronto, Dominion Bank, Toronto Dominion Bank (after merger).
1964–1967	Historical Statistics of Canada. Link: https://www150.statcan.gc.ca/n1/pub/11-516-x/3000140-eng.htm. Tables J181-201 and J261-272.
1968–1987	Bank of Canada Review (various years). Table A4 of the February or March issue.
1988–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	Canadian Bankers Association. Database of Domestic Banks' Financial Results. Fiscal year-end 2006-2015, 8 banks.

Denmark

Table 3.B.5. Data sources: Denmark

Year	Data source
Bank profitability	
1872–1920	Danmarks Statistik (1969). Statistike Underslogelser Nr. 24 Kreditmarkedsstatistik. Link: http://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=19918&sid=kreditm. Table: Bankernes samlede status inden for hovedlandsdele og for hele landet.
1921–1985	Statistical Yearbook (various years).
1986–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	Finansrådet (2015). The sector in figures. Table: Accounting figures.
Bank P&L components	
1875–1920	Abildgren (2017). A chart & data book on the monetary and financial history of Denmark. Working Paper. Sheet S081A
1920–1978	Statistical Yearbook (various years).
1979–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010-2015	Finansrådet (2015). The sector in figures. Table: Accounting figures.
Bank dividends	
1872–1920	Danmarks Statistik (1969). Statistike Underslogelser Nr. 24 Kreditmarkedsstatistik. Link: http://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=19918&sid=kreditm. Table: Bankernes samlede status inden for hovedlandsdele og for hele landet.
1921–1978	Beretning om de danske bankers virksomhed (various years). Official government publication with statistics on all commercial banks.
1979–2004	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.

Finland

Table 3.B.6. Data sources: Finland

Year	Data source
Bank profitability	
1870–2010	Herrala (1999). Banking crises vs depositor crises: the era of the finnish markka. Scandinavian Economic History Review. Vol 47, No 2, 5-22. Banking sector balance sheets and income statements in Finland: selected figures. Data continued by the author for the latter years. Data kindly shared by the author.
2011–2016	Statistics Finland Online. Link: http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin_Passiivi/StatFin_Passiivi__rah__llai/. Change website to Finnish to access data prior to 2014.
Bank P&L components	
1870–1990	Herrala (1999). Banking crises vs depositor crises: the era of the finnish markka. Scandinavian Economic History Review. Vol 47, No 2, 5-22. Banking sector balance sheets and income statements in Finland: selected figures. Data continued by the author for the latter years. Data kindly shared by the author.
1991–2000	Statistical Yearbook of Finland (various years). Talletuspankit, Depositionsbanker (deposit taking institutions).
2001–2016	Statistics Finland Online. Link: http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin_Passiivi/StatFin_Passiivi__rah__llai/. Change website to Finnish to access data prior to 2014.
Bank dividends	
1870–1955	Aaku (1957). Suomen Liikepankit 1862-1955. Commercial banks.
1979–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.

France

Table 3.B.7. Data sources: France

Year	Data source
Bank profitability	
1870–1914	Bouvier, Furet and Gillet (1965). Le mouvement du profit en France au 19e siècle. Paris et La Haye. Data of individual banks is aggregated.
1915–1947	Annual Reports of major banks (various years): Credit Lyonnais and Societe Generale.
1953–1980	Commission de controle de banques (various years). Rapport Annuel.
1980–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.
Bank P&L components	
1980–2006	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2007–2015	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.
Bank dividends	
1870–1913	Bouvier, Furet and Gillet (1965). Le mouvement du profit en France au 19e siècle. Paris et La Haye. Data of individual banks is aggregated.
1988–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.

Germany

Table 3.B.8. Data sources: Germany

Year	Data source
Bank profitability	
1871–1882	Annual Reports of major banks (various years): Commerzbank and Deutsche Bank for 1871-1872, Commerzbank, Dresdener Bank and Deutsche Bank for 1873-1882.
1883–1920	Die Deutschen Banken im Jahre (various years). Special publication of 'Der Oekonomist'. Covers largest 50-150 commercial banks.
1925–1944	Annual Reports of major banks (various years): Commerzbank, Dresdener Bank and Deutsche Bank.
1952–1968	Annual Reports of major banks (various years): Commerzbank and Deutsche Bank.
1969–2016	Bundesbank Online. Statistics of banks' profit and loss accounts. Link: https://www.bundesbank.de/Navigation/EN/Statistics/Banks_and_other_financial_institutions/Banks/Statistics_of_the_banks_profit_and_loss_accounts/tables/tabellen.html. Table guv_tab8_en.
Bank P&L components	
1883–1920	Die Deutschen Banken im Jahre (various years). Special publication of 'Der Oekonomist'. Covers largest 50-150 commercial banks.
1969–2016	Bundesbank Online. Statistics of banks' profit and loss accounts. Link: https://www.bundesbank.de/Navigation/EN/Statistics/Banks_and_other_financial_institutions/Banks/Statistics_of_the_banks_profit_and_loss_accounts/tables/tabellen.html. Table guv_tab8_en.
Bank dividends	
1883–1920	Die Deutschen Banken im Jahre (various years). Special publication of 'Der Oekonomist'. Covers largest 50-150 commercial banks.
1979–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.

Italy

Table 3.B.9. Data sources: Italy

Year	Data source
Bank profitability	
1890–1973	Natoli, Piselli, Triglia and Vercelli (2016). Historical archive of credit in Italy. Bank of Italy, Economic History Working Papers No. 36.
1974–1992	Annual report of the Bank of Italy (various years).
1993–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	Bank of Italy – Statistical Database. Link: https://www.bancaditalia.it/statistiche/basi-dati/bds/index.html?com.dotmarketing.htmlpage.language=1. All banks.
Bank P&L components	
1974–1992	Annual report of the Bank of Italy (various years).
1993–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	Bank of Italy – Statistical Database. Link: https://www.bancaditalia.it/statistiche/basi-dati/bds/index.html?com.dotmarketing.htmlpage.language=1. All banks.
Bank dividends	
1984–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2010–2015	Bank of Italy – Statistical Database. Link: https://www.bancaditalia.it/statistiche/basi-dati/bds/index.html?com.dotmarketing.htmlpage.language=1. All banks.

Japan

Table 3.B.10. Data sources: Japan

Year	Data source
Bank profitability	
1930–1956	Economic Statistics Annual (1972). Statistics Department, Bank of Japan. Ordinary banks.
1957–1979	Bank of Japan, File CDAB0540. Ordinary Banks.
1980–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2009–2015	IMF Online. Financial Soundness Indicators. Link: data.imf.org/FSI.
Bank P&L components	
1930–1956	Economic Statistics Annual (1972). Statistics Department, Bank of Japan. Income and expenses of ordinary banks.
1956–1979	Bank of Japan, File CDAB0540. Ordinary Banks.
1980–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
Bank dividends	
1980–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.

Netherlands

Table 3.B.11. Data sources: Netherlands

Year	Data source
Bank profitability, P&L components and dividends	
1870–1941	Annual Reports of major banks (various years): 1909-1941: Incassobank, Rotterdamsche Bank, Amsterdamsche Bank, Twentsche Bank. 1877-1908: Twentsche Bank, Ontvang- en Betaalkas, Handel en Maatschappij. 1870-1976: Twentsche Bank. Sources: Eisfeld (1916). Das Niederländische Bankwesen. Den Haag. Kiliani (1923). Die Großbanken Entwicklung in Holland und die Mitteleuropäische Wirtschaft. Verlag von Felix Meiner in Leipzig. De Graaf (2012). Voor Handel en Maatschappij – Geschiedenis van de Nederlandsche Handel-Maatschappij, 1824-1964.
1948–1980	Centraal Bureau voor de Statistiek (various years). Maandstatistiek van het financiewezen. Commercial banks.
1981–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2008–2017	De Nederlandsche Bank Online. Link: https://statistiek.dnb.nl/en/downloads/index.aspx#/details/balance-sheet-of-the-dutch-banking-sector-consolidated/dataset/dcb6775e-1afa-4a45-bee0-669be22f8bd5/resource/ebb838b3-fe5f-422d-b6b2-2021ba06b4c98. Balance sheet and income statement of the Dutch banking sector.

Norway

Table 3.B.12. Data sources: Norway

Year	Data source
Bank profitability and dividends	
1874–1944	Statistics Norway Online. Various publications. Link: https://www.ssb.no/a/en/histstat/, section 13. Money and credit – Norges private aksjebanker og sparebanker.
1947–1975	Statistical Yearbook of Norway (various years). Forretningsbanker. Driftsregnskap.
1976–1980	Statistical Yearbook of Norway (various years). Offentlige og private banker. Resultatregnskap. Norske forretningsbanker og Norges sparebanker.
1980–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database.
2010–2017	Statistics Norway Online. Link: https://www.ssb.no/en/statbank/table/07880/tableViewLayout1/?rxid=e8526cc9-a688-4b75-857d-2c79e5112586.
Bank P&L components	
1900–1944	Statistics Norway Online. Various publications. Link: https://www.ssb.no/a/en/histstat/, section 13. Money and credit – Norges private aksjebanker og sparebanker.
1976–1980	Statistical Yearbook of Norway (various years). Offentlige og private banker. Resultatregnskap. Norske forretningsbanker og Norges sparebanker.
1980–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database.
2010–2017	Statistics Norway Online. Link: https://www.ssb.no/en/statbank/table/07880/tableViewLayout1/?rxid=e8526cc9-a688-4b75-857d-2c79e5112586.

Portugal

Table 3.B.13. Data sources: Portugal

Year	Data source
Bank profitability	
1931–1961	Instituto Nacional de Estatistica, Estatisticas Financeiras (various issues). Bancos, Casas Bancarias e Caixas Economicas.
1962–1978	Instituto Nacional de Estatistica, Estatisticas Monetaria Financeiras (various issues). Group of "Bancos e casas bancario" less "Banco Formento" and "Bank of Portugal".
1980–2007	OECD Banking Statistics. Income statement and balance sheet. Online Database.
2008–2016	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.
Bank P&L components	
1980–2007	OECD Banking Statistics. Income statement and balance sheet. Online Database.
2008–2016	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.

Spain

Table 3.B.14. Data sources: Spain

Year	Data source
Bank profitability	
1901–1978	Tafunell (2000). La rentabilidad financiera de la empresa española, 1880-1981: una estimación en perspectiva sectorial. Revista de Historia Industrial 18: 71-112.
1979–2009	OECD Banking Statistics. Income statement and balance sheet. Online Database.
2010–2015	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.
Bank P&L components	
1979–2007	OECD Banking Statistics. Income statement and balance sheet. Online Database.
2008–2015	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.
Bank dividends	
1979–2007	OECD Banking Statistics. Income statement and balance sheet. Online Database.

Sweden

Table 3.B.15. Data sources: Sweden

Year	Data source
Bank profitability and dividends	
1870–1997	Swedish Riksbank. Bank Lending and Borrowing 1870-2006. Data source: Hortlund (2005). The long-term relationship between capital and earnings in banking. SSE/EFI Working Paper Series in Economics and Finance No. 611.
1997–2015	Statistics Sweden Online. Link: http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__FM__FM0402/?rxid=3d618be3-5da4-4cb7-9934-972462441227. Financial Markets – Financial Enterprises. Balance sheets and income statement for all banks.
Bank P&L components	
1870–1997	Swedish Riksbank. Bank Lending and Borrowing 1870-2006. Data source: Hortlund (2005). The long-term relationship between capital and earnings in banking. SSE/EFI Working Paper Series in Economics and Finance No. 611.
1988–1995	Riksbank Yearbook (various years). Banking sector balance sheets and profit and loss account. Available funds and their distribution. All banks.
1997–2015	Statistics Sweden Online. Link: http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__FM__FM0402/?rxid=3d618be3-5da4-4cb7-9934-972462441227. Financial Markets – Financial Enterprises. Balance sheets and income statement for all banks.

Switzerland

Table 3.B.16. Data sources: Switzerland

Year	Data source
Bank profitability and dividends	
1870–1905	Historical Statistics of Switzerland Online. Link: https://www.fsw.uzh.ch/histstat/main.php. Table O.12. Diskontobanken, Kantonalbanken und übrige Emissionsbanken: Passiven, Aktiven und Gewinnrechnung 1826-1910.
1906–2002	Schweizerische Nationalbank. Historische Zeitreihen. Die Banken in der Schweiz. Link: https://www.snb.ch/de/iabout/stat/statrep/statpubdis/id/statpub_histz_arch. Balance sheet data from Table 9. Net profit after taxes from Tables 29.1 and 29.2.
1996–2016	Schweizerische Nationalbank Online. Link: https://data.snb.ch. Annual banking statistics. All banks.
Bank P&L components	
1870–1905	Historical Statistics of Switzerland Online. Link: https://www.fsw.uzh.ch/histstat/main.php. Table O.12. Diskontobanken, Kantonalbanken und übrige Emissionsbanken: Passiven, Aktiven und Gewinnrechnung 1826-1910.
1906–1995	Schweizerische Nationalbank. Historische Zeitreihen. Die Banken in der Schweiz. Link: https://www.snb.ch/de/iabout/stat/statrep/statpubdis/id/statpub_histz_arch. Balance sheet data from Table 9. Income components from Tables 29.1 and 29.2.
1906–1992	Historical Statistics of Switzerland Online. Link: https://www.fsw.uzh.ch/histstat/main.php. Table O.15. Banken (1): Gewinn- und Verlustrechnung 1906-1992.
1993–1995	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues. All banks.
1996–2016	Schweizerische Nationalbank Online. Link: https://data.snb.ch. Annual banking statistics. All banks.

United Kingdom

Table 3.B.17. Data sources: United Kingdom

Year	Data source
Bank profitability, P&L components and dividends	
1870–1920	Capie and Webber (1985). Profits and profitability in british banking, 1870-1939. Centre for Banking and International Finance Discussion Paper 18. Series: English and Welsh Joint Stock Banks – Aggregate Profits.
1920–1967	Capie and Billings (2004). Evidence on competition in English commercial banking, 1920—1970. Financial History Review. Volume 11 / Issue 01 / pp 69 - 103.
1968	Ackrill and Hannah (2001). Barclays, The Business of Banking 1690-1996. Cambridge University Press. Tables B1, B2, B4, B6.
1969–1976	CLCB Statistical Unit. London Clearings Banks 1966-1976. Profit and balance sheet statistics. Consolidated accounts.
1977–1979	Ackrill and Hannah (2001). Barclays, The Business of Banking 1690-1996. Cambridge University Press. Tables B1, B2, B4, B6.
1980–2008	OECD Banking Statistics. Income statement and balance sheet. Online Database and print issues.
2009–2015	European Central Bank Online. Consolidated banking data. Link: https://www.ecb.europa.eu/stats/supervisory_prudential_statistics/consolidated_banking_data/html/index.en.html. Domestic banking groups and stand alone banks, foreign (EU and non-EU) controlled subsidiaries and foreign (EU and non-EU) controlled branches.

United States

Table 3.B.18. Data sources: United States

Year	Data source
Bank profitability	
1870–1918	Historical Statistics of the United States. Link: https://hsus.cambridge.org/HSUSWeb/HSUSEntryServlet. Table: National banks – number, earnings, and expenses: 1869—1998 Cj238-250.
1919–1950	Banking and Monetary Statistics 1914-1941 and 1941-1970. Tables: Member bank earnings, expenses and dividends, 1919-1941. Member bank income, expenses and dividends 1941-70. All FDIC insured commercial banks.
1951–2015	FDIC Online. Historical Statistics on Banking. Link: https://www5.fdic.gov/hsob/HSOBRpt.asp. All FDIC insured commercial banks.
Bank P&L components	
1870–1935	Historical Statistics of the United States. Link: https://hsus.cambridge.org/HSUSWeb/HSUSEntryServlet. Table: National banks – number, earnings, and expenses: 1869—1998 Cj238-250.
1935–1966	FDIC Online. Historical Statistics on Banking. Link: https://www5.fdic.gov/hsob/HSOBRpt.asp. All FDIC insured commercial banks.
1967–2015	FDIC Online. Historical Statistics on Banking. Link: https://www5.fdic.gov/hsob/HSOBRpt.asp. All FDIC insured commercial banks.
Bank dividends	
1870–1918	Historical Statistics of the United States. Link: https://hsus.cambridge.org/HSUSWeb/HSUSEntryServlet. Table: National banks – number, earnings, and expenses: 1869—1998 Cj238-250.
1919–1945	Banking and Monetary Statistics 1914-1941 and 1941-1970. Tables: Member bank earnings, expenses and dividends, 1919-1941. Member bank income, expenses and dividends 1941-70. All FDIC insured commercial banks.
1946–1966	FDIC Online. Historical Statistics on Banking. Link: https://www5.fdic.gov/hsob/HSOBRpt.asp. All FDIC insured commercial banks.
1967–2015	FDIC Online. Historical Statistics on Banking. Link: https://www5.fdic.gov/hsob/HSOBRpt.asp. All FDIC insured commercial banks.

References

Adrian, Tobias, Emanuel Moench, and Hyun S. Shin. 2019. "Dynamic Leverage Asset Pricing." CEPR Discussion Paper 11466.

Aikman, David, Andrew G. Haldane, and Benjamin D. Nelson. 2015. "Curbing the Credit Cycle." *Economic Journal* 125 (585): 1072–1109.

Akinci, Ozge, and Jane Olmstead-Rumsey. 2018. "How Effective are Macroprudential Policies? An Empirical Investigation." *Journal of Financial Intermediation* 33: 33–57.

Baron, Matthew, Emil Verner, and Wei Xiong. 2020. "Banking Crises without Panics." *Quarterly Journal of Economics* 136 (1): 51–113.

Baron, Matthew, and Wei Xiong. 2017. "Credit Expansion and Neglected Crash Risk." *Quarterly Journal of Economics* 132 (2): 713–764.

Bernanke, Ben S., and Mark Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79 (1): 14–31.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review* 110 (9): 2748–2782.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2018. "Diagnostic Expectations and Credit Cycles." *Journal of Finance* 73 (1): 199–227.

Bordalo, Pedro, Nicola Gennaioli, Andrei Shleifer, and Stephen J. Terry. 2019. "Real Credit Cycles." Working Paper.

Brunnermeier, Markus K., and Yuliy Sannikov. 2014. "A Macroeconomic Model with a Financial Sector." *American Economic Review* 104 (2): 379–421.

Capie, Forrest, and Mark Billings. 2001. "Profitability in English Banking in the Twentieth Century." *European Review of Economic History* 5 (3): 367–401.

Carvalho, Daniel, Janet Gao, and Pengfei Ma. 2020. "Loan Spreads and Credit Cycles: The Role of Lenders' Personal Economic Experiences." Working Paper.

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

Danielsson, Jón, Marcela Valenzuela, and Ilknur Zer. 2018. "Learning from History: Volatility and Financial Crises." *Review of Financial Studies* 31: 2774–2805.

De Stefani, Alessia. 2020. "House price history, biased expectations and credit cycles: The role of housing investors." *Real Estate Economics*, Forthcoming.

Dell'Ariccia, Giovanni, Deniz Igan, Luc Laeven, and Hui Tong. 2016. "Credit Booms and Macrofinancial Stability." *Economic Policy* 31 (86): 299–355.

Duke CFO Global Business Outlook. 2018. Accessed: 2018-05-25.

Fahlenbrach, Rüdiger, Robert Prilmeier, and René M Stulz. 2017. "Why Does Fast Loan Growth Predict Poor Performance for Banks?" *Review of Financial Studies* 31 (3): 1014–1063.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer. 2016. "Expectations and Investment." *NBER Macroeconomics Annual* 30 (1): 379–431.

Goldstein, Itay, and Ady Pauzner. 2005. "Demand–Deposit Contracts and the Probability of Bank Runs." *Journal of Finance* 60 (3): 1293–1327.

Gorton, Gary, and Guillermo Ordonez. 2020. "Good Booms, Bad Booms." *Journal of the European Economic Association* 18 (2): 618–665.

Greenwood, Robin, and Samuel G. Hanson. 2013. "Issuer Quality and Corporate Bond Returns." *Review of Financial Studies* 26 (6): 1483–1525.

Greenwood, Robin, and Samuel G. Hanson. 2015. "Waves in Ship Prices and Investment." *Quarterly Journal of Economics* 130 (1): 55–109.

Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of Returns and Expected Returns." *Review of Financial Studies* 27 (3): 714–746.

Greenwood, Robin Marc, Samuel Gregory Hanson, and Lawrence J. Jin. 2019. "Reflexivity in Credit Markets." NBER Working Paper 25747.

He, Zhiguo, and Wei Xiong. 2012. "Dynamic Debt Runs." *Review of Financial Studies* 25 (6): 1799–1843.

Holmstrom, Bengt, and Jean Tirole. 1997. "Financial Intermediation, Loanable Funds, and the Real Sector." *Quarterly Journal of Economics* 112 (3): 663–691.

Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina. 2017. "Macroprudential Policy, Countercyclical Bank Capital Buffers, and Credit Supply: Evidence from the Spanish Dynamic Provisioning Experiments." *Journal of Political Economy* 125 (6): 2126–2177.

Jordà, Òscar, Björn Richter, Moritz Schularick, and Alan M. Taylor. 2021. "Bank Capital Redux: Solvency, Liquidity, and Crisis." *Review of Economic Studies* 88 (1): 260–286.

Jordà, Òscar, Moritz Schularick, Alan M Taylor, and Felix Ward. 2019. "Global Financial Cycles and Risk Premiums." *IMF Economic Review* 67 (1): 109–150.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2017. "Macrofinancial History and the New Business Cycle Facts." *NBER Macroeconomics Annual* 31 (1): 213–263.

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante. 2020. "The Housing Boom and Bust: Model Meets Evidence." *Journal of Political Economy* 128 (9): 3285–3345.

Kindleberger, Charles P. 1978. *Manias, Panics and Crashes: A History of Financial Crises*. Palgrave Macmillan.

Kirti, Divya. 2020. "Lending Standards and Output Growth." Working Paper.

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

Krishnamurthy, Arvind, and Wenhao Li. 2020. "Dissecting Mechanisms of Financial Crises: Intermediation and Sentiment." NBER Working Paper 24932.

Krishnamurthy, Arvind, and Tyler Muir. 2017. "How Credit Cycles Across a Financial Crisis." NBER Working Paper 23850.

Kuchler, Theresa, and Basit Zafar. 2019. "Personal Experiences and Expectations about Aggregate Outcomes." *Journal of Finance* 74 (5): 2491–2542.

Kuvshinov, Dmitry. 2018. "The Time Varying Risk Puzzle." Working Paper.

Landier, Augustin, Yueran Ma, and David Thesmar. 2018. "New Experimental Evidence on Expectations Formation." CEPR Discussion Paper 12527.

Meiselman, Ben S., Stefan Nagel, and Amiyatosh K. Purnanandam. 2018. "Judging Banks' Risk by the Profits They Report." Working Paper.

Mian, Atif, and Amir Sufi. 2018. "Finance and Business Cycles: the Credit-Driven Household Demand Channel." *Journal of Economic Perspectives* 32 (3): 31–58.

Mian, Atif R., Amir Sufi, and Emil Verner. 2017. "Household Debt and Business Cycles Worldwide." *Quarterly Journal of Economics* 132 (4): 1755–1817.

Minsky, Hyman P. 1977. "The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to "Standard" Theory." *Challenge* 20 (1): 20–27.

Newey, Whitney K., and Kenneth D. West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies* 61 (4): 631–653.

Rey, Hélène. 2016. "International Channels of Transmission of Monetary Policy and the Mundellian Trilemma." *IMF Economic Review* 64 (1): 6–35.

Richter, Björn, Moritz Schularick, and Paul Wachtel. 2020. "When to Lean Against the Wind." *Journal of Money, Credit and Banking*, Forthcoming.

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): 1029–1061.

Shiller, Robert. 1981. “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71 (3): 421–36.

Simsek, Alp. 2013. “Belief Disagreements and Collateral Constraints.” *Econometrica* 81 (1): 1–53.

Chapter 4

Sovereigns Going Bust: Estimating the Cost of Default*

Joint with Dmitry Kuvshinov

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

* 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). An early version of this research project entitled "The Bitter Pill: Estimating the Average Treatment Effect of Sovereign Default" was submitted as a master's thesis by Kaspar Zimmermann. Kuvshinov: University of Bonn, email: dmitry.kuvshinov@uni-bonn.de. Zimmermann: University of Bonn, email: kaspar.zimmermann@uni-bonn.de.

Zdzienicka, 2012).[1] Second, we do not yet know what exactly generates the default 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 compre-

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).

hensive 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.

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, 2018).

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 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).

Table 4.1. Existing estimates of the cost of sovereign default

Paper	Default cost, % GDP		Method
	First year	Medium term	
Historical unconditional estimates:			
Reinhart and Rogoff (2011b)	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 + counterfactual 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 (2011b) 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).

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. 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 (2018) 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 (2011a) 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 4.1. Whilst historical studies (see Tomz and Wright, 2007; Reinhart and Rogoff, 2011b) 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.

Table 4.2. 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.

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.

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 4.2 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:

$$\widehat{PD}_{i,t} \;=\; \Lambda(X^P_{t-1}, \tilde{X}^P_{t-1}, \tilde{X}^P_{t-2}, \hat{\beta}) \tag{4.1}$$

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 (4.1). 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):

$$\Delta y_{i,t+h} \;=\; \alpha_i + \theta_h \delta_{i,t} + \Gamma_{h,1} X^C_{i,t-1} + \tilde{\Gamma}_{h,1} \tilde{X}^C_{i,t-1} + \tilde{\Gamma}_{h,2} \tilde{X}^C_{i,t-2} + \varepsilon_{i,t} \quad h \in \{0, \ldots, 9\} \tag{4.2}$$

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 (4.1) than controls in stage 2 (4.2), 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 (4.2) 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:

$$ATE_h(\delta)^{\text{IPSWRA}} = \frac{1}{n_{\text{Def}}} \sum_i \sum_t \frac{\Delta\hat{y}_{i,t+h} * \delta_{i,t}}{\widehat{PD_{i,t}}} - \frac{1}{n_{\text{NoDef}}} \sum_i \sum_t \frac{\Delta\hat{y}_{i,t+h} * (1-\delta_{i,t})}{1-\widehat{PD}_{i,t}} \tag{4.3}$$

Here, $\Delta\hat{y}_{i,t+h}$ is the forecast obtained by estimating (4.2), $\delta_{i,t}$ 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:

$$\Delta y_{i,t+h} = \alpha_i + \theta_h^w \delta_{i,t} * IPW_{i,t} + \Gamma_{h,1} X_{i,t-1}^C * IPW_{i,t} + \tilde{\Gamma}_{h,1} \tilde{X}_{i,t-1}^C * IPW_{i,t} + \tilde{\Gamma}_{h,2} \tilde{X}_{i,t-2}^C * IPW_{i,t} + \varepsilon_{i,t}, \tag{4.4}$$

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]

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

This is crucial, since the fallout from macroeconomic crises in emerging markets tends 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 (4.2) imposes little structure on the data, and allows the response $ATE_h(\delta)^{\text{IPSWRA}}$ to vary in a non-linear manner over the forecast horizon h – unlike, say, a VAR which carries a linear structure of the form $ATE_h(\delta)^{\text{VAR}} = F * ATE_{h-1}(\delta)$, 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 (4.1) and (4.2) 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 (4.1) and local projection controls in (4.2), 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

$$\Delta y_{i,t+h}(\delta) \perp\!\!\!\perp \delta_{i,t} \mid X^P_{t-1}, \tilde{X}^P_{t-1}, \tilde{X}^P_{t-2}, \beta \quad h \in \{0, \ldots, 9\}, \tag{4.5}$$

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 (4.1).

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.

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 (4.5). 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.

Table 4.3. Alternative definitions of sovereign default

Source	Definition	Key criteria
Standard & Poor's (baseline)	Failure to make a payment; distressed restructurings	Legal
Reinhart and Rogoff (2011a)	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

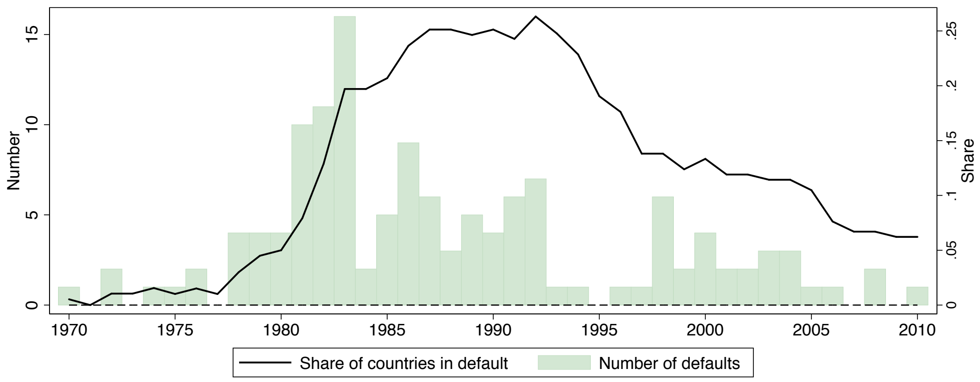
4.4 A dataset of sovereign defaults and their drivers

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 defining what constitutes a sovereign default. The literature has proposed several such definitions, which are summarised in Table 4.3. 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 (2011a) 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

Figure 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.

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 (2011a) and 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, 2011b), 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 4.3 (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 (2011a), Reinhart and Trebesch (2016) and Beers and Nadeau (2015). Appendix Section 4.A.2 and Table 4.A.5 provide further details on the data construction, and Appendix Figure 4.A.2 provides a timeline of sovereign defaults under the five alternative definitions for each country in our sample.

Figure 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

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.

all independent countries.[6] Appendix Figure 4.A.1 shows the corresponding trends for the four alternative default definitions, which paint a similar picture to Figure 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 4.1, 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 exam-

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.

ple 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 4.4 (bottom row) and Figure 4.2 (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 4.4 middle row, Figure 4.2 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]

7. As shown in equation (4.4), 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.

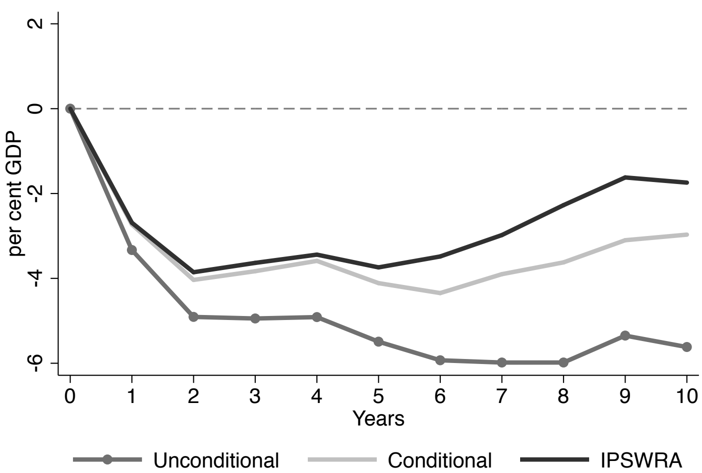
8. Using a "sandwich" estimator, we find that the conditional and unconditional paths are significantly different, at 10% level over the full horizon, with higher significance levels for individual years 1, 7, 8, 9 and 10. We cannot test for the difference between the IPSWRA and unconditional paths because 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 4.4. 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 4.2. 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.

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 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 es-

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.

timate 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 4.1 and Cerra and Saxena, 2008). 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 4.2 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 (2011b) 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

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".

using the methodology of Cantor and Packer (1996) (see Appendix 4.A.1 for further detail).

Table 4.5 presents the results. Panel (a) limits the regressions to a smaller sample – effectively, the 1990s – but uses the more accurate raw data on credit 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 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 4.4, 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 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.

Table 4.5. 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 4.4, 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 4.4. 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.

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 and variable choices, are potentially subject to further selection biases which we examine next.

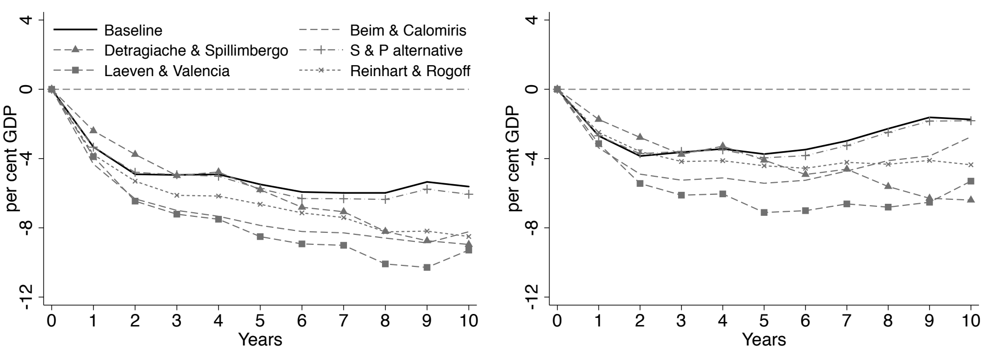
4.5.2 Other considerations

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 4.3 shows default cost estimates under the five alternative definitions listed in Section 4.4 Table 4.3. 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 4.3a shows the unconditional estimates with country fixed effects only, and Figure 4.3b 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 (2011a) 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.

Figure 4.3. Cost estimates for different definitions of sovereign default



(a) Unconditional **(b)** IPSWRA

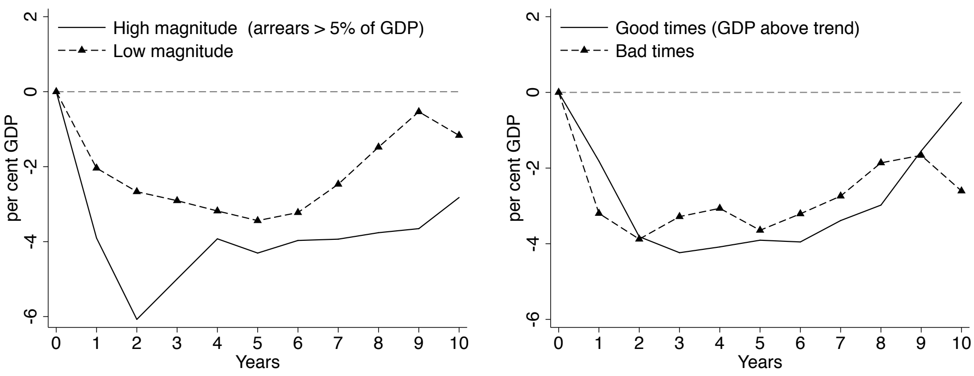
Notes: The baseline definition uses *Standard & Poor's* data, Beim & Calomiris group default spells less than 5 years apart together, Detragiache & Spillimbergo definition is based on arrears to total debt, Laeven & Valencia focus on sovereign crises, Reinhart & Rogoff takes the data on defaults and distressed restructurings from Reinhart and Rogoff (2011a), and S & P alternative drops defaults which occur when the country is still in default on another type of debt.

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 4.4a 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

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 4.4. Impact of default magnitude and economic situation on the cost of default



(a) Default magnitude

(b) Defaulting in good and bad times

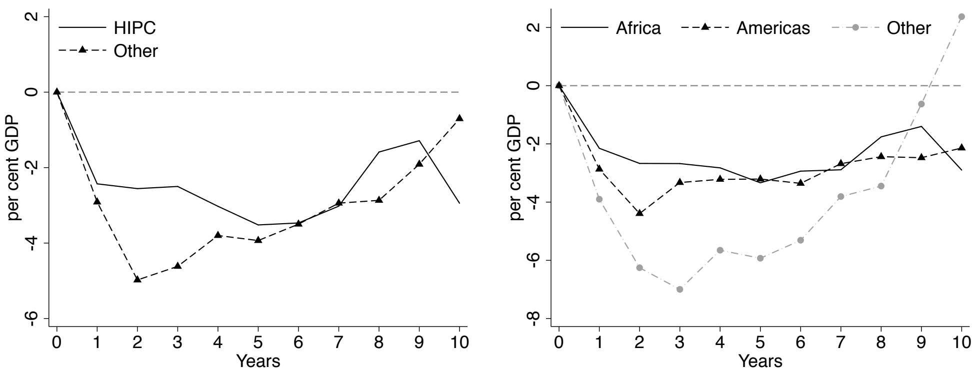
Notes: IPSWRA estimates. Left panel shows the cost of high-magnitude compared to low-magnitude defaults. Default is classified as high-magnitude if debt arrears to private creditors exceed 5% of GDP in the year of default. Right panel shows the cost of defaulting during times of good, or bad economic performance. Good economic performance means that GDP growth is on average above trend during the three years preceding default.

could still be driven by a small subsample of countries with bad economic fundamentals, both before and after 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 4.4b 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 4.4b provide further details.

Default cost among different groups of countries. Figure 4.4b 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 4.5a 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 4.5b estimates the default costs across different continents and again finds that these are similar, with

Figure 4.5. 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.

defaults in Asian and European countries, included in the "Other" group, having a somewhat higher cost.

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 Morais and Wright, 2008; Reinhart and Rogoff, 2011a), 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 Gennaioli, Martin, and Rossi, 2014; Bocola, 2016; Jordà, Schularick, and Taylor, 2016; Gennaioli, Martin, and Rossi, 2018; Reinhart and Rogoff, 2011a). 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 default events. The list includes both those defaults where sovereign distress directly

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

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.7 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, 2011a), 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, 2021). 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.4 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 and Figure 4.6 present the IPSWRA estimates of the cost for those defaults that are followed by a systemic banking crises (Table 4.6 bottom row, Figure 4.6b), compared to those which are not (Table 4.6 top row, Figure 4.6a). 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 civil wars in

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.

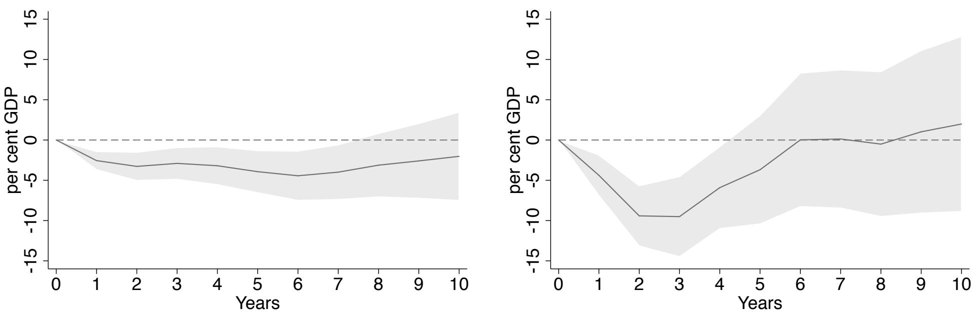
14. As in Section 4.5, we use a "sandwich" estimator to test for joint difference in the treatment effect estimates for standalone defaults vs sovereign-banking crisis, in all of the years 1–10.

Table 4.6. 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.

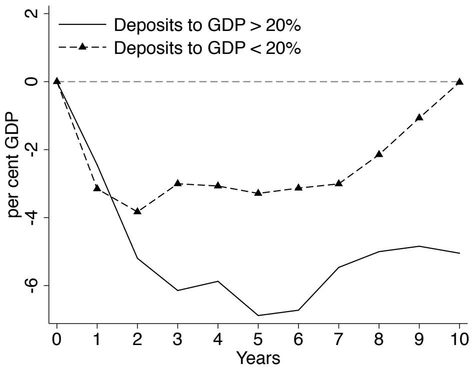
Figure 4.6. Cost of sovereign default and systemic banking crises



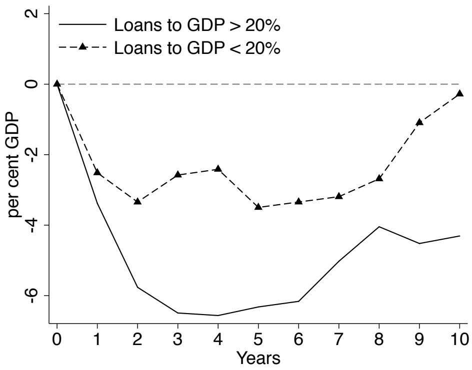
(a) Default and no systemic banking crisis **(b)** Default and systemic banking crisis

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 4.7. 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.

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.5 and Table 4.B.11 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 4.7 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, Kaminsky, 2006; De Paoli, Hoggarth, and Saporta, 2009; Reinhart and Rogoff, 2011a). 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.6.

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.12 and 4.B.13. 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 4.6b). 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

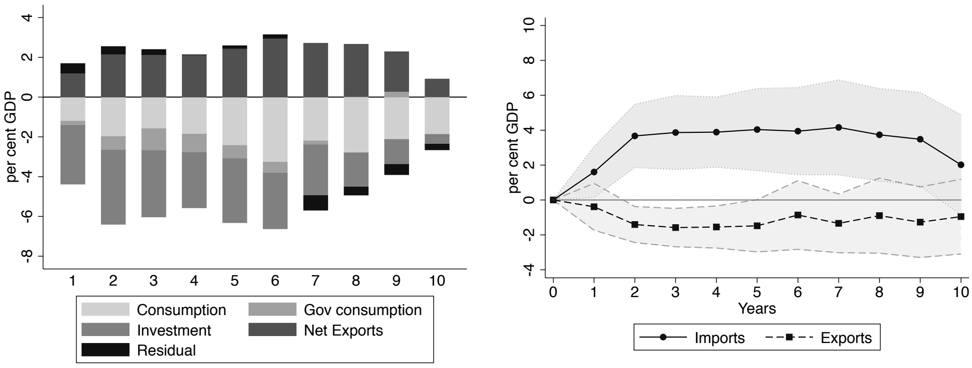
Theoretical models have proposed a number of channels though which sovereign default can harm the economy (see Section 4.2). These transmission channels generally have an asymmetric impact on different economic sectors: for example, banking disruption may disproportionately affect investment, while trade 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 4.8a breaks down the aggregate GDP cost in Table 4.4 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 lists the point estimates and standard errors underlying Figure 4.8. In order to interpret these figures one has to take into account the share of each component in GDP (Appendix Table 4.B.14). 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.14) – this represents a drop of more than one-fifth in relative terms. The

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 4.8. 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.

fall in private consumption is modest, especially given its GDP share of 70%. In 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. 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

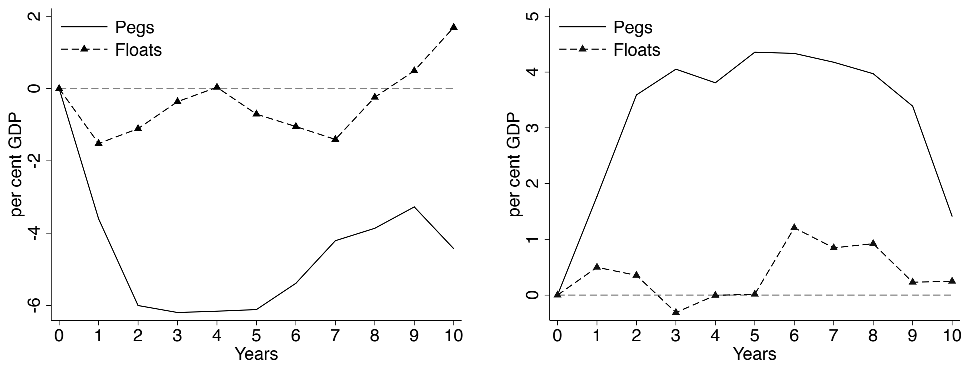
pegged and floating exchange rate regimes in the year before default, using the classification in Ilzetzki, Reinhart, and Rogoff (2019).[18]

Figure 4.9a 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.15. 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 4.9b 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 4.9. 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.15.

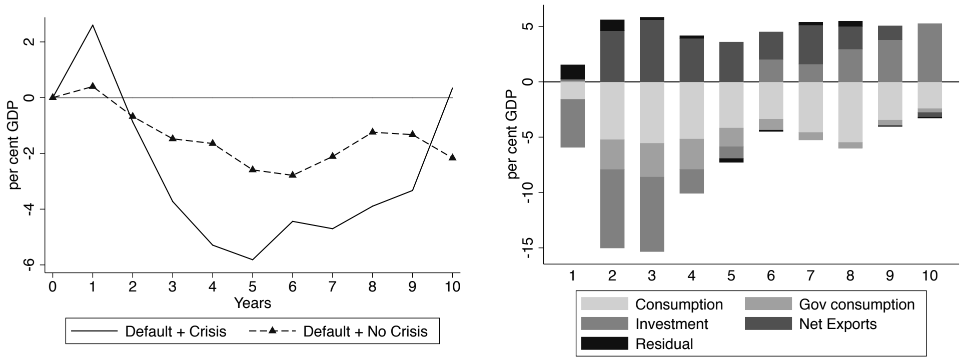
exchange rate is fixed. In fact, exports actually decline after the default, and all of the external adjustment in Figure 4.9b takes the form of lower imports (see Appendix Figure 4.B.6). This adjustment seems to impose substantial costs on the economy in the form of lower GDP.

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 4.8a may, therefore, be directly con-

Figure 4.10. 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 4.10a 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 4.10b 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 in the same way that Figure 4.8a does for our baseline estimates. Appendix Table 4.B.16 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 4.10b 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, Eisert, Eufinger, and Hirsch (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 (Manasse, Roubini, and Schimmelpfennig, 2003; Manasse and Roubini, 2009; Reinhart and Rogoff, 2011a). 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 (2011a) 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 (2011a), 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, Kuvshinov, Schularick, and Taylor (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 IPSWRA, 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 (2019)	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 (2019). 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.5. Alternative default definitions

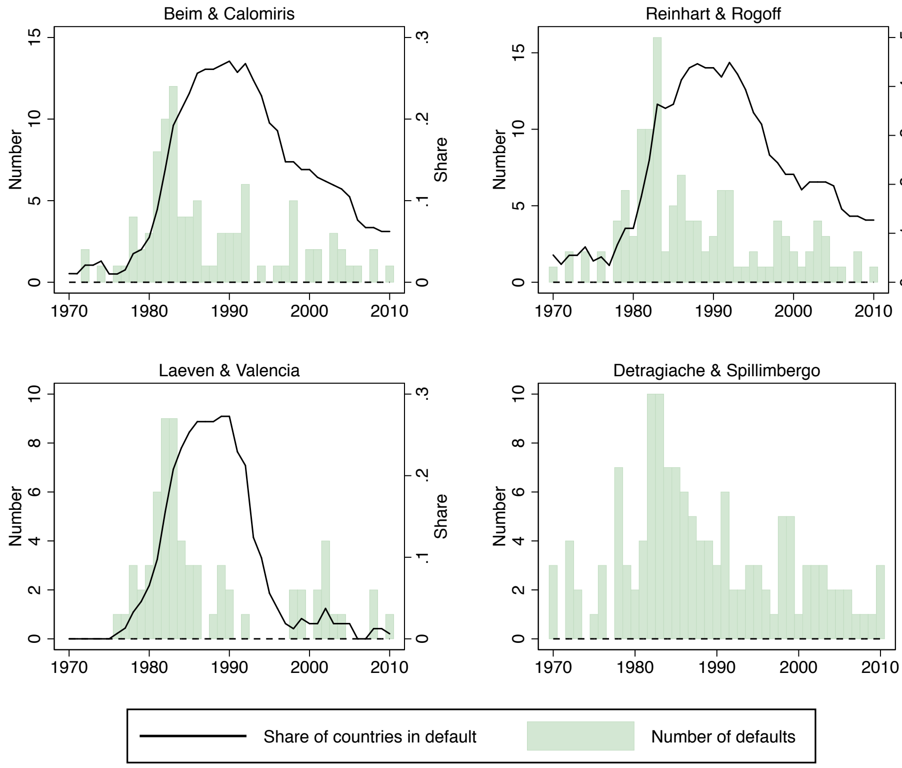
Source	Original definiton	Extension
Reinhart and Rogoff (2011a)	Failure to make a payment; distressed restructurings	Use original data, extended by Reinhart and Trebesch (2016)
Beim and Calomiris (2000)	Failure to make a payment over > 6 months, no political defaults, group default spells less than 5 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.5 details the construction of each alternative default definition variable in our sample. The Reinhart and Rogoff (2011a) 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 (2011a) 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.5, 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 4.1 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 (2011a). 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 (2011a) 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 (2011a) 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.2 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.3 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.

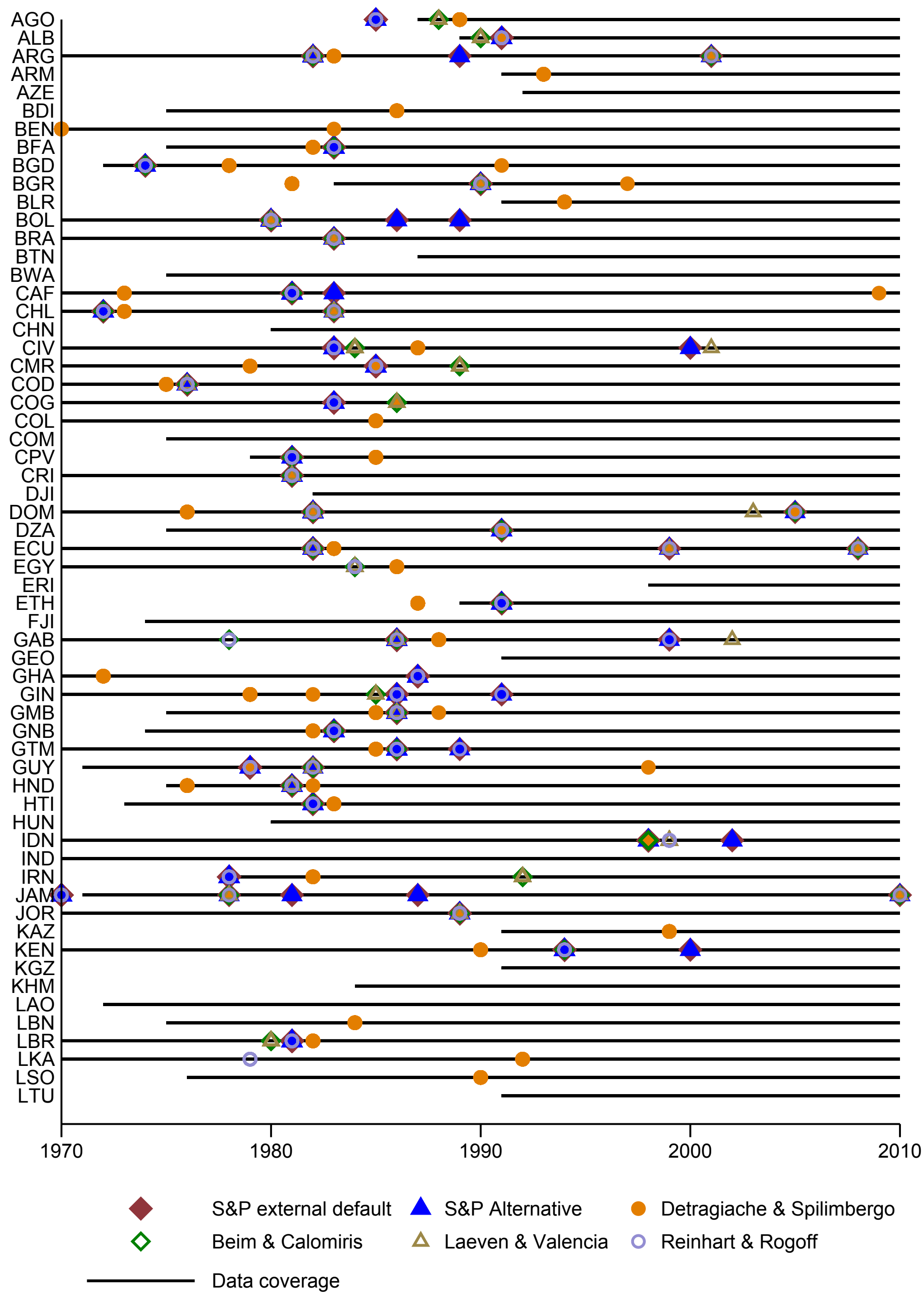
Table 4.A.6 summarises the joint occurrence of the different crisis events throughout our historical sample. We include 92 defaults, 17 of which coincide with systemic banking crises, 35 – with currency crises, 5 – with both banking and currency crises, and between 10 and 20 – with various types of political 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.6. 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.2. The timeline of sovereign defaults for individual countries



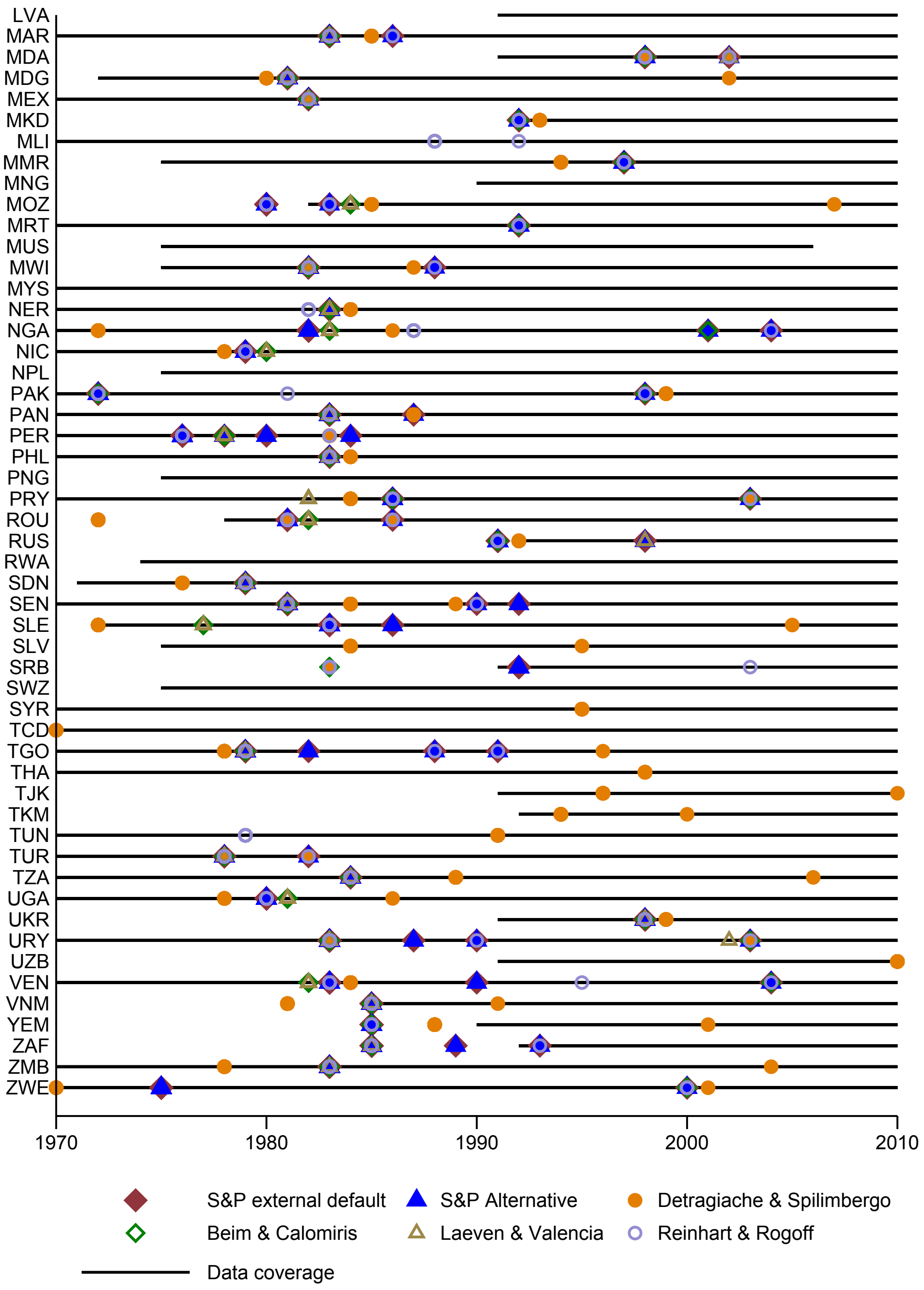
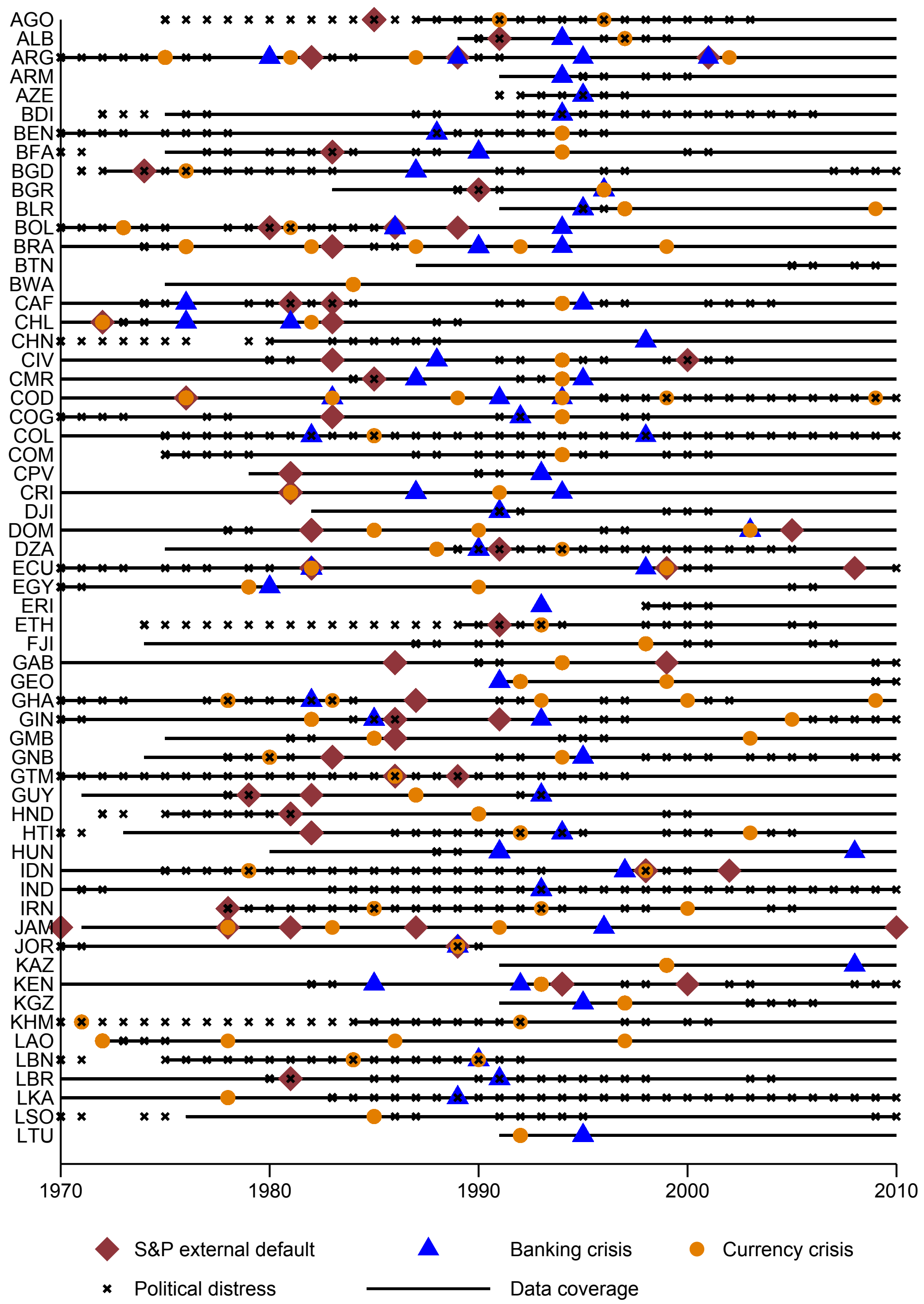
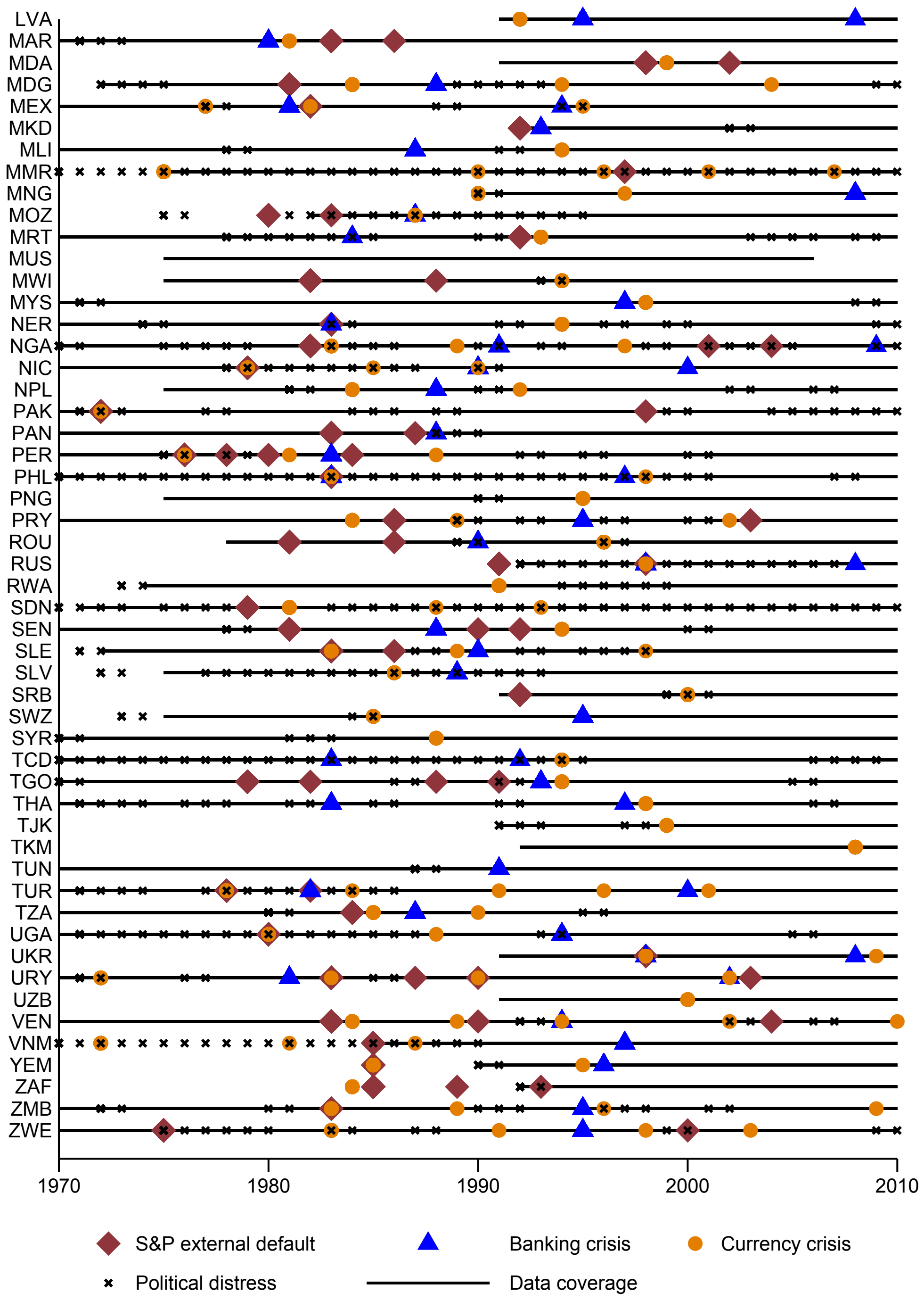

LVA
MAR
MDA
MDG
MEX
MKD
MLI
MMR
MNG
MOZ
MRT
MUS
MWI
MYS
NER
NGA
NIC
NPL
PAK
PAN
PER
PHL
PNG
PRY
ROU
RUS
RWA
SDN
SEN
SLE
SLV
SRB
SWZ
SYR
TCD
TGO
THA
TJK
TKM
TUN
TUR
TZA
UGA
UKR
URY
UZB
VEN
VNM
YEM
ZAF
ZMB
ZWE
1970
1980
1990
2000
2010
S&P external default
S&P Alternative
Detragiache & Spilimbergo
Beim & Calomiris
Laeven & Valencia
Reinhart & Rogoff
Data coverage

Figure 4.A.3. 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.7 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 and Figure 4.6. 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.7. List of joint sovereign-bankign crisis events, where the sovereign default preceded or was not caused by the banking crisis

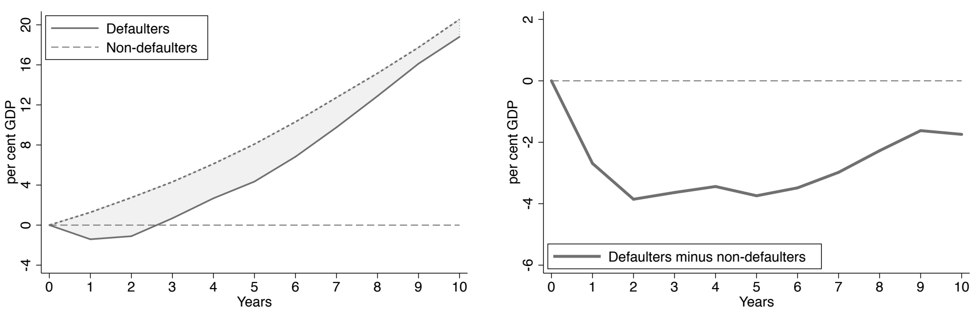
Episode	Narrative of events
Sovereign default and banking crisis in the same year:	
Turkey, 1982	Sovereign finance problems starting in 1970s, default in 1978, coup in 1980, another default in 1982, accompanied by a banking crisis.
Niger, 1983	Boom in uranium prices accompanied by sovereign and private 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 (4.2). 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 (4.1)) 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 4.2, 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 4.2. 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 (4.4)) – 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 4.4 (bottom row) and Figure 4.2 (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

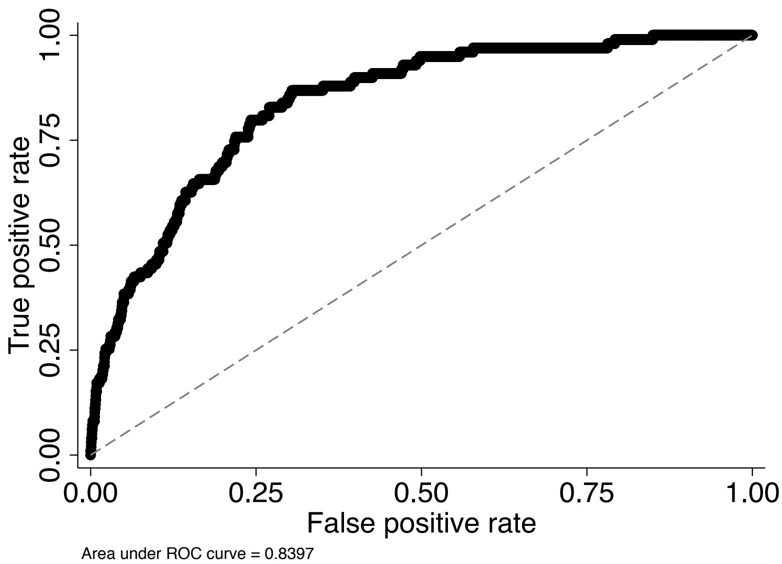


Table 4.B.3. IPSWRA second stage: IPS-weighted regression results

	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.

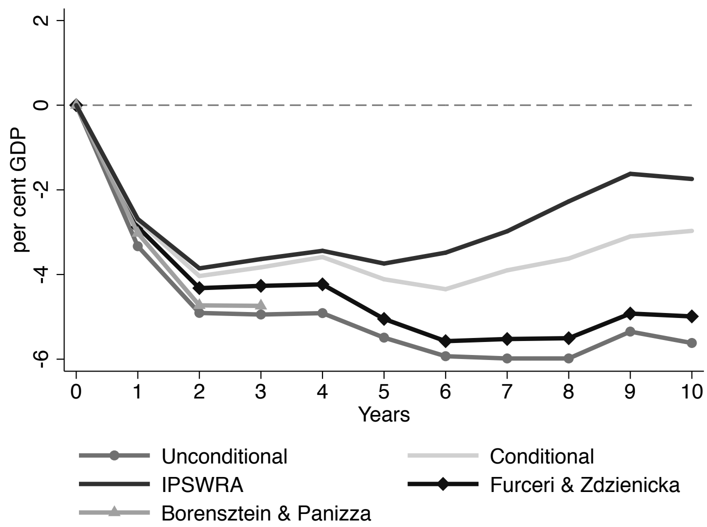
4.B.3 Comparison of baseline results with existing literature

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 IPSWRA 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.5, it includes the baseline estimates of Table 4.4, 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 (2011a), 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 4.4a 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 4.4b 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 bad times, which seems to go against the assumptions often made in theoretical literature.

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.

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 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

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.

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.

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

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

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

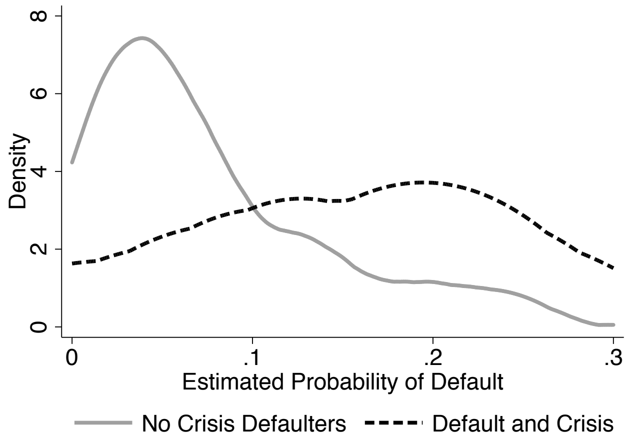
Estimating the cost of joint sovereign-banking crises faces two main challenges. First, our control and predictor variables are primarily selected to forecast sovereign defaults, and may thus do a poor job of forecasting banking crises and joint sovereign-banking crisis events. Second, we want to exclude 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.4 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.5 and Table 4.B.11 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.5 shows both the unconditional and IPSWRA cost estimates, and Table 4.B.11 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.5). The 67 standalone defaults in the sample are, on average, around 1 percentage point less

Figure 4.B.4. Predicted default probabilities for "standalone" sovereign defaults, and those followed by a systemic banking crisis



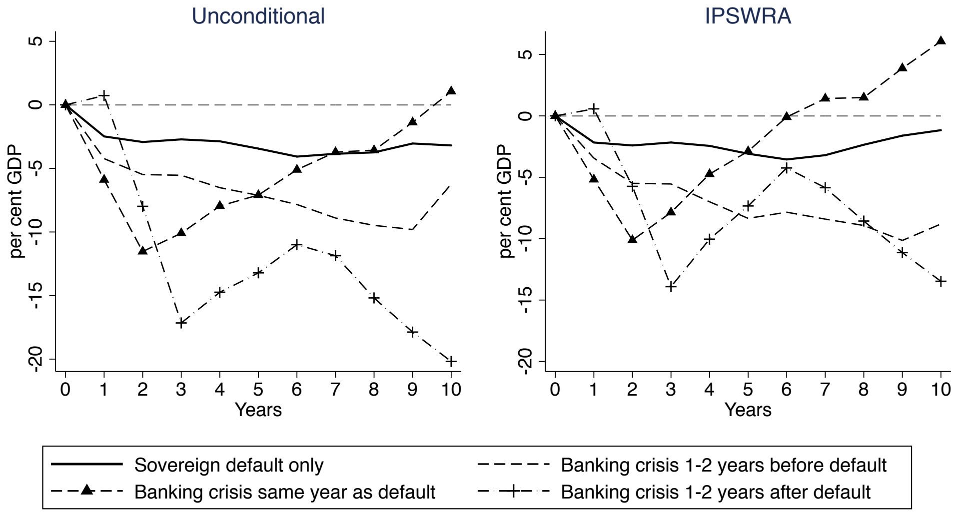
Notes: Kernel density plot of the predicted default probability based on IPSWRA first-stage logit model (as specified in equation (4.1), predictors as in Table 4.A.1.). The "default and crisis" observations are those where a sovereign default is followed by a systemic banking crisis within two years. The "no crisis defaulters" are defaults not followed by a systemic banking crisis.

costly than our baseline estimate in Table 4.4. Taken together, these results suggest 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.12 and 4.B.13 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.

Figure 4.B.5. 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.11. 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.12. 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.13. 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.14. 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

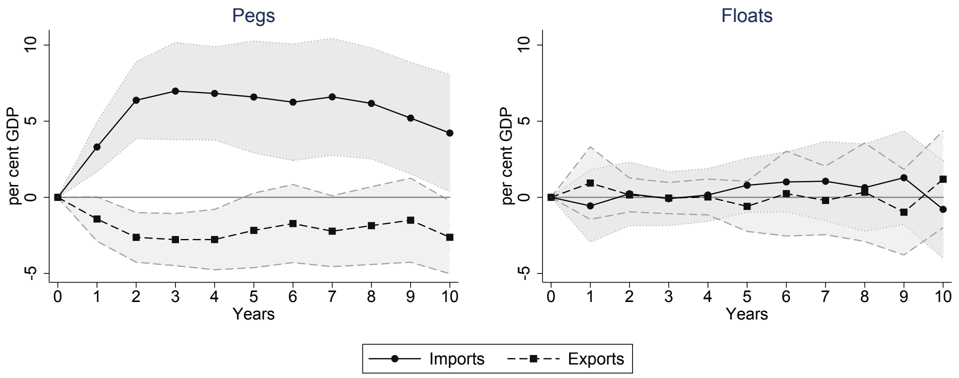
The shares refer to the year before the default episode.

This section provides additional results to help with interpretation of the GDP component cost decomposition in Section 4.7. Table 4.B.14 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 4.8a, 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.6 and Table 4.B.15 provide further detail on how the default cost varies according to the exchange rate regime. Figure 4.B.6 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.15 decomposes the cost for pegs into different components of GDP. As in the baseline specification (Table 4.7), most of the adjustment takes place via investment and imports.

Table 4.B.16 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.6. 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.15. 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.16. 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. 2018. “The Costs of Sovereign Default: Evidence from the Stock Market.” *Review of Financial Studies* 31 (5): 1707–1751.

Angrist, Joshua D., Òscar Jordà, and Guido Kuersteiner. 2017. “Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited.” *Journal of Business & Economic Statistics* 36 (3): 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 Jean-Sébastien Nadeau. 2015. “Database of Sovereign Defaults, 2015.” Technical Report 101.

Beers, David T., and John Chambers. 2006. “Default Study: Sovereign Defaults at 26-Year Low, to Show Little Change in 2007.”

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): 37–53.

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): 3119–3145.

Ilzetzki, Ethan, Carmen M Reinhart, and Kenneth S Rogoff. 2019. “Exchange Arrangements Entering the Twenty-first Century: Which Anchor Will Hold?” *Quarterly Journal of Economics* 134 (2): 599–646.

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* 134 (3): 1225–1298.

Jordà, Òscar, Björn Richter, Moritz Schularick, and Alan M. Taylor. 2021. “Bank Capital Redux: Solvency, Liquidity, and Crisis.” *Review of Economic Studies* 88 (1): 260–286.

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. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy* 126 (2): 850–901.

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. "From Financial Crash to Debt Crisis." *American Economic Review* 101 (5): 1676–1706.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011b. "The Forgotten History of Domestic Debt." *Economic Journal* 121 (551): 319–350.

Reinhart, Carmen M., and Christoph Trebesch. 2016. "The International Monetary Fund: 70 Years of Reinvention." *Journal of Economic Perspectives* 30 (1): 3–28.

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): 1029–1061.

Sosa-Padilla, Cesar. 2018. "Sovereign Defaults and Banking Crises." *Journal of Monetary Economics* 99: 88–105.

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.