Location, Location, Location: Long-run Trends in Housing Markets and Regional Economic Development

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch die

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

vorgelegt von

Martin Dohmen

aus Simmerath

Bonn

2022

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Moritz Schularick
Zweitreferent:	Prof. Dr. Benjamin Born
Tag der mündlichen Prüfung:	27. Juli 2022

Acknowledgements

I am grateful for all the generous support I received during writing this dissertation. First of all, I want to thank my supervisor and co-author Moritz Schularick for all the helpful feedback, support, freedom, resources and the environment that enabled me to make this work possible. Without all the challenges you gave me and all the demands for improvements and revisions this dissertation would definitely not have the quality it has right now and I would have learned far less.

I also want to thank my second supervisor Benjamin Born for providing an important second opinion on many questions and for always being friendly and providing an open ear for all the questions and concerns anytime. I would also like to thank all the other members of the current and former Bonn faculty – in particular Christian Bayer, Thomas Dohmen, Keith Kuester, Moritz Kuhn, Farzad Saidi and Lingwei Wu – for their advice and help at various stages of my research.

I am also deeply grateful for being part of team of the Macrofinance & Macrohistory Lab and all the feedback and advice I received from all members of the team - I think the atmosphere in the team was extraordinary great. A special thanks to Kaspar Zimmermann and Dmitry Kuvshinov, who have always been my in-official mentors and secret role models and were providing irreplaceable help especially during the earlier phase of my dissertation.

This dissertation originated within the framework of the Bonn Graduate School of Economics, the Institute for Macroeconomics and Econometrics and the Cluster of Excellence ECONtribute. I wish to acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy – EXC 2126/1 – 390838866 and by the DFG Research Training Group 2281, 'The Macroeconomics of Inequality'. I wish to thank Heide Baumung, Ravena Penning and Beate Roters for helping me deal with countless administrative matters. I also wish to thank numerous research assistants that supported me with all the data work that became part of this dissertation.

Maybe most importantly, I want to express my great thankfulness to my friend and co-author Francisco Amaral. I am deeply grateful to have been able to work with you and I am a little sorry for all the harder times you had to experience and endure during working with me. Together, however, we have always been able to find good solutions for seemly insolvable problems and also found the time to have a lot of fun.

I also do not want to forget my other co-author Sebastian Kohl. It was not only very helpful to have someone with a different background and a lot of knowledge about data collection in the team, but I think we also had a good time working together and at various research trips.

Finally, I would like to thank my friends Timo, Matthias, Luca, Martin K., Christina and all the other doctoral colleagues at the BGSE and ECONtribute, who made the PhD experience a pleasant one and helped me to relax and switch-off in-between nearly six years of work. I am also deeply grateful to my family, who always supported me and provided a safe harbour. A special thanks goes to Matilda, who shared the good and the bad times with me and always reminded me that there are other important things in life next to my dissertation.

Contents

List of Figures

List of Tables

Introduction

> *Und nachsinnend überdenken wir Freud' und Leid dieser Gegenden, die Schicksale nicht bloß der Menschen, nein, auch ihrer Werke, der Häuser, die sich da endlos vor uns dehnen, eins fast genau wie das andre, und von denen doch jedes seine eigene Entwicklung hat. Und wieder bleibt unser Gedanke haften an dem wogenden Auf- und Absteigen der Grundstückswerte, an dem so viele Menschenschicksale hängen, und wir sehen die Häuser an und fragen sie: was habt ihr hier draußen davon erlebt, was war euer Schicksal? Ja, wenn sie reden könnten! Von wie mancher glücklich ergatterten, unverdienten Million, die nun mit dem Schweiß und den Entbehrungen ganzer Generationen verzinst werden muß, aber auch von wie mancher mißglückten Spekulation, zerstörten Hoffnungen, verschwundenen Ersparnissen würden sie erzählen! Aber sie bleiben stumm, und wieder müssen wir das mühselige Werk beginnen, allerlei Zahlen aus allerlei Schriften zusammenzusuchen, um zu erfahren, was auf diesen Blättern des Schicksals geschrieben steht.*
>
> —Mangoldt (1907)

Economic activity is distributed unequally across space. Today's distribution is the outcome of developments that went on for several decades or even centuries. Between countries, the complex interaction of the so-called fundamental causes of economic growth seems to determine which places developed rapidly and which stayed behind. A large strand of economic literature tries to explain why some countries industrialized quickly and are rich today, whereas others are poor (Acemoglu, Johnson, and Robinson, 2005; Clark, 2014; Fuchs-Schündeln and Hassan, 2016). Within countries, the interplay between agglomeration and congestion forces is responsible for the geographic distribution of economic activity. Agglomeration forces drive the formation of large urban centers and smaller cities, whereas congestion forces ensure that economic activity does not collapse to a single point in space. This is the foundation of many spatial models of economic activity in the tradition of Rosen (1979) and Roback (1982) and a large body of empirical research in the urban economic literature (Glaeser, 2010; Ahlfeldt and Pietrostefani, 2019).

The most relevant congestion force is housing. The fact that people need shelter somewhere close to their workplace ensures that not all people are able to work at the same location. Higher housing prices cancel out rising wages in places of growing productivity, such that the unequal distribution of economic activity is mirrored in the unequal distribution of housing market outcomes. Simultaneously housing is also the single most important asset in household portfolios (Badarinza, Campbell, and Ramadorai, 2016) and housing wealth accounts for roughly one-half of national wealth in a typical economy (Piketty, 2014). This ensures that housing prices not only work as a proxy for the relative fortunes of different regions and as such are even able to predict local voting behavior (Adler and Ansell, 2019; Ansell, 2019), but are also fundamental to individual fortunes and shape the development of the wealth distribution (Kuhn, Schularick, and Steins, 2020). The local nature of the asset housing ties back together changes in the spatial distribution of economic activity and the dynamics of the wealth distribution.

This thesis tries to make some progress in understanding long-run trends in the spatial behavior of housing markets and economic development. To achieve this, it introduces and analyses new regional long-run data to uncover new stylized facts and combines state-of the art econometric techniques with parsimonious, target-oriented economic theory. The questions this thesis wants to advance on are: How do returns on housing investments vary across space (Chapter 1)? Does the nationwide fall in the risk-free rate have the potential to explain the increasing dispersion in housing prices relative to rents within developed countries (Chapter 2)? Can institutional innovations and an increase in market size help to explain why some regions in Europe industrialized quickly and are developed today whereas others stayed behind (Chapter 3)?

Chapter 1 - "Superstar Returns: The Geography of Housing Market Risk", which is joint work with Francisco Amaral, Sebastian Kohl and Moritz Schularick, documents, for the first time, substantial spatial variation in housing market return premia. For this purpose, the chapter introduces a novel international data set covering 27 large cities from 15 developed countries spanning a period of nearly 150 years. Next to city-level housing price and rent series, it contains annual total housing return series as the sum of capital gains and rent returns. The data set uncovers large variation in total housing returns across locations. Total returns in large "superstar" cities are close to 100 basis points lower per year than in the rest of the country. As shown in previous studies, house prices tend to grow faster in large cities. Rent returns, however, are substantially smaller within large cities, resulting in lower long-run total returns.

This key finding can be rationalized in a standard asset-pricing framework where excess returns are a compensation for housing investment risk. Suppose that everything that makes a large "superstar" city – its diversified economy, its large and liquid housing market, its productivity, its amenities – also makes it a safer place as an investment. A consequence would be that buyers are willing to pay a higher price and accept a lower return for housing investments in these cities. The rest of the chapter provides empirical support for this interpretation of the (negative) large city premium. On the one hand, we present evidence that the co-variance between housing returns and income growth is lower in large cities. On the other hand, we show that idiosyncratic housing risk is larger outside the large cities, which seems to be related to the fact that housing markets in large cities are considerably more liquid.

Chapter 2 - "The Fall in the Risk-Free Rate and Rising House Price Dispersion", also written jointly with Francisco Amaral, Sebastian Kohl and Moritz Schularick, offers a new explanation for the increasing dispersion of regional housing prices over the last decades. Existing explanations mainly from the urban economics literature - like Gyourko, Mayer, and Sinai (2013) or Nieuwerburgh and Weill (2010) - try to explain the rising spatial heterogeneity in housing prices with local housing market fundamentals. A combination of rising local demand and inelastic housing supply implies an increase in the value of housing services in supply-constrained cities. As housing prices, in equilibrium, equal the discounted future value of housing services, this translates into growing housing prices in these cities. Recent empirical literature and our new data, however, show that rent dispersion has increased considerably less than housing price dispersion. This limits the potential of local fundamentals to explain the rising dispersion in housing prices.

The chapter develops a new spatial Gordon growth model that shows that a nationwide fall in the risk-free rate combined with initial regional differences in rent-price ratios increases housing price dispersion without affecting rent dispersion. The initial difference in rent-price ratios is evident in our new data. In accordance with Chapter 1, we argue that it is probably due to lower local housing risk-premia in larger cities. We calibrate our model to the 1985 values assuming a difference in risk-premia that coincides with the return differences uncovered in Chapter 1. Under these assumptions, a uniform fall in discount rates of only 1.3 percentage points is able to generate the increase in levels as well as dispersion in housing prices observed in the data in 2018.

Chapter 3 - "Freedom of Enterprise and Economic Development in the German Industrial Take-Off" turns away from housing markets and instead analyzes differences in long-run trends of economic development. It contributes to the literature on the fundamental causes of economic growth by examining the interaction between institutional innovations and increasing market size as drivers of the industrial revolution. In the literature, institutions are debated as a fundamental cause of economic development, because inclusive institutions might be necessary to create the incentives required to invest in and industrialize production processes. On the other hand, demand needs to be sufficiently high and markets sufficiently large to enable increasing returns to scale technologies to break even, such that a sector industrializes.

In the Chapter, I use the division of the Kingdom of Westphalia as a natural experiment to show that only reforms in both dimensions combined stimulated economic growth during the German industrial take-off. Homogeneous counties were allocated quasi-randomly to Prussia or the Electorate of Hesse as part of a package deal at the Congress of Vienna. In Prussian counties, the *Gewerbefreiheit* (freedom of enterprise) and the abolition of guilds increased incentives to industrialize manufacturing production, yet these counties and those under the guild system developed similarly. Only the establishment of the German *Zollverein* (customs union), which increased market size considerably, enabled counties featuring the *Gewerbefreiheit* to experience significantly higher growth.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "Chapter 6 Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth*. Elsevier, 385–472. DOI: 10.1016/s1574-0684(05)01006-3. [1]

Adler, David, and Ben Ansell. 2019. "Housing and Populism." *West European Politics* 43 (2): 344–65. DOI: 10.1080/01402382.2019.1615322. [2]

Ahlfeldt, Gabriel M., and Elisabetta Pietrostefani. 2019. "The Economic Effects of Density: A Synthesis." *Journal of Urban Economics* 111 (May): 93–107. DOI: 10.1016/j.jue.2019.04.006. [1]

Ansell, Ben W. 2019. "The Politics of Housing." *Annual Review of Political Science* 22 (1): 165–85. DOI: 10.1146/annurev-polisci-050317-071146. [2]

Badarinza, Cristian, John Y. Campbell, and Tarun Ramadorai. 2016. "International Comparative Household Finance." *Annual Review of Economics* 8 (1): 111–44. DOI: 10.1146/annurev-economics-080315-015425. [1]

Clark, Gregory. 2014. "The Industrial Revolution." In *Handbook of Economic Growth*. Elsevier, 217–62. DOI: 10.1016/b978-0-444-53538-2.00005-8. [1]

Fuchs-Schündeln, N., and T.A. Hassan. 2016. "Natural Experiments in Macroeconomics." In *Handbook of Macroeconomics*. Elsevier, 923–1012. DOI: 10.1016/bs.hesmac.2016.03.008. [1]

Glaeser, Edward L., editor. 2010. *Agglomeration Economics*. University of Chicago Press. [1]

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar Cities." *American Economic Journal: Economic Policy* 5 (4): 167–99. DOI: 10.1257/pol.5.4.167. [3]

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins. 2020. "Income and Wealth Inequality in America, 1949-2016." *Journal of Political Economy* 128 (9): 3469–519. [2]

Mangoldt, Karl von. 1907. *Die städtische Bodenfrage: eine Untersuchung über Tatsachen, Ursachen und Abhilfe*. Vol. 8, Die Wohnungsfrage und das Reich. Vandenhoeck & Ruprecht. [1]

Nieuwerburgh, Stijn van, and Pierre-Olivier Weill. 2010. "Why Has House Price Dispersion Gone Up?" *Review of Economic Studies* 77 (4): 1567–606. DOI: 10.1111/j.1467-937x.2010.00611.x. [3]

Piketty, Thomas. 15, 2014. *Capital in the Twenty-First Century*. Harvard University Press. URL: https://www.ebook.de/de/product/21683807/thomas_piketty_capital_in_the_twenty_first_century.html. [2]

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78. DOI: 10.1086/261120. [1]

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." *Current Issues in Urban Economics*, 74–104. [1]

Chapter 1

Superstar Returns: The Geography of Housing Market Risk*

Joint with Francisco Amaral, Sebastian Kohl, and Moritz Schularick

1.1 Introduction

Residential real estate is the most important asset in household portfolios, the main collateral of bank lending, and plays a central role in current macroeconomic models of aggregate fluctuations in which asset structure and household borrowing interact with business cycle fluctuations and monetary policy (Mian and Sufi, 2011; Berger, Guerrieri, Lorenzoni, and Vavra, 2018; Kaplan, Moll, and Violante, 2018; Cloyne, Ferreira, and Surico, 2020). While housing markets were a side show for a long time, they are now at the center of a research agenda that studies the con-

* This work is part of a larger project supported by the European Research Council Grant (ERC-2017-COG 772332). We are indebted to a large number of researchers who helped with the data on individual cities, especially Jacques Friggit, Matthijs Korevaar, Avner Offer and Chihiro Shimizu. We are grateful to Andrea Balestracci, Melina Cosentino, Gregor Jost, Dimitrios Kanelis, Gerwin Kiessling, Liudmila Kiseleva, Carla Revilla, Markus Schick and Masashi Tanigaki for their excellent research assistance. For helpful feedback we thank Gabriel Ahlfeldt, Benjamin Born, Esther Eiling, Marco Giacoletti, Arpit Gupta, Christian Hilber, Nir Jaimovich, Christine Laudenbach, Lara Loewenstein, Ronan Lyons, Jonathan Moreno Medina, Stijn Van Nieuwerburgh, Marco Pagano, Monika Piazzesi, Joseph E. Stiglitz, Daniel Sturm, and Sofie Waltl. We wish to acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy – EXC 2126/1 – 390838866. We also thank conference participants at the "Housing in the 21st Century" Conference and at the 15th North American Meeting of the Urban Economics Association and workshop participants of the "Advances in Research on Housing Markets" workshop and of the "Financial History Workshop 2021" in Antwerp as well as seminar participants at the Harvard Joint Center for Housing Studies, the LSE Urban and Regional Economics Seminar and the Bonn Finance and Insurance ECONtribute Seminar for valuable comments. All remaining errors are our own. The views expressed herein are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of New York or the Federal Reserve System.

sequences of heterogeneous household portfolios for the macroeconomy providing valuable insights into the transmission of shocks and the dynamics of the wealth distribution (Favilukis, Ludvigson, and Van Nieuwerburgh, 2017; Kaplan, Mitman, and Violante, 2020; Kuhn, Schularick, and Steins, 2020; Greenwald, Leombroni, Lustig, and Van Nieuwerburgh, 2021). Yet despite the importance of housing for household portfolios, the risk and return properties of housing investments remain largely unexplored.

What are the risks a homeowner faces when buying real estate in a specific market? How do expected returns on housing assets vary across space? Some progress has been made in recent years in macrofinance and economic history to quantify long-run returns on different asset classes, including nation-wide residential real estate markets (Shiller, 2015; Knoll, Schularick, and Steger, 2017; Jordà, Knoll, Kuvshinov, Schularick, and Taylor, 2019). However, households do not hold diversified claims on parts of the national housing stock, but individual properties in specific locations. The "housing market" is a collection of markets that differ by many attributes (Piazzesi, Schneider, and Stroebel, 2020). It is often said that "location, location, location" is the number one rule of the real estate business. So far we know very little about the spatial distribution of housing returns and the corresponding risk profiles of household portfolios across locations. Such within asset class differences in returns potentially hold important insights for a growing literature on the role of asset returns for the dynamics of wealth inequality (Fagereng, Guiso, Malacrino, and Pistaferri, 2020), but also for spatial economic models as the sorting of workers across locations can have important effects on household portfolio risk.

In this paper, we make the first comprehensive quantitative attempt to study the heterogeneity of return premia in housing markets. We introduce and analyze a new city-level data set covering the main agglomerations in 15 OECD countries over the past 150 years. It contains annual house prices and rents for 27 large cities.[1] For each city, we calculate long-run total returns on residential real estate investments as the sum of price appreciation and rent returns – and compare them to returns in other parts of the country. For the construction of the data set, we could partly draw on existing historical research for individual cities. In most cases, however, we hand-collected new house price and rent series from city yearbooks or primary sources such as newspapers, tax records, and notary archives.[2]

Our central finding is that, over the long-run, there exists systematic spatial variation in returns between large cities and other parts of the same country. The large agglomerations have witnessed lower total returns on residential real estate. With

1. These cities are sometimes referred to as "superstar cities." The superstar terminology goes back to a well-known paper by Gyourko, Mayer, and Sinai (2013) for the U.S. We study the main economic agglomerations in 15 countries.

2. The construction of the series and their sources are documented in a comprehensive Data Appendix.

5.72 log points per year, average total returns have been smaller compared to the national average of 6.68 log points. In other words, an investment in large cities comes with a negative return premium of approximately 95 basis points annually relative to national returns (including the largest cities) and about 100 basis points lower than the rest of the country (excluding the largest cities). These return differences are a robust feature of the data across countries and time periods, and statistically highly significant. A negative return premium of around 1 percentage point accumulates to substantial return differences in the long run. For instance, an investment in the large city portfolio earned only about half the cumulative return than the national portfolio over the past 70 years.

We corroborate this finding by studying the U.S. and German housing markets, two economies with different housing market structures and policy regimes, for which we have comprehensive return data across the entire city-distribution for the post-World War II period. For the U.S., we combine the data set constructed by Gyourko, Mayer, and Sinai (2013) with data from the American Community Survey for the 2010-2018 period. For Germany, we hand-collected a data set on housing returns covering 127 small and large German cities. In both countries we find that total returns to housing decrease with city size. The return premium of small vs. large MSAs in the U.S. amounts to about 80 basis points annually and about 60 basis points in Germany. Both estimates are statistically highly significant.

What explains spatial variation in housing return premia? Our key finding can be rationalized in a standard framework where excess returns are a compensation for risk. Observable long-run return differences between different assets must be attributable to differences in risk, or to violations of standard assumptions (such as persistent behavioral biases in expectations). The urban economics literature has documented how large agglomerations differ from the rest of the country (Black and Henderson, 1999; Eeckhout, Pinheiro, and Schmidheiny, 2014; Desmet and Henderson, 2015). Suppose that everything that makes a large "superstar" city – its diversified economy, its large market, its productivity, its amenities, the international linkages – also makes it a safer place as an investment. A consequence would be that the present value of future housing services will be subject to less risk so that buyers are willing to pay a higher price and accept a lower return for housing investments in large agglomerations. In turn, higher returns outside the large cities would be a compensation for higher risk. For remote locations to attract capital, they have to offer higher returns.

The second part of the paper provides theoretical and empirical support for this interpretation of the (negative) premium on large city real estate. On the one hand, the co-variance between housing returns and income growth is lower in large cities. Between 1950 and 2018, the co-variance between U.S. MSA-level income growth and MSA-level housing returns has been significantly larger in smaller MSAs. On the other hand, households typically do not hold diversified housing portfolios and, therefore, are also exposed to idiosyncratic risk. We show that idiosyncratic housing risk is considerably higher outside the large cities. Using U.S. transaction-level data from Corelogic, we find that the idiosyncratic component of housing risk decreases with MSA size. As liquidity is low, home owners in thinner markets face a greater risk of not realizing the local market return at the point of sale. Real estate search engine data confirm a significant increase of housing liquidity with city size. These findings mesh with recent work by Giacoletti (2021), Sagi (2021) and Kotova and Zhang (2019) who show a strong relationship between idiosyncratic risk and housing market liquidity.

The result that large cities witnessed lower housing returns might seem counterintuitive at first. House price appreciation in cities like New York, London, Paris, Sydney, or Amsterdam has been eye-catching in recent decades. In many countries, the gap between the highest priced locations and the rest of the country has grown (Arundel and Hochstenbach, 2019), contributing to the perception of an increasing economic and social gap between the successful agglomerations and more remote areas (Ansell, 2019).

We confirm that house price appreciation tends to be higher in large cities than in the rest of the country. Over the long run, we estimate an average annual difference of up to 60 basis points annually between our 27 agglomerations and other parts of the country (albeit with mixed statistical significance). At the same time, rent returns are considerably higher outside the large cities. The negative spatial correlation between capital gains and rent returns has also been documented for the largest 30 U.S. MSAs between 1986 and 2014 by Demers and Eisfeldt (2021). Our long-run estimates put the mean rent return differential at 160 basis points per annum. The differences in rent returns exceed the differences in capital gains in the long run and lead to lower total housing returns.

Differences in housing risk can rationalize the divergent patterns in capital gains and rent returns. Assuming housing risk is lower in large cities, investors will be willing to pay a higher price for the safer rental cash-flow in large cities. In equilibrium, the difference in rent returns between small and large cities will be bigger than the difference in capital gains. At the same time, an increase in housing demand leads to higher house price appreciation in supply-constrained superstar cities compared to smaller cities (Gyourko, Mayer, and Sinai, 2013; Hilber and Vermeulen, 2016), explaining higher price gains in the large cities *and* lower returns overall.

We perform a large number of robustness checks to back-up our key results. We use different rental yield benchmarks, study different sub-periods and the effects of rent regulations, and vary the definitions of large cities. *First*, as our core finding is driven by substantial differences in rent returns, we rebuild our main data set using independent, country specific, current day rental yield benchmarks. The overall results remain very similar. *Second*, although we are interested in long-run returns, we want to make sure that they are not driven by specific time periods. We separate the early historical parts of the sample, and also split the sample period in 1990. The same patterns can be found in the historical period as well as during the last three decades. *Third*, we divide our data set into different rent regulation regimes. It turns out that our results are not driven by periods with strict rent controls. On the contrary, the differences in capital gains are highest during periods of strict rent controls and lowest for total returns. *Last*, we tried different definitions of large cities, and experimented with different size cut-offs in different eras. Once more, none of this altered the core findings in a systematic way.

Previous literature: Our work is the first to put together international long-run return series for different cities and regions. This adds a regional dimension to the existing literature on long-run house prices (Knoll, Schularick, and Steger, 2017), and returns on housing portfolios (Jordà, Knoll, et al., 2019) and an international dimension to individual papers on long-run housing returns in individual regions (Eichholtz, Korevaar, Lindenthal, and Tallec, 2020; Keely and Lyons, 2020; Demers and Eisfeldt, 2021). It also helps alleviate the lack of data that remains a central challenge to research on housing markets (Piazzesi, 2018) and asset pricing with housing (Lustig and Van Nieuwerburgh, 2005; Piazzesi, Schneider, and Tuzel, 2007; Jordà, Schularick, and Taylor, 2019).

Our paper also complements the existing urban economics literature by bringing together house price data with rental yields, housing returns and measures of local housing risk. While the existing literature has focused on the spatial distribution of economic activity (Glaeser, 2010) and implications for house prices (Saiz, 2010; Gyourko, Mayer, and Sinai, 2013; Hilber and Vermeulen, 2016), we show that the spatial distribution of housing returns is different – a fact that we explain with differences in local housing market risks. We uncover another consequence of agglomeration: Other than having higher productivity and wage levels (Ahlfeldt and Pietrostefani, 2019), better consumption amenities (Diamond, 2016), less concentrated labour markets (Desmet and Rossi-Hansberg, 2013), higher elasticities of urban costs (Combes, Duranton, and Gobillon, 2019) and more diversified industry compositions (Duranton and Puga, 2000), large cities also feature less housing investment risk.

Finally, our paper is part of a nascent literature on the risk-return relation in housing markets (Demers and Eisfeldt, 2021; Giacoletti, 2021; Sagi, 2021), and complements recent work by Hilber and Mense (2021), who show that higher cyclical increases in price-rent ratios in superstar cities are due to lower supply elasticities and serially correlated housing demand shocks. Our paper points to persistent differences in price-rent ratios between cities.

The paper is organized as follows. The next section describes our new long-run data set and provides an overview of the series and various consistency checks (also see the detailed documentation in the Data Appendix). In the third section, we describe the main long-run stylized facts emerging from our data set and compare city-level and national housing returns to establish our key finding that total returns are lower in large cities. The next section introduces two granular data sets for the U.S. and Germany and studies housing returns over the entire city-size distribution in both countries. In section five, we turn to the differences in housing risk as an explanation for the return differences. Using multiple U.S. data sets, we show that housing risk is lower in large cities, both in terms of co-variance risk between excess returns and local income as well as due to smaller idiosyncratic shocks in more liquid markets. The last section concludes.

1.2 A new long run city-level housing returns data set

This section introduces our new historical city-level data set. The data covers 27 cities over the long run: London, New York, Paris, Berlin, Tokyo, Hamburg, Naples, Barcelona, Madrid, Amsterdam, Milan, Melbourne, Sydney, Copenhagen, Rome, Cologne, Frankfurt, Turin, Stockholm, Oslo, Toronto, Zurich, Gothenburg, Basel, Bern, Helsinki, and Vancouver. Figure 1.1 shows the geographical distribution of the cities included in our long-run sample. Our city-level data set contains house prices and rents as well as rental yields for every city. In the following, we briefly discuss the criteria we employed for the choice of cities and the methods used to construct the series. Details on sources for each city can be found in the Data Appendix.

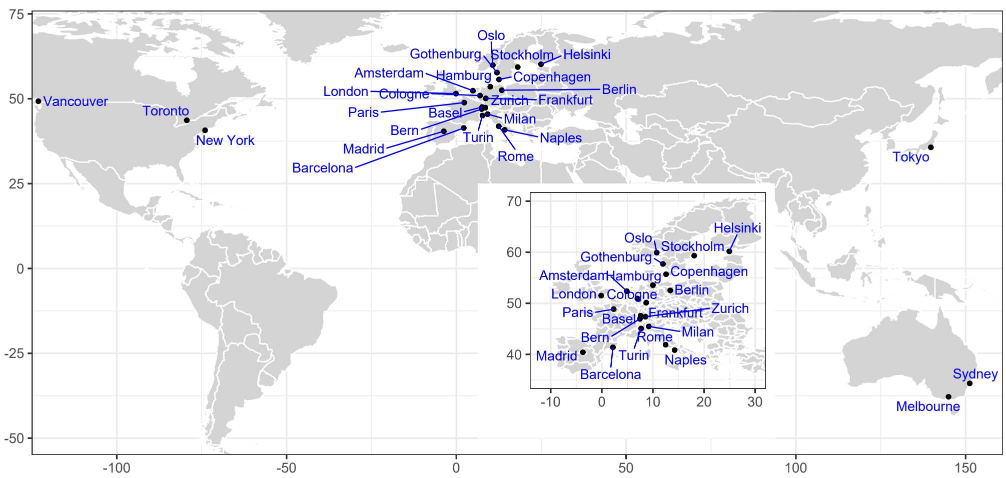

Figure 1.1. Geographical distribution of our city sample

Notes: Latitude and longitude are given on the y- and x-axis, respectively. The map was built using the shape file in Becker et al. (2018).

1.2.1 City sample

We focused our data collection on the largest cities within 15 developed countries that are covered by the Jorda-Schularick-Taylor Database and Jordà, Knoll, et al. (2019) as we want to compare our data with national data contained in the database.

We also rely on the database for macroeconomic data such as long-run consumer price indices.

For each country, we define the largest cities in terms of 1900 population and include cities with a population share of more than 1% in 1900. To the extent possible, we also aimed to cover at least 10% of the 1900 country population in order to analyze a relevant share of the countries' housing markets.[3] Selecting cities based on the population in 1900, instead of using current population, circumvents the problem of survivorship bias. A detailed discussion of city choice by country is provided in the Data Appendix. Urban systems evolve over time and so do the boundaries of cities. Over time, all cities and local housing markets grow either through incorporation of more and more suburbs or through the creation of metropolitan regions. We follow the administrative definitions in our sources which makes our city definition consistent *within* country. City definitions are mostly identical for the rental and ownership markets.

The sample is summarized in Table 1.1. Data coverage of price and rent data is shown in columns 5 and 6. The sample starts in 1870, but some gaps remain. We have 7 decades of data for all cities and a balanced panel for the post-1950 period. Column 3 shows the cities' share of the country population in 1900 and column 4 the aggregated share of country population in 1900 that is covered by our sample cities.

1.2.2 Sources and methodology

This section describes the sources of the data and the construction of the total return series. For all cities in our sample, we construct annual house price indices, rent indices and calculate total return series.

1.2.2.1 House price and rent indices

Whenever possible and of sufficient quality, we use house price and rent indices from existing research. An example is the return series for Amsterdam described in

3. We included fewer cities in countries such as France where a dominant urban center was well established by the 19th century. In some cases, there were not enough cities with more than 1% of population to reach the 10% target. We then included cities in descending order of total 1900 population, starting with the largest city. For some countries, however, we were forced to deviate from this rule for specific reasons. An extreme example of this is the case of Germany. The list of the largest cities in 1900 is (in descending order): Berlin, Hamburg, Dresden, Leipzig, Munich, Cologne, Wroclaw and Frankfurt. Of these cities, only Berlin and Hamburg hit the 1% target. The geographical area of Germany, however, changed drastically several times after 1900. This means that we do not include Wroclaw, which no longer belongs to Germany and Leipzig and Dresden, which were part of Eastern Germany between 1945 and 1990 and hence market price and rent data is missing for a considerable time period. From the remaining cities, there does not exist sufficient data coverage for Munich. To still get close to the 10% target and as Germany covered a larger area in 1900 compared to today, we chose to include all cities up to Frankfurt in our sample.

Table 1.1. City choice and data coverage

City	Pop1900	Share pop	Country	House prices	Rents
London	6480	0.157	0.157	1895–2018	1870–2018
New York	4242	0.056	0.056	1920–2018	1914–2018
Paris	3330	0.082	0.082	1870–2018	1870–2018
Berlin	2707	0.048	0.078	1870–2018	1870–2018
Tokyo	1497	0.034	0.034	1950–2018	1950–2018
Hamburg	895	0.016	0.078	1870–2018	1870–2018
Naples	563	0.017	0.054	1950–2018	1950–2018
Barcelona	552	0.030	0.059	1950–2018	1947–2018
Madrid	539	0.029	0.059	1950–2018	1947–2018
Amsterdam	510	0.099	0.099	1870–2018	1870–2018
Milan	491	0.015	0.054	1950–2018	1950–2018
Melbourne	485	0.130	0.257	1880–2018	1901–2018
Sydney	478	0.128	0.257	1880–2018	1901–2018
Copenhagen	462	0.180	0.180	1938–2018	1885–2018
Rome	438	0.013	0.054	1950–2018	1950–2018
Cologne*	437	0.008	0.078	1902–2018	1890–2018
Frankfurt*	350	0.006	0.078	1897–2018	1895–2018
Turin	330	0.010	0.054	1950–2018	1950–2018
Stockholm	300	0.059	0.084	1875–2018	1894–2018
Oslo	227	0.102	0.102	1870–2018	1892–2018
Toronto	205	0.038	0.050	1900–2018	1921–2018
Zurich	150	0.045	0.098	1905–2018	1890–2018
Gothenburg	130	0.025	0.084	1875–2018	1914–2018
Basel	109	0.033	0.098	1912–2018	1889–2018
Helsinki	97	0.037	0.037	1946–2018	1946–2018
Vancouver*	69	0.013	0.050	1950–2018	1950–2018
Bern	64	0.019	0.098	1912–2018	1890–2018

Notes: Column 2 shows city-level population in 1900 in 1000 inhabitants. Column 3 describes the share of each city's population of total country population in 1900. Column 4 describes the cumulative share from all cities in a respective country in our data set. Columns 5 and 6 describe data coverage from earliest to latest year of price and rent indices in our data set. For some cities there are gaps in the data coverage because of missing data, e.g. during periods of war and hyperinflation, see the Data Appendix. City-level population data is taken from Reba, Reitsma, and Seto (2016) and country-level population from Jordà, Schularick, and Taylor (2017). For Cologne and Frankfurt, city-level population was below 1% of country population in 1900. However, the German Empire in 1900 had a considerably different area compared to Germany today. In 1950, the population in both Frankfurt and Cologne was above 1% of Germany's total population. The estimate for Vancouver is taken as the sum of Burrard and Vancouver city from the Canadian population census from 1901. Burrard became officially part of Vancouver in 1904.

Eichholtz, Korevaar, Lindenthal, and Tallec (2020). In most cases, however, house price and rent indices are not readily available or the quality is insufficient. To construct the series, we first used data from a broad range of secondary sources such as city yearbooks, but in many cases we had to hand-collect new data from diverse primary sources. These consisted of newspapers, tax records, notaries, archives of real estate agents, and diverse other archival data. About half of the series are newly constructed.

The criteria to select appropriate sources mainly depended on data representativeness and availability. Whenever we had multiple choices, we used the source which provided the best coverage and the most details. The case of London provides an illustration where we could partly rely on data from previous research but had to close a large gap after World War II. The existing house price series cover the years before 1946 and after 1969. To connect the series, we hand-collected asking prices from real estate advertisement sections in newspapers. We focused on sales ads that provided enough information to build quality-adjusted indices. Figure 1.2 panel (a) shows an example of the property ad section of the *Kensington Post* in 1965. The marked advertisements are used in our final index.

(a) London, 1965 **(b)** Frankfurt, 1895-1910

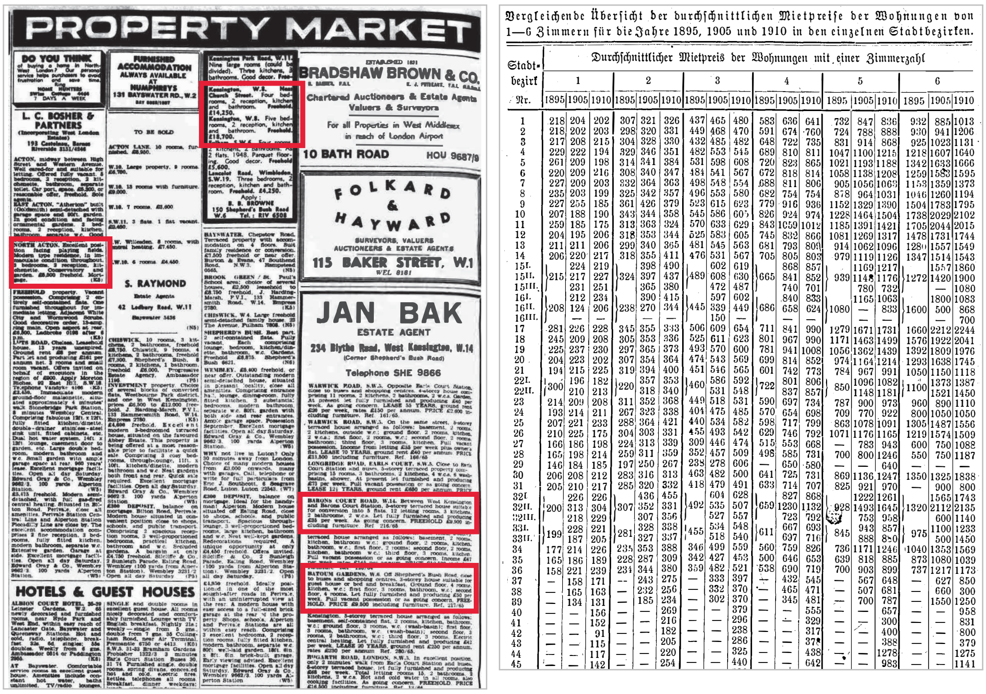

Figure 1.2. Examples of primary and secondary sources

Notes: Panel (a): Extract of the real estate part of the ad section for the 14th of May 1965 from the newspaper *Kensington Post*. Panel (b): From *Beiträge zur Statistik der Stadt Frankfurt am Main 11. NF (1919)* published in Busch (1919).

Index-construction depends on the type and quality of the available data. Whenever micro-data was available, we relied on repeat-sales or hedonic regression methods. For instance, for Frankfurt we built a hedonic house price index from 1960-2018 using transaction level data from public sources and their archives. Whenever micro-level data was not available, we used data disaggregated by housing types and location inside a city to construct stratification indices.

Regarding the construction of rent indices, one concern is that the rise of urban home ownership might have made rents less important than they have been historically. In cities like Oslo or Rome, private rentals occupy a small segment of the market, and the majority of homes are in private ownership. However, Figure 1.3 shows that the majority of urban households are tenants, as in all German-speaking cities but also in New York and Paris.

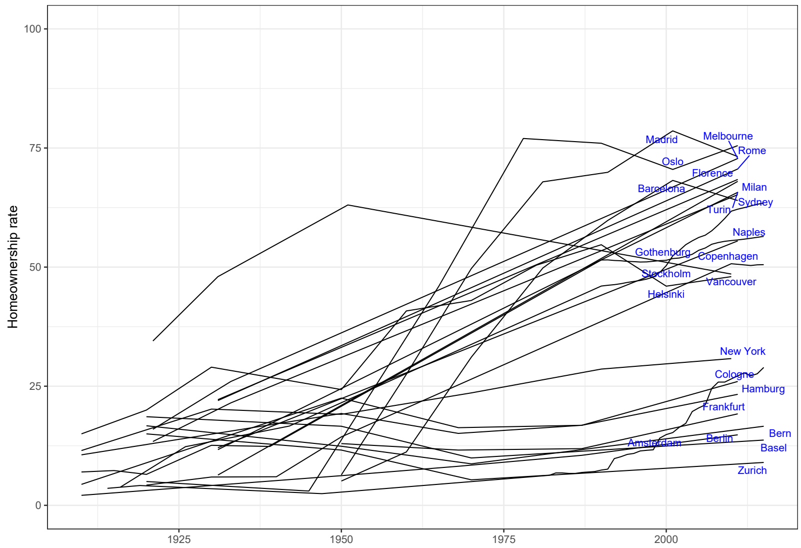

Figure 1.3. Urban home ownership rates, long sample cities

Notes: The city definition refers to administrative cities, not metropolitan regions. Source: Kohl and Sørvoll (2021).

That said, rent data was often harder to find than house price data. We rely on rent indices from statistical agencies as rent indices that were constructed by city statistical offices for (city-level) CPI data. They mainly use repeated rents methodology. In other cases, when we were able to collect micro-level data, we relied on hedonic methods. For example, for the city of Oslo, we constructed a hedonic rent index for the period between 1950 and 1970 from newspaper rental advertisements. In other cases, we constructed stratification indices whenever possible, mainly relying on statistical publications. For example, in the case of Stockholm we used average rent by size of dwelling to construct a chained stratification rent index.

In general, the quality of house price data improves over time.[4] For rent data, whenever the quality is not already high, we benchmark our rent indices with rents surveyed in housing censuses. Historically, such censuses were taken roughly every

4. After 1970, the majority of house price indices rely on hedonic, repeat sales or SPAR methods.

ten years and typically covered all rental units, providing a precise picture of the universal level of rents in a specific city. Figure 1.2 panel (b) depicts an example of data for Frankfurt from historical housing censuses. In this example, the publication contains data on average rent prices for apartments by number of rooms and by city neighborhood, the so-called *Stadtbezirke*. We use this data to benchmark our rent series between 1895 and 1910. Table 1.2 displays an overview of the new series we constructed including the sources we used. We collected additional data on 20 out of our 27 sample cities; 13 of them had to be constructed from scratch.

Table 1.2. Overview of the new series

City	Series	Period	Source	City	Series	Period	Source
London	house	1946-1969	newspaper	Cologne	house	1966-2018	trans. records
London	rent	1946-1998	newspaper	Cologne	rent	1904-1972	stat. yearbook
Paris	house	1950-1958	newspaper	Cologne	rent	1973-2018	market reports
Berlin	house	1870-1964	stat. yearbook	Frankfurt	house	1897-1959	stat. yearbook
Berlin	house	1965-2018	trans. records	Frankfurt	house	1960-2018	trans. records
Berlin	rent	1870-2018	stat. yearbook	Frankfurt	rent	1895-1965	stat. yearbook
Tokyo	house	1950-1975	newspaper	Frankfurt	rent	1972-2018	market reports
Hamburg	house	1870-1970	stat. yearbook	Turin	house	1927-1996	stat. yearbook
Hamburg	house	1971-2018	market reports	Turin	rent	1927-1996	stat. yearbook
Hamburg	rent	1870-1966	stat. yearbook	Stockholm	rent	1894-2018	stat. yearbook
Hamburg	rent	1972-2018	market reports	Oslo	rent	1950-1970	newspaper
Naples	house	1927-1996	stat. yearbook	Toronto	house	1900-1991	newspaper
Naples	rent	1927-1996	stat. yearbook	Toronto	rent	1921-1991	newspaper
Barcelona	house	1960-2008	newspaper	Zurich	house	1905-2018	stat. yearbook
Milan	house	1956-1966	newspaper	Zurich	rent	1915-2018	stat. yearbook
Milan	house	1967-1996	stat. yearbook	Gothenburg	rent	1914-2018	stat. yearbook
Milan	rent	1950-1996	stat. yearbook	Basel	house	1912-1981	stat. yearbook
Rome	house	1927-1996	stat. yearbook	Basel	rent	1920-2018	stat. yearbook
Rome	rent	1914-1996	stat. yearbook	Bern	house	1912-2018	stat. yearbook
Cologne	house	1870-1965	stat. yearbook	Bern	rent	1915-2018	stat. yearbook

Notes: This table lists all new series we constructed ourselves. Some of these series we had to construct from scratch, others were taken from contemporaneous statistical publications, which we combined to build long-run indices. More details about the sources and methods used to construct these series and on all the other series from various authors we used can be found in the Data Appendix.

To build the new series we follow the handbook on residential property price indices from Eurostat (2013) and the methodology in Hill (2012). All price and rent indices are deflated using country-level CPI data from Jordà, Schularick, and Taylor (2017). For details on source and index construction by city please refer to the Data Appendix. The resulting series cover a representative city-level housing portfolio that approximates the behavior of the value weighted housing market within a city. This being said, for some cities and time periods we are only able to cover specific market segments due to data limitations. We are then making the assumption that the market segment trends are representative of trends in the city overall.[5]

5. In appendix 1.A.1 we compare hedonic house price indices for different market segments for Cologne. We show that over a period of 30 years, trends for all residential market segments have been similar.

1.2.2.2 Housing return series

We use our house price and rent indices to construct housing returns series. As an asset, a house delivers two types of returns. First, the price of a house can change and this generates a capital gain (or loss). Secondly, a house delivers a consumption stream in form of housing services. These can be sold to receive a cash flow by renting out the house. Alternatively, they can be consumed; in this case the owner receives the replication value as a cash-flow. Total returns on housing can be computed as:

$$\text{Total return}_t \quad = \quad \underbrace{\frac{P_t - P_{t-1}}{P_{t-1}}}_{\textbf{Capital gain}} + \underbrace{\frac{R_t(1-c)}{P_{t-1}}}_{\textbf{Net rent return}}, \tag{1.1}$$

where P_t is the house price at time t, R_t is the gross rent payment at time t and c are the total net operating costs as a share of R_t, which we describe in more detail below. Following this equation, the construction of city-wide (real) capital gains is straightforward using our house price indices. To construct rent return series, we estimated rent-price ratios, which we adjusted to nominal house price growth in the following manner: $\textit{Rent return} = \frac{R_t}{P_t} * \frac{HPI_t^{nom}}{HPI_{t-1}^{nom}}$.

Rent-price ratio estimates are constructed following the rent-price approach used in Jordà, Knoll, et al. (2019) and Brounen, Eichholtz, Staetmans, and Theebe (2013). To do so, we first use benchmark rent-price ratios for the end of our sample period in 2018. We again follow Jordà, Knoll, et al. (2019) and use benchmarks calculated from realized net operating income yields of real estate investors. These were provided by *MSCI* that collect data from a variety of real estate investors for large cities around the world. Yields are defined net of total operating costs, which are composed of maintenance and property taxes as well as other costs. Other costs included are management costs as well as cost of vacancies, letting and rent review fees, ground rents and bad debt write-offs. Finally, we use our rent and price indices to calculate rent-price ratios over time:

$$\frac{RI_{t+1}}{HPI_{t+1}} \quad = \quad \left(\frac{RI_{t+1}/RI_t}{HPI_{t+1}/HPI_t}\right)\frac{RI_t}{HPI_t}. \tag{1.2}$$

The disadvantage of this methodology is that possible measurement errors accumulate over time due to extrapolation. To account for this, we collected historical rent-price ratios to verify our rental yield series. Whenever the rent-price approach estimates diverge from these historical sources, we adjust the estimates to the historical measures of rent-price ratios as detailed in the Data Appendix.

For historical rental yield benchmarks, we predominantly relied on secondary sources or newspapers. For all sources, we aimed at collecting rental yield estimates out of rent and price data for the same buildings. All benchmark rent-price ratios are constructed net of depreciation and running costs. Whenever direct estimates for these costs were not available, we rely on estimates for depreciation and running costs in percentage of gross rent inside the country in question from Jordà, Knoll, et al. (2019). Another potential bias in our return series could arise from the ratio of net to gross income. Although we control for regional differences in the ratio of net to gross rental income by benchmarking our series to the 2018 *MSCI* estimates, it is not clear that these differences stayed constant over time. Nevertheless, evidence in Jordà, Knoll, et al. (2019) and Demers and Eisfeldt (2021) shows that the ratio of net to gross stayed relatively constant over time and that there are very small differences across regions over the last 30 years.

Last but not least, throughout the paper we follow the existing literature and measure housing returns in log points instead of percentage points. The main reason is that log returns are time compoundable, whereas percentage returns are not. Moreover, log returns have preferable distributional features and are approximately equal to percentage returns for small numbers. For a full rationalization please refer to appendix 1.A.2.[6]

1.3 Returns in large cities

In this section, we first establish the main stylized facts on long-run housing returns in large cities. We then proceed to analyze trends in capital gains and rent returns, as well as their contributions to total returns, and compare large cities to the rest of the country.

We start with summary statistics on real log housing returns and its components for our new data set. The left-hand panel of Figure 1.4 shows average log housing returns for the full time period and the right-hand panel for the period post 1950.[7] City-level total housing returns have been in the four to six log point range per year, with some differences across the cities in our sample. Toronto, Amsterdam, Gothenburg, Tokyo and Sydney are the cities with the highest long-run returns. The panel on the right shows that housing returns have been higher in the post 1950 period and reached about 6 log points.

Figure 1.5 plots the distribution of annual log real housing returns for the pre- and post-1950 period. While housing returns were on average lower in the pre-1950 period, they also displayed a higher standard deviation than in the post-1950 period,

6. To briefly see why, consider the following example: In city A house prices increase by 50% in period 1 and fall by 1/3 in period 2, in city B house prices stay constant. Using simple returns, average capital gains in city A are approximately 8.3% per year, but zero in city B. In fact, after two periods, prices in both cities are the same as they initially were and an investor holding a house for both periods realized a capital gain of zero. Using log returns, average capital gains for both example cities are zero.

7. Appendix Table 1.A.2 shows summary statistics by city in numbers including standard deviations. Appendix Table 1.A.3 adds summary statistics with average percentage point (simple) returns for comparison to other literature.

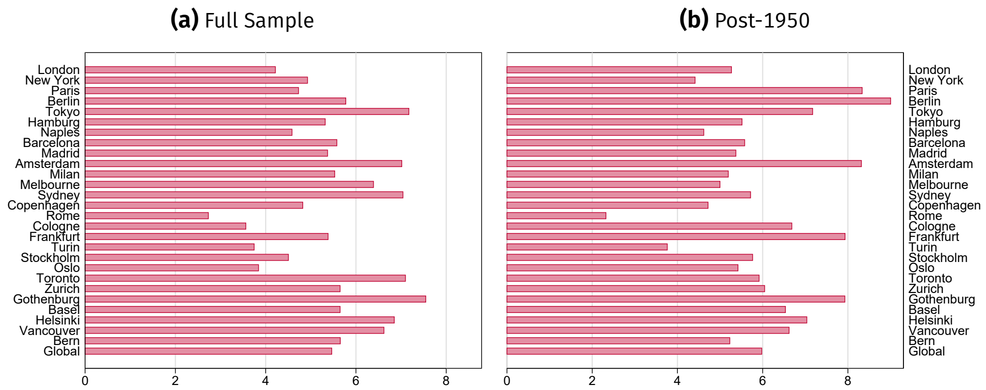

Figure 1.4. City-level real average total housing returns (log points)

Notes: The figure shows average real total housing returns in log points for all cities in our main sample. The series have been deflated using the national CPI series from Jordà, Schularick, and Taylor (2017). Panel (a) covers the entire sample for return data in our main data set, which is the subset of years for which rent and house price data (minus 1 year) exist, compare Table 1.1. Panel (b) shows average housing return data by city starting in 1950.

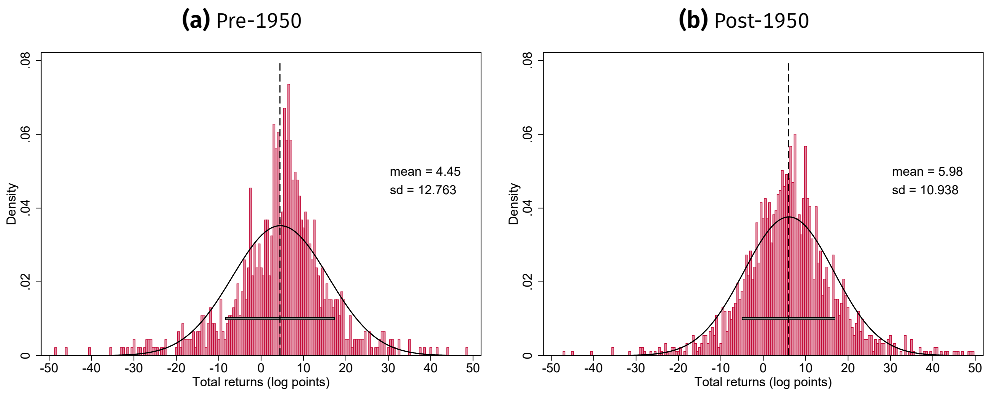

Figure 1.5. Distribution of annual real housing returns (log points)

Notes: The figure shows the distribution of annual total housing returns in log points for all cities in our main sample. The series have been deflated using the national CPI series from Jordà, Schularick, and Taylor (2017). Panel (a) covers the entire sample of cities until 1950, compare 1.1. Panel (b) covers the entire sample of cities after 1950.

apparent in a thicker left-tail in the pre-1950 period. This does not come as a surprise, considering that this period featured two World Wars, the Great Depression and large variations in housing policies. Post-1950 large city returns were close to 2 percentage points higher with a lower standard deviation.

Panel (a) of Figure 1.6 plots 10-year lagged moving averages of log real housing returns averaged over the 27 large cities over time. Housing returns dropped

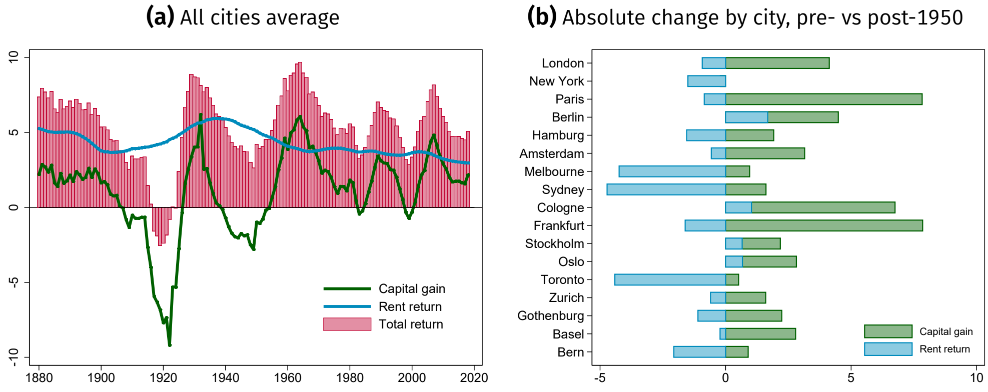

Figure 1.6. City-level average returns (log points)

Notes: Panel (a) shows unweighted averages of city-level log housing returns and its components over time. All cities get an equal weight. The displayed series are 10-year lagged moving averages, e.g. the total returns for the year 2010 are the average of total returns between 2000 and 2010. Panel (b) shows the absolute difference in post- and pre-1950 mean log real capital gains and rent returns by city. A positive (negative) difference means that the mean was higher (lower) in the post-1950 period as compared to the pre-1950 period. We exclude all cities for which we lack pre-1950 data from this graph.

sharply around the World Wars as house prices dropped. The effect is particularly pronounced after World War I, as many governments introduced rent freezes in high inflation environments as discussed by Pooley (1992) and White, Snowden, and Fishback (2014).

The post-1950 period features periods of pronounced cross-country co-movement in housing returns. Most noticeable are the high returns in the postwar boom as well as the low returns in the 1990s in the wake of real estate crises in Japan and Scandinavia, as well as a period of negative house price growth in a number of European cities in the 1990s. Figure 1.6 panel (a) also shows that average capital gains and average rent returns fluctuate at different frequencies. Whereas rent returns vary little and slowly decreased after World War II, average capital gains have been much more volatile.

Higher mean annual total returns in the post-1950 period are driven by substantially higher capital gains. While mean annual capital gains were, on average, 2.95 log points higher in the post-1950 period when compared to the pre-1950 period, mean annual rent returns were, on average, 1.22 log points lower. Panel (b) of Figure 1.6 shows the difference in means between the post- and pre-1950 period by city. Across the majority of cities in our sample, mean capital gains were substantially higher in the post-1950 period, while mean rent returns were lower. The increase in mean capital gains is mostly driven by the largely negative capital gains during the war periods. Kuvshinov and Zimmermann (2020) discuss that the fall in rent

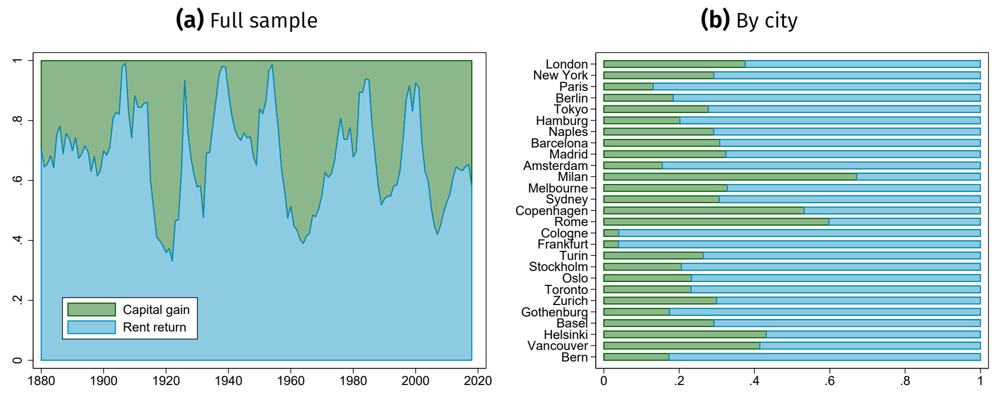

Figure 1.7. Share of log total returns, 1870-2018

Notes: Panel (a): The displayed series are 10-year lagged moving averages, e.g. the share of capital gains for the year 2010 is the average share of capital gains between 2000 and 2010. All cities get an equal weight. This panel shows the share of log capital gains and log rent returns in the sum of both. In the few cases when moving average log capital gains have been negative, we take the absolute value of the moving average log capital gains instead. Panel (b): Average share of log real capital gains and log rent returns by city for the whole period for which we have data for the city.

returns mirrors the secular decline in the yield component of stock returns over the same period.

Rent returns represent approximately 67% of total housing returns over the last 150 years. Panel (a) of Figure 1.7 shows that, although the relative share of rent returns has been quite volatile over time, it has remained by and large the main contributor to total housing returns. In fact, for all cities in our sample, with the exception of Milan, rent returns represent more than 50% of total housing returns in the long-run. This result is in line with the findings in Jordà, Knoll, et al. (2019) and Demers and Eisfeldt (2021).

1.3.1 Large agglomerations vs. national housing markets

In the next step, we merge our city-level data set with national housing returns from Jordà, Knoll, et al. (2019) in order to compare returns in the large to those in the rest of the country. Jordà, Knoll, et al. (2019) compiled data on capital gains, rent returns and total housing returns for nationally diversified housing portfolios that represent the weighted sum of housing markers within a specific country. The weighting is done by value shares, such that more expensive places get a larger weight in the portfolio. We extended their data to 2018 using country-level house price and rent indices from national statistical agencies and substituted house price series for Japan after 2008 and for Sweden after 1952, because series with better methodology and coverage became available. For details see appendix 1.B.1.

The national housing portfolios in Jordà, Knoll, et al. (2019) include the large cities in our sample. For transparency and comparability reasons, we will still compare the large city returns to the national series from that study. But we also calculate returns of a "rest of the country" portfolio as the weighted average of the housing returns in the other locations in the country. National returns can be expressed as:

$$\textit{National return}_t \quad = \quad w_{t-1} * \textit{Large city return}_t \quad + \quad (1 - w_{t-1}) * \textit{RoC return}_t, \quad (1.3)$$

where w is the relative weight of the large city in the respective national housing series. Using equation 1.3 and our large cities return series, we can approximate the housing returns in the rest of the country (*RoC return*) by subtracting the large cities in our data set from the national series. As data on market capitalization are lacking, we use population shares as portfolio weights to construct return series for the rest of the country (excluding the large cities). All city-level and national population data for this calculation are taken from United Nations (2018). House prices tend to be higher in the large cities and using population shares as weights will give a smaller weight to the large cities than a market capitalization weighted index. As a consequence, the rest of the country returns that we back-out from national series in Jordà, Knoll, et al. (2019) likely mark a lower bound.

In some cases, the geographical coverage of the national housing series is too narrow in the pre-World War II era to allow a meaningful comparison between the large cities and the rest of the country. Appendix Table 1.B.1 shows the geographical coverage of the national house price series by country. For the comparison between large city returns and the rest of the country, we will therefore focus on the 70-year period between 1950-2018 for which the national housing series have a wide enough geographical coverage. This being said, the overall results are very similar when we study returns over the entire sample period (see Appendix 1.B.3).

To guide the reader through the results, we start with an example of an undisputed large "superstar" city for which we have high quality data: Paris. Our data show that an investor who bought an apartment in Paris in 1950 realized an average yearly capital gain of 4.85 log points over the period until 2018. The annual rent return in Paris was 3.66 log points on average, resulting in a healthy total annual return of 8.33 log points. This means, for instance, that investments in Parisian residential real estate beat investments in the French equity market by a substantial margin, even on an unleveraged basis.

How does this investment return compare to the rest of France? According to Jordà, Knoll, et al. (2019), an investment in the French national housing portfolio over the same 70-year period saw annual capital appreciation of 4.48 log points, somewhat lower than Paris. As Paris is a substantial part of the French national portfolio, the difference must be driven by other regions in France, in which house prices have risen about half a percentage point less per year than in Paris. However, the picture changes when we bring in rent returns, which were substantially higher in the rest of the country (5.06 vs. 3.66) and more than offset Paris' advantage with respect to capital gains. Total housing returns were 9.15 per annum for the rest of France and thus about 85 basis points per year higher than in Paris. Despite higher capital appreciation, Paris underperformed the rest of France with respect to total returns on housing investment.

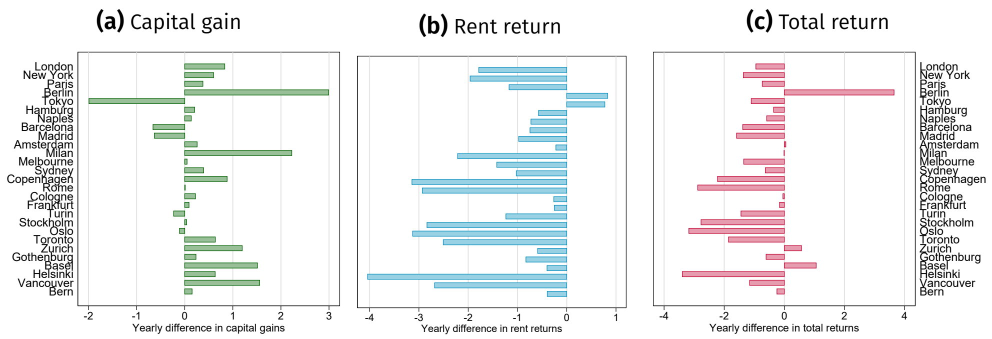

Figure 1.8. Average differences in city-level and national returns (log points), 1950-2018

Notes: This graph shows the mean difference in log capital gains (Panel (a)), log rent returns (Panel (b)) and log total returns (Panel (c)) between the city-level and the respective national portfolio by city. The period covered is 1950 to 2018, except for German cities, Tokyo and Toronto, because the national data only starts in 1963, 1960 and 1957 respectively.

In Figure 1.8, we broaden the perspective to all 27 large cities in the sample and compare them to their national real estate markets. Figure 1.8 shows differences in capital gains (left), rent returns (middle) and total housing returns (right) between 1950 and 2018 for each city relative to the national returns from Jordà, Knoll, et al. (2019).[8] A general pattern can be easily discerned. Just like in the French case, capital gains are higher in nearly all large cities. The only major exception is Tokyo – a city that experienced a severe real estate crisis in the early 1990s. Real house prices in Tokyo were still only one third of their 1990 level in 2018, while house prices in other parts of the country stand at 65% of the 1990 level. Rent returns are generally much lower in the big agglomerations, and overall returns are lower.[9]

8. Appendix Table 1.B.4 presents the numbers including standard errors of paired t-tests.

9. The main exception is (West) Berlin. As data for East Berlin is missing between 1945 and 1990, the Berlin portfolio covers only West Berlin after World War II. The higher housing return in West Berlin might, however, not be surprising when considering the unique history of the city. Prior to the fall of the Berlin Wall and the reunification of Germany in 1990, Berlin was not only heavily supply constrained, but also potentially a very risky place to invest in taking the political tensions between the Soviet Union and the West into account. Additionally, the reunification of Germany itself could be regarded as a very large positive shock to (West) Berlin potentially keeping housing returns off equilibrium for several years. The other outliers are much smaller and typically featured exceptionally high capital gains compared to the respective national index. These, in turn, might be driven by large

Table 1.3. City-level and national yearly housing returns (log points), 1950-2018

	27 large cities				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	2.25	1.82	0.43* (0.23)	1.64	0.61** (0.26)
Rent return	3.55	4.94	-1.39*** (0.04)	5.21	-1.65*** (0.05)
Total return	5.72	6.68	-0.95*** (0.23)	6.76	-1.04*** (0.26)
N	1767				
	Only largest city/country				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	2.45	2.12	0.33 (0.30)	1.99	0.46 (0.34)
Rent return	3.53	5.17	-1.63*** (0.06)	5.41	-1.88*** (0.07)
Total return	5.89	7.18	-1.29*** (0.30)	7.30	-1.41*** (0.34)
N	1061				

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. Rest of country (RoC) returns are calculated as national housing portfolio returns share after taking out the returns of the 27 national large cities. We use previous year population shares as weights of the portfolio share of our cities, such that the estimate should be interpreted a lower bound. The upper panel shows the results averaged over all 27 cities in our main data set. The lower panel shows the results only for the cities, which had the largest population in their respective countries in 1950 in our data. $*: p < 0.1$; $**: p < 0.05$; $***: p < 0.01$.

Table 1.3 formalizes the analysis of different large city/national housing portfolio definitions, together with paired t-tests for the equality of means between city and national housing portfolios: the table shows capital gains, rent returns and total returns at the large city-level (Column 1) and for the national housing portfolios as defined in Jordà, Knoll, et al. (2019) (Column 2). Column (4) shows the population-weighted return for the rest of country (excluding the large cities), as defined above. The lower panel narrows the large city definition to the single largest city in each country (New York, London, Paris, etc.), providing an even stronger large city vs. rest of country comparison.[10]

The results essentially mirror those in the Parisian example before. At 2.25 log points capital gains have been about 43 basis points higher in the 27 large cities than in the national portfolio, and 61 basis points higher than in the rest of the country.

positive shocks to the city development. The main example is Basel, which had a rapidly growing economy since World War II and now is the region with the highest GDP per capita in Switzerland. Within Switzerland, the Canton Basel-Stadt (Nuts-2 region) had by far the largest GDP per capita in 2018, which was nearly twice as high as that of the Canton Zurich, (source: Federal Statistical Office Switzerland, Table je-e-04.02.06.03, published 21.01.2021).

10. We use the largest city per country within our data set. This implies that Toronto is included although Montreal was the largest city in 1950, because housing data for Montreal is missing.

Rent returns, in contrast, have been lower in the large cities with a difference of 1.39 or 1.65 log points, depending on the comparison portfolio. The higher rent returns outside the large cities more than compensate for the lower rate of capital appreciation. Our overall benchmark estimate is that in the long-run total returns in the large cities were 95-100 basis points lower per year than in the national portfolio and rest of the country.

The lower panel of Table 1.3 focuses only on the largest city within each country (measured by 1950 population). If the negative return premium we found is related to the size or the national *importance* of a city, we could expect the effect to become even larger. This is indeed what we observe. For the narrower sample, the average difference between the city-level and the rest of the country grows to 1.41 log points per year. The return difference is not only significant at the 1% level but also economically large. The average total return of the national housing portfolio is around 7% per annum so that large city returns are about 15% lower. Over the long run, housing investments in the most important national cities – like London, New York, Paris or Rome – performed substantially worse than investments in the rest of the country.

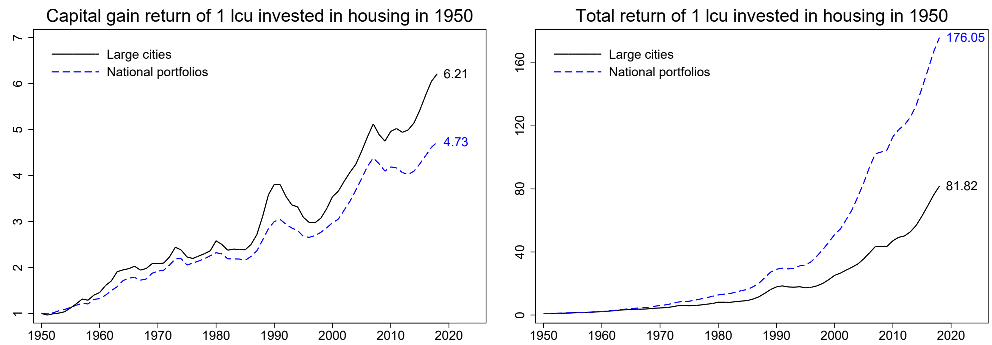

Figure 1.9. Cumulative portfolio returns of city-level vs national portfolios

Notes: The figure shows cumulative real returns for a portfolio with equal investment in each city in our main data set in 1950 (black) compared to a portfolio with investment in each national portfolio weighted by the numbers of cities in our data set (blue). As the national series for Canada, Japan and Germany start only in 1957, 1960 and 1963, respectively, we assume an investment in each of these cities of the average value of the cities in the portfolio in the respective start year and accordingly for the national series. Panel (a) shows the cumulative capital gains of the city and national portfolio. Panel (b) shows the cumulative total returns assuming reinvestment of rental returns.

At first sight the return differences of 95-100 basis points between the large cities and the rest might seem small. But differences in yearly returns can accumulate quickly and generate substantial return differences over time. Figure 1.9 shows that the city-level portfolio increased much faster in value than the national one, because house prices in the large cities appreciated more, namely by a factor of 6.2 compared to 4.7 for the national portfolio.[11] Once again, the picture changes when rent returns are added in the right panel of Figure 1.9. Assuming reinvestment of rent returns, the national portfolio outperforms the city-level portfolio by a large margin. From 1950 to 2018, the cumulative return on a national housing portfolio has been more than twice as high as the returns in the large cities.

1.3.2 Further tests

Capital gains are higher in large cities, but they are more than offset by lower rent returns, resulting in lower overall returns. In the following, we will subject this core finding to a number of additional tests. First, we use alternative rental yield benchmarks. Secondly, we show that our results hold in the historical period as well as in more recent decades. Finally, we study the potential role of rent regulations. Moreover, a discussion of the effect of taxation can be found in appendix 1.D where we show that differences in real estate taxation do not affect our results.

1.3.2.1 Alternative rental yield benchmarks

The data used to calculate rent returns is assembled by professional real estate investors. They are based on rental yield benchmarks net of maintenance, management and other costs. As our core finding rests on the differences in rent returns between large cities and the rest of the economy, we recalculate returns with alternative rental yield benchmarks taken from country-specific sources or from the user driven online database Numbeo.com. The alternative estimates potentially provide a broader coverage of the housing market but might be less precise.

Table 1.4 shows the results with alternative rent return data. If anything, the alternative data accentuate the differences in rent returns and suggests that the differences between market segments within cities do not play a major role. In appendix 1.B.5 we show the summary statistics of our main data set and individual city returns with the alternative rental yield benchmarks. The differences are minor.

1.3.2.2 Subperiods

Driven by limited data availability, most of the recent literature on housing returns focused on developments in the last two or three decades. A natural question to ask is whether our results also hold for the most recent period that saw a particularly

11. In this exercise, we take the perspective of an investor who bought a housing portfolio in 1950 and passively followed the performance of this portfolio over time in real local currency units. We do not reweigh the portfolio, except for adding new cities in cases where the national series starts later than 1950 (Germany, Canada and Japan). This implies that cities that experienced a higher price growth in the beginning of the period get a larger share in the portfolio subsequently. Due to this effect, the numbers do not exactly add up to the averages shown in Table 1.3.

Table 1.4. Yearly housing returns (log points) using alternative rental yield benchmarks, 1950-2018

	27 large cities			Only largest city/country		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.25	1.82	0.43* (0.23)	2.57	2.13	0.45 (0.29)
Rent return	3.32	4.94	-1.62*** (0.04)	3.42	5.20	-1.78*** (0.06)
Total return	5.50	6.68	-1.18*** (0.23)	5.90	7.22	-1.32*** (0.29)
N	1767			1004		

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference using alternative rental yield benchmarks from country specific sources. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. The left-hand side shows the results averaged over all cities in our main data set. The right-hand side shows the results for the cities, which had the largest population in their respective countries in 1950. $*: p < 0.1$; $**: p < 0.05$; $***: p < 0.01$.

pronounced increase in real estate prices (Knoll, Schularick, and Steger, 2017) as well as the emergence of global superstar cities?

For a first test, we split our sample period in 1990. Table 1.5 shows the results for the 27 large cities relative to the national index. Our key results also hold for the most recent period: large city returns have also been significantly lower in the post 1990 era. The same is true for the largest city in each country. Additionally, Appendix Table 1.B.6 presents results for other sub-periods as well as moving averages over the entire time period.

Table 1.5. Yearly housing returns (log points) for 27 large cities until and post 1990

	Until 1990			Post 1990		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.67	2.21	0.46 (0.37)	1.69	1.31	0.38* (0.22)
Rent return	3.73	5.37	-1.63*** (0.07)	3.31	4.36	-1.06*** (0.04)
Total return	6.31	7.47	-1.16*** (0.37)	4.94	5.62	-0.68*** (0.22)
N	1011			756		

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference for the largest city in 1950 within each sample country. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. The left-hand side shows the results for the years from 1950 to 1990. The right-hand side shows the results for the years from 1991 to 2018. $*: p < 0.1$; $**: p < 0.05$; $***: p < 0.01$.

1.3.2.3 Rent regulations

Could stricter rent regulations in large cities account for the lower rent returns compared to the rest of the country? To start with, from an asset pricing perspective, rent regulations should not by themselves have an effect on housing returns since they only regulate the cash-flow received from the asset. As the price of an asset is determined by the discounted value of future expected cash-flows, we would expect house prices to adapt to different cash-flows, such that rent-price ratios will be unaffected. Rent controls could, however, influence expectations about future rents, which could affect house prices and current returns. This could effect our results in the time periods when investors can expect rent control regimes to change.

As an empirical test for the effects of rent controls on returns, we use the rent control index from the *Rental Market Index (ReMaIn) Database*. The database compiled by Kholodilin (2020) uses rent legislation since 1914 in 64 countries to create standardized indices measuring the existence and intensity of rent control, tenant protection and housing rationing. The index ranges from zero to one, with higher values corresponding to stricter rent controls. We divide our sample into cities with weak and strict rent protection and reproduce our main analysis for the rent control regimes.

The results in Table 1.6 confirm that, independently of rent control regimes, capital gains are higher and rent returns lower in the large cities compared to the national average. The absolute difference between large cities and the national returns is even slightly higher in stricter rent control regimes.

Table 1.6. Difference in yearly housing returns (log points) depending on rent regulation, 1950-2018

Sample	Capital gain	Rent return	Total return	N
Weak rent reg.	0.52* (0.30)	-1.64*** (0.08)	-1.11*** (0.30)	497
Strict rent reg.	0.47 (0.44)	-1.74*** (0.08)	-1.26*** (0.44)	687

Notes: The table shows the mean difference between city-level and national log housing returns, log capital gains and log rent returns. Standard errors (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. The first row shows the results for weak national rent regulations defined as a rent law index below one third, the second row the results for strict national rent regulation with a rent law index of at least two thirds. $*: p < 0.1$; $**: p < 0.05$; $***: p < 0.01$.

1.4 Housing returns over the city-size distribution

In this section, we study housing returns across the entire cross-section of cities within two countries, the U.S. and Germany. The choice of these countries is ultimately data driven but allows us to analyze two national real estate markets that belong to two different "housing regimes" (Kohl, 2017): U.S. cities are dominated by owner-occupied, single-family dwellings with light rent regulation but comparatively strong home ownership subsidies. The German housing market is characterized by tenant-occupied, multi-storied buildings and a soft rent-control regime without much home ownership support (Kholodilin and Kohl, 2021). In typologies of housing regimes (Schwartz and Seabrooke, 2008), these two countries often end up on opposite sides and are seen as representative for different approaches in housing policy (Kemeny, 1995).

We use two different data sets that cover the complete size distribution of cities. The approach and the methodology are the same within data sets. The central question is whether the findings from the long-run comparison of large cities with other parts of the country apply more broadly across the city distribution: Are total returns lower in larger cities despite higher capital gains, because rent returns are higher in smaller urban markets?

1.4.1 U.S. superstars redux

For the US we rely on the data set compiled by Gyourko, Mayer, and Sinai (2013), to which we add two additional observations for 2010 and 2018 from the *American Community Survey* (ACS).[12] Their original data cover the near-universe of MSAs from 1950 to 2000 at decadal frequency from the *Census on Housing and Population*. Gyourko, Mayer, and Sinai (2013) find large differences in house price appreciation across metropolitan areas over a period of 50 years.

Due to the decadal frequency of the data, we calculate total housing returns as averages of capital gains and rental yields over 10-year periods. Moreover, we use rental yields instead of rent returns, because the decadal data does not allow us to precisely calculate rent returns. Rental yields are the inverse of the price-rent ratios calculated by Gyourko, Mayer, and Sinai (2013) and adjusted downwards for maintenance costs and depreciation following Jordà, Knoll, et al. (2019). More precisely, we assume that one third of gross rents is spent on these costs across all locations.

We define the largest cities as being the largest five percent of sampled MSAs in terms of 1950 population. Choosing the largest 5% as the cutoff allows us to focus on exceptionally large and economically important cities. The size of these cities will be far from the mass point of cities, as the city size distribution is approximately

12. A disadvantage of adding this new data source is the lower county coverage compared to the census data. To make the data comparable, we build MSA level aggregates using the official borders from 1990, as done by Gyourko, Mayer, and Sinai (2013). All our results stay virtually the same when we restrict our analysis to the original data set covering only the years until 2000 and, if anything, become stronger if we restrict the sample to only MSAs with a full county coverage in 2010 and 2018. Results are available on request. All the data is on MSA-level, but to simplify we still refer to them as "cities" here. For details about data construction please refer to Appendix 1.C.1 and Gyourko, Mayer, and Sinai (2013).

a Pareto distribution.[13] In the following, we compare these top-5% of cities to all other MSAs in the data set and, secondly, to the smallest 5% of MSAs. But our overall results do not depend on these cutoffs.

Table 1.7. Difference in housing returns (log points) for 316 US MSAs, 1950-2018

Sample	Capital gain	Rental yield	Total return	N
Large vs rest	0.13 (0.21)	-0.67*** (0.16)	-0.52*** (0.15)	2184
Large vs small	-0.20 (0.25)	-0.63*** (0.20)	-0.80*** (0.20)	217
GMS superst. vs rest	0.53*** (0.13)	-0.68*** (0.11)	-0.17* (0.10)	1936
GMS superst. vs small	0.44** (0.19)	-0.55*** (0.18)	-0.13 (0.18)	347

Notes: The table shows differences in housing returns between large cities and the rest of the sample or small cities. It covers 316 MSAs on decadal frequency between 1950 and 2010 and additionally the year 2018. Differences are measured as coefficients in a random effects panel regression of the dependent variable (log capital gain, log rental yield and log total housing return respectively) on a large city dummy and year fixed effects. Standard errors (in parenthesis) are clustered at the MSA-level. Large cities are defined as being at or above the 95th percentile of the MSA population distribution in 1950 from census data. The second row shows the same, but comparing large cities only to small cities, which are defined as being at or below the 5th percentile of the MSA population distribution in 1950. The third row compares the superstar cities defined in Gyourko, Mayer, and Sinai (2013) to the other MSAs. In this comparison, we reduced the sample to the 279 MSAs included in the original analysis of the aforementioned authors. Note that we use rental yields instead of rent returns, because using decadal data rent returns cannot accurately be calculated. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$

Table 1.7 presents by now familiar patterns. Rental yields are considerably lower in large cities compared to all other cities or to small cities. Despite somewhat higher capital gains (in most specifications), total housing returns have been significantly lower in large cities within the U.S. since the 1950s. The absolute difference in total returns is estimated between 50 and 80 basis points per year and hence somewhat smaller than in the international sample. This can be expected as we include more large cities compared to above where we only focused on the very largest agglomerations within each country.

The third row shows the comparison of the superstar cities as defined in Gyourko, Mayer, and Sinai (2013) with the rest of the city distribution,[14] but extended to 2018. Using this city sample, the difference in capital gains is significantly positive. This is not surprising, because the authors sample their superstar cities based on exceptionally high house price growth. For these cities too the difference in rental yields is significantly negative and larger in absolute values than the difference in capital gains. Even for a city sample selected by Gyourko, Mayer, and Sinai (2013) on the basis of high capital gains, the total return difference is negative and significant at the 10% level.

13. See e.g. Eeckhout (2004) or Duranton (2007).

14. We use the *ever_superstar* variable of the original data set, extended to the years 1960, 2010 and 2018. The authors exclude MSAs that do not meet the population threshold of 50,000 in 1950.

Thanks to the detailed data, we can also study housing returns over the entire city-size distribution and investigate the relation between city size and returns in more detail. We sort the cities into size deciles ordered from smallest to largest MSA. We also split the first and last decile again to get a more precise picture of the tails of the distribution.

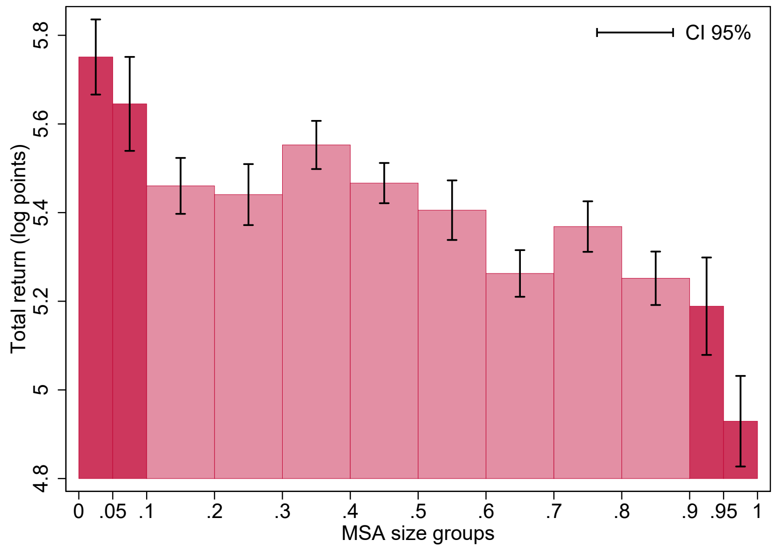

Figure 1.10. Total returns for 316 MSAs (log points) by population size, 1950-2018

Notes: All returns are log returns. Cities are divided into bins based on the size of MSA population in 1950. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins split the smallest and largest deciles in half. As the data for American MSAs only exist in decadal steps, we are not able to construct rent returns. Rental yields are, however, used as a decent approximation of rent returns.

Average log total returns within each bin are plotted in Figure 1.10, which shows that overall housing returns decrease with city size. The relation is not perfectly monotonic across all size bins, but clearly visible overall.[15] Appendix Table 1.C.3 shows the decomposition of housing returns within size bins. It demonstrates that the differences between the largest bins and the others are again driven by considerably lower rental yields in large cities. The relation between capital gains and city size is less clear in the U.S. data.

1.4.2 German cities

For Germany, we constructed a novel data set for this study that covers 42 (West) German cities between 1974 and 2018 at annual frequency. The data set covers only comparably large cities that correspond to urban municipalities excluding rural

15. Results for equity markets are similar. The "big vs small" factor is also not linear across all the size bins and is much stronger for the tails of the distribution; compare Fama and French (1993).

hinterlands.[16] We extend the data to 127 (West) German cities from 1992 onward in a data set that covers the near-universe of (West) German cities. We exclude Eastern Germany, because data coverage mostly started later and Eastern German cities might be fundamentally different to West German ones at the beginning of our sample period. The data set is constructed using market reports of the German Real Estate Association and one of its predecessors.[17] These market reports surveyed local real estate agents and collected city-level observations for various market and quality segments. For the period from 1989 onward, the source allows us to directly use annual estimates for rental yields, such that we only have to rely on the rent-price approach discussed above for some years. We provide more information on the data sources and methods in Appendix 1.C.2. We start with the comparison of large cities and other cities (or the smallest 5% of cities). To do this, we sort cities by their 1975 population.[18] As for the U.S., we define large cities as being at or above the 5% largest of the size distribution.

Table 1.8. Difference in housing returns (log points) for 42 German cities, 1975-2018

Sample	Capital gain	Rent returns	Total return	N
Large vs rest	0.47 (0.57)	-0.91*** (0.34)	-0.45* (0.25)	1848
Large vs small	1.03 (0.72)	-1.58*** (0.43)	-0.57* (0.35)	264

Notes: The table shows differences in annual housing returns between large cities and the rest of the sample or small cities. It covers 42 major German cities between 1975 and 2018. Differences are measured as coefficients in a random effects panel regression of the dependent variable (log capital gain, log rent return and log total housing return respectively) on a large city dummy and year fixed effects. Standard errors (in parenthesis) are clustered at the city-level. Large cities are defined as being at or above the 95th percentile of the city population distribution in 1975. The second row shows the same, but comparing large cities only to small cities, which are defined as being at or below the 5th percentile of the city population distribution in 1975. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$

Table 1.8 confirms the identical pattern for Germany: capital gains tend to be higher in large cities, although the difference is not significant in the German case. Rent returns, by contrast, are considerably lower. Taken together, this leads to lower total returns in the largest cities. As expected the return gap becomes larger when only comparing large to small cities. We also study the more comprehensive housing return data starting in 1992 to compare housing returns over the city size distribution. The results are shown in Appendix Table 1.C.4: once more, rent returns are monotonically decreasing with city size, while capital gains are higher in large cities. Figure 1.11 plots city-level gross rental yields across the German city distribution, as calculated by local real estate agents in 2018. Although gross rental yields vary

16. The average size of cities covered is approximately 418,000 inhabitants in 1975, with a standard deviation around 414,000 and a minimum of approximately 31,000.

17. The *Immobilienverband Deutschland (IVD)* and its predecessor *Ring deutscher Makler (RDM)*.

18. Source: Statistical office of Germany: *Gemeindeverzeichnis, Gebietsstand: 31.12.1975, Statistisches Bundesamt.*

considerably within size bins, a clear negative relation between city size and gross rental yields is visible in the raw data.[19]

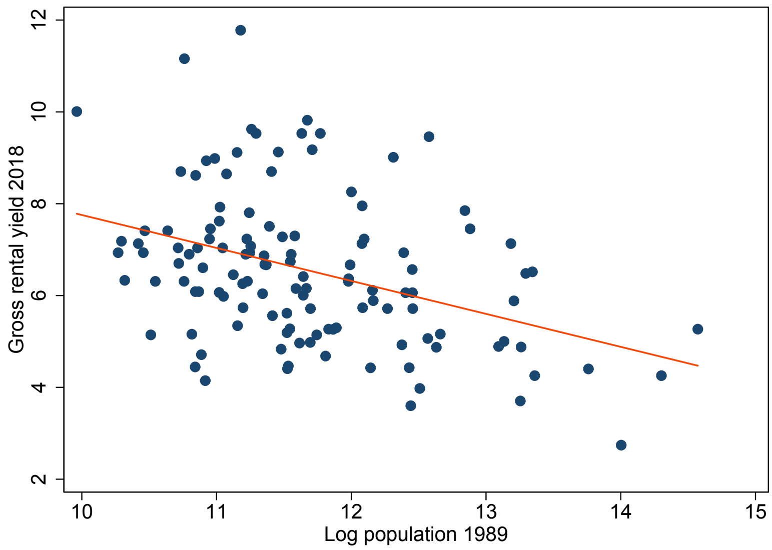

Figure 1.11. Correlation of gross rental yields (log points) in 2018 and log population size

Notes: The figure shows city-level log gross rental yields from IVD by population in 1989 for 127 West German cities. Population data is taken from the "*Gemeindeverzeichnis*" of the German Statistical Office.

Using data for the cross-section of cities in the U.S. and Germany, we have confirmed that the largest cities tend to have lower total housing returns than other housing markets in the same country. A key takeaway from this exercise is that our finding from the long-run national large city data set is confirmed by more comprehensive data for individual countries that point to a close relation between total returns and city size: housing returns are not only the lowest for the largest cities but also tend to be particularly high for very small cities. In the next section we discuss a framework that rationalizes these findings with differences in risk and present supportive empirical evidence.

1.5 Housing risk and return

Both in the long-run historical data and for the city-size distribution in the U.S. and Germany we found that: (i) capital gains are higher in large cities, (ii) rent returns are lower in the big agglomerations, (iii) the difference in rent returns is larger than the difference in capital gains, and, consequently, (iv) total returns are lower in large

19. Hilber and Mense (2021) show that, although the gap in rental yields between London and the rest of England changes over the cycle, rental yields are always smaller in London, even at the trough of the cycle.

cities. In this section we demonstrate that differences in housing risk between large and small cities can account for these findings from the perspective of a rational expectations asset market equilibrium.

Before we go into greater detail, it is important to note that our focus on the rational expectations benchmark is not meant to imply that behavioral factors are not important. Recent research has shown that behavioral factors are important in household decision making and home ownership decisions (for instance, Rozsypal and Schlafmann (2020)). Expectational biases also matter on the macro level and help explain excessive cyclical variation of asset prices and repeated credit booms (Bordalo, Gennaioli, Porta, and Shleifer, 2019). In our setting, it is possible that expectations for house price appreciation are systematically too optimistic in large cities, or that investors myopically focus on higher capital gains in the large cities and neglect the rent return component in total housing returns. In appendix 1.E we use the framework of diagnostic expectations to explore the potential effects of behavioral biases.

In a rational expectation setting, we start with a parsimonious two-city model with housing investments in a large city A and a small city B. We assume that housing risk is lower in the large city A compared to the small city B. In an asset market equilibrium with rational expectations, risk-adjusted total returns need to equalize between cities, such that investors are indifferent between investing in city A or city B:

$$\left(\frac{R^A_{t+1}}{P^A_t} + cg^A_{t+1}\right) * \frac{1}{\delta^A} = \left(\frac{R^B_{t+1}}{P^B_t} + cg^B_{t+1}\right) * \frac{1}{\delta^B}, \tag{1.4}$$

with P^l_t being the house price at time t in location l, R the rent payment, and $cg^l = \frac{P^l_{t+1} - P^l_t}{P^l_t}$ the capital gain. δ^l is the location-specific discount rate. As housing risk is lower in city A, risk-averse investors will discount future payments in A at a lower rate than in B: $\delta_A < \delta_B \iff \frac{1}{\delta_A} > \frac{1}{\delta_B}$. This holds as long as investors have some degree of risk aversion and implies that, in order to attract investors, risky city B will need to offer higher housing returns than safe city A:

$$\frac{R^A_{t+1}}{P^A_t} + cg^A_{t+1} < \frac{R^B_{t+1}}{P^B_t} + cg^B_{t+1}. \tag{1.5}$$

For simplicity of exposition, we assume that houses in both cities feature the same expected future rental cash-flow: $R^A_{t+1} = R^B_{t+1}$. Note that the same result holds under the potentially more realistic assumption that future rents are expected to rise faster in the large city.[20] In order for the equilibrium condition (1.4) to hold, current

20. This is because investors will be willing to pay a higher price for a house with the same *current* rental income, which leads to a lower rent return in city A.

prices will adjust. Investors will be willing to pay a higher price for the safer rental cash-flow, because future payments are discounted at a lower rate. This implies that rent returns would be lower in city A compared to B:

$$\frac{R^A_{t+1}}{P^A_t} < \frac{R^B_{t+1}}{P^B_t}. \tag{1.6}$$

This helps rationalize the empirical finding (ii) that rent returns are lower in large cities. In a next step, we can rewrite inequality (1.5) as:

$$cg^A_{t+1} - cg^B_{t+1} < \frac{R^B_{t+1}}{P^B_t} - \frac{R^A_{t+1}}{P^A_t}, \tag{1.7}$$

which shows that, in equilibrium, the difference in rent returns between city B and city A will be larger than the difference in capital gains between A and B. This, in turn, would rationalize our third stylized fact that the difference in rent returns in favor of small cities exceeds the difference in capital gains between large and small cities.

We know that the right-hand side of inequality (1.7) is larger than zero. This does not, however, pin down the difference in capital gains. It could be the case that risky cities have higher capital gains than safer cities or vice versa. Yet the empirical evidence clearly points to higher capital gains in large cities. To rationalize this finding we need to combine the asset market perspective with insights from the urban economics literature.

It is well established that larger cities have greater supply constraints (Saiz, 2010). National population growth as well as urbanization tendencies increased the demand for housing in cities over the last decades. Highly inelastic housing supply did not meet the surging demand, driving up house prices in the largest cities. A similar mechanism is described in more detail in Gyourko, Mayer, and Sinai (2013). Hilber and Vermeulen (2016) show empirically that more inelastic housing supply causes stronger house price growth in reaction to rising demand. Under the realistic assumption that supply constraints are more binding in large cities, our first empirical finding – higher capital gains in large cities – can hence also be rationalized.

In short, a parsimonious model that features differences in housing risk between large and small cities can account for the key empirical facts established in the previous parts: lower overall returns in large cities despite higher capital gains, driven by lower rent returns. We will now explore the empirical evidence that housing risk is indeed lower in large cities and, if so, what risks investors outside are compensated for with higher returns.

1.5.1 Two sources of housing risk

Examining the difference in housing investment risk between cities, we will consider two separate sources of risk. On the one hand, in some cities the variation in local housing market returns might be more correlated with consumption due to differences in the local economies. On the other hand, idiosyncratic shocks to property-level housing returns might be larger due to differences in the structure of housing markets. Both types of risk are conceptually independent. We will first give some guidance on the two concepts and discuss how we measure both types of risk. Afterwards, we will present empirical evidence that large cities are less risky than smaller ones along both dimensions.

In standard asset pricing, risk premia arise as a result of the co-variance between asset returns and marginal utility, where the latter is typically approximated by consumption growth (Cochrane, 2009). In the case of housing, it could be the case that the co-variance of local housing returns and consumption differs across large and small cities. For instance, one could expect that large cities have more diversified economies, less exposure to industry-specific shocks and a weaker co-variance between housing returns and consumption growth. Additionally, the stronger exposure of large cities to foreign or out-of-town investors could decrease the co-variance in these cities, since these investors are typically less concerned with local risk.[21]

To test this hypothesis, we approximate consumption growth with local income growth and calculate the co-variance between income growth and excess housing returns over the period 1950 to 2018 at the MSA-level in the U.S. We find evidence that the co-variance is lower in large cities, implying that the variation in local market-level housing returns carries more risk in smaller cities.

A second source of risk is idiosyncratic housing risk. In the case of housing, there are good reasons to think that idiosyncratic risk is priced. This is because houses are large, indivisible and illiquid assets and most home-buyers are owner-occupiers that own one house in a specific location and not a diversified housing portfolio (Piazzesi and Schneider, 2016; Giacoletti, 2021). Core assumptions of models of diversified portfolios do not necessarily apply in housing markets.[22]

Higher returns in small cities could be a compensation for higher exposure to idiosyncratic risk. To test whether this is in fact true, we calculate the idiosyncratic component of house price risk following the approach pioneered by Giacoletti (2021) for a large cross-section of U.S. MSAs for the period between 1990 and 2020. The upshot will be that idiosyncratic house price risk decreases with MSA size. Moreover, we will argue that the estimated differences in idiosyncratic risk are linked to liquidity. More liquid housing markets reduce the exposure of homeowners to

21. Recent literature has shown that foreign investors push-up house prices in large international cities like London, New York or Paris, and that their investment decisions are mostly driven by economic factors in their hometowns (Badarinza and Ramadorai, 2018; Cvijanović and Spaenjers, 2021).

22. This lack of diversification suggests that idiosyncratic risk is priced as Merton (1987) showed.

sales-specific shocks, making housing investments in larger cities safer along this dimension too.

1.5.2 Co-variance risk

In a standard asset pricing model, the following holds for a utility-maximizing household that allocates resources between consumption and different investment opportunities:

$$ln\mathbb{E}[R_{t+1}] - lnR_f = \gamma \; Cov\left[ln\left(\frac{C_{t+1}}{C_t}\right), lnR_{t+1} - lnR_f\right], \qquad (1.8)$$

where R_{t+1} is the total return on the asset next period, R_f is the return on the risk-free asset, γ the risk-aversion parameter and $\frac{C_{t+1}}{C_t}$ is consumption growth. In other words, an asset that has a greater co-movement with consumption features a higher risk and, therefore, risk averse agents ($\gamma > 0$) request a higher excess return. Unfortunately, to the best of our knowledge long-run data on consumption at the regional level does not exist. Instead, we approximate consumption growth with regional income growth. An asset is riskier when it has a higher correlation with future income as it cannot be used to hedge income shocks or amplifies them.

To calculate the co-variance between MSA-level income growth and MSA-level excess housing returns, we use the US Census data, described above in section 1.4.1. These data provide a measure of total housing returns and of family income at the MSA-level, which we use to measure the growth of income over time. It is important to note that the data have decadal frequency. This implies that we compare the correlation of log excess housing returns and log income growth over long time periods.[23]

We first calculate MSA-specific co-variances as:

$$Cov_s = Cov(R_s - R_f, y_s),$$

where R_s is total real log housing return for MSA s, R_f is the risk-free rate approximated by total real log returns on short-term U.S. t-bills and y_s is average real log income growth in MSA s. Hence, $R_s - R_f$ is the excess return on housing in MSA s. We calculate the co-variances for the period between 1950 and 2018.[24] We then test whether these co-variances are smaller in large MSAs. The results are depicted in

23. By focusing on the 10-year averages, we are averaging out the cyclical evolution in consumption growth. This is in line with Parker and Julliard (2005), who show that the co-variance between current asset returns and cumulative consumption growth explains the cross-section of expected returns to a much greater extent than the co-variance between the asset's return and contemporaneous consumption growth.

24. Note that given the decadal frequency of the data, we have overall 7 data points for each variable MSA combination.

Table 1.9 row 1. The co-variances of income and excess housing returns are significantly smaller in large MSAs compared to the rest. The difference in co-variances becomes larger when we compare the largest MSAs to only the smallest ones. Appendix 1.H.2 shows results for the entire distribution of MSAs as well as estimated betas from a consumption based asset pricing model (CCAPM).

In the same spirit, the data allow us to test whether high return MSAs exhibit higher co-variances between housing returns and income, as the CCAPM predicts. In the lower half of Table 1.9 we sort MSAs by housing returns and compare co-variances for high and low return MSAs. We find evidence that return co-variances with income are lower in low return cities. This being said, the statistical significance is mixed. The results are borderline significant at the 10%-level ($p = 0.105$) only in the last column where we compare the lowest return MSAs (that tend to be the largest MSAs in terms of population) with all other MSAs. As can be seen in the middle column, the mean difference between co-variances in high vs. low return markets is particularly large, but it is not significant in the decadal data that we have at our disposal. Future research will have to rely on new types of data sets with more granular consumption series and higher frequency return data to pin down these differences more firmly. For now, we conclude that the available data for the U.S. suggest that housing risks are higher in small cities as income co-varies more with local housing returns in these places. Markets also appear to price this risk as there is evidence that MSAs with low returns also tend to have smaller co-variances between returns and income growth than others.

Table 1.9. Differences in co-variances for different MSA sortings, 1950-2018

Sorting	Large vs rest	Large vs small	Rest vs small
By MSA size	-0.55** (0.273)	-1.94*** (0.573)	-1.49*** (0.496)
By total returns	0.36 (0.416)	0.60 (0.448)	0.27 (0.167)
N	316	31	316

Notes: The first row of the table shows differences in the co-variance between income growth and log excess total returns by MSA size. Large MSAs are defined as being at or above the 95th percentile of the MSA population distribution in 1950. Small MSAs are defined as being at or below the 5th percentile of the MSA population distribution in 1950. In column 3 we show the differences between small MSAs and all the rest of MSAs. The second row shows differences in total log housing returns between MSAs with a large total housing returns and the rest of the sample or MSAs with smaller total housing returns. MSAs with large returns are defined as being at or above the 95th percentile of the MSA average total log housing returns distribution between 1950 and 2018. MSAs with small returns are defined as being at or below the 5th percentile of the MSA total log housing returns distribution between 1950 and 2018. In column 3 we show the differences between small MSAs and the rest. Differences are measured as coefficients in a cross-sectional regression of the dependent variable (co-variance) on a large MSA dummy (columns 1 and 2) or on a rest MSA dummy (column 3). Robust standard errors in parenthesis. Overall, we use estimates for 316 MSAs between 1950 and 2018. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

1.5.3 Idiosyncratic house price risk

Using a combination of transaction-level price data from Corelogic and county-level house price indices from FHFA and Zillow.com, we can estimate idiosyncratic risk for American MSAs for the last 30 years. The focus will be on the U.S. because, to the best of our knowledge, equally detailed and micro-level house price transaction data sets do not exist for other countries.

Importantly, these estimates of idiosyncratic risk build on sales data. Including income streams is unlikely to affect property-level return variation as Sagi (2021) has shown for commercial real estate markets. In appendix 1.I, we demonstrate that rental markets are substantially more liquid in larger cities. Rental vacancy rates are lower and less volatile in large cities, decreasing the uncertainty that a landlord face over his future income stream. In all likelihood, estimates using sales data likely mark a lower bound of the idiosyncratic risk differences between large and small markets.

We estimate idiosyncratic house price risk as the unexplained variation in sales-level capital gains after controlling for: (i) market-level price changes (at the county level), and (ii) common house and transaction characteristics in the following equation:[25]

$$\Delta p_{i,l,t} = \Delta v_{l,t} + BX_i + \sigma_{l,idiosyncratic}\ \varepsilon_{i,t}, \tag{1.9}$$

where $\Delta v_{l,t}$ is the growth in local county house prices, BX_i is a vector of house and transaction characteristics, which includes zip-code and time fixed effects, and $\sigma_{l,idiosyncratic}\ \varepsilon_{i,t}$ is a sales-specific shock. We then measure idiosyncratic risk as the standard deviation of sales specific shocks for properties within a specific MSA. Using data from *Corelogic* on single-family repeat-sales for the period between 1990 and 2020, we can estimate annual idiosyncratic risk aggregated at the MSA-level for 248 MSAs, covering around 86% of US population in 1990. We describe the data sources and the methods used to estimate idiosyncratic house price risk in more detail in Appendices 1.F and 1.G.

To calculate the growth in local county house prices ($\Delta v_{l,t}$), we build house price indices from January 1990 to December 2020 combining repeat-sales indices from FHFA, which cover the period between 1990 and 1996, and hedonic price indices from Zillow.com, which cover the period after 1996.

At the county level, the standard deviation appears to slightly increase for the largest MSAs, likely as a result of tighter supply constraints that lead to stronger

25. Giacoletti (2021) studies local market risk at the zip-code level. Our definition of local markets relates to individual counties. The estimates of idiosyncratic risk that we obtain at the MSA level are, however, very similar to the ones we obtain at the zip code level for MSAs for which we have sufficient observations to use both approaches.

house price reactions to positive demand shocks (Hilber and Vermeulen, 2016). However, as shown in the last section, overall house price growth co-varies less with income in the largest MSAs.[26] Idiosyncratic risk represents the largest share of total house price growth variation as can be seen in Appendix 1.H.3.

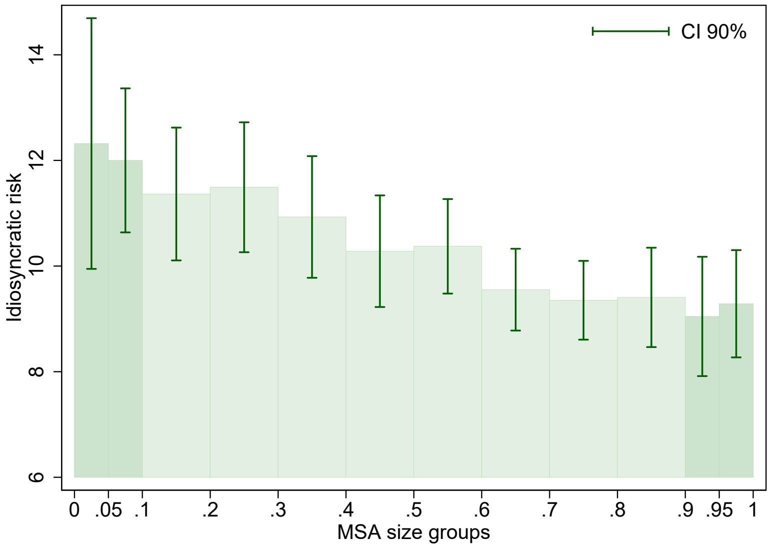

Figure 1.12. Annual idiosyncratic house price risk by MSA size, 1990-2020

Notes: The figure shows average annual idiosyncratic house price risk for different MSA size groups for the period between 1990 and 2020. MSAs are divided into bins based on the size of MSA population in 1990. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins split the smallest and largest deciles in half. All series are real and annualized.

Figure 1.12 plots our measure of idiosyncratic house price risk across different MSA-size bins. It shows that idiosyncratic risk decreases substantially with MSA size. Between 1990 and 2020, average idiosyncratic risk in the smallest MSAs was 12.34% of the house sales price, but about 25% lower in the largest MSAs at 9.28%.

Our measure of idiosyncratic house price risk is orthogonal to local housing market fluctuations. Hence, this risk measure is independent to the co-variance risk that we measured above. However, this does not imply that city-wide factors are irrelevant for idiosyncratic housing risk. Realizations of sales-specific shocks are idiosyncratic by nature. But the distribution from which these sales-specific shocks are drawn is arguably the same for similar houses and will be determined by local housing market characteristics. In other words, $\sigma_{l,idiosyncratic}$ is not

26. Moreover, tighter supply constraints imply that house price increases will be higher in reaction to positive demand shocks. As housing supply cannot be decreased easily in all cities, the effect of negative demand shocks will be much more comparable between constrained and unconstrained cities. Tighter supply constraints, therefore, are comparable to an option value for positive demand shocks without bearing a higher risk if demand shocks are negative.

a fixed parameter, but will be a function of local housing market and other factors: $\sigma_{l,idiosyncratic}(market\ characteristics_l, ...)$. The rest of the section will shed some light on the factors that determine idiosyncratic risk and give some evidence on why it differs so strongly over the MSA-size distribution.

As shown in Equation 1.9, idiosyncratic risk measures the dispersion of individual house price changes around the local house price index, controlling for observed house and transaction characteristics. A standard result from the housing search literature is that more liquid markets have lower price dispersion (Han and Strange, 2015). The intuition is straightforward: in more liquid markets, price information on similar assets will be more readily available, reducing the gap between actual and expected prices. Moreover, higher competition between alternative sellers, or buyers respectively, will reduce price dispersion.

The recent real estate finance literature has established a close relation between idiosyncratic risk and housing liquidity. Empirical work by Giacoletti (2021) and Sagi (2021) shows that the dynamics of idiosyncratic house price risk do not follow a random walk. Instead, they find that risk is realized at the point of sale and resale, indicating that matching frictions in housing markets (i.e., liquidity) are an important source of idiosyncratic risk. For instance, Sagi (2021) develops a structural model that matches the dynamics of commercial real estate risk and market matching frictions linking housing risk and liquidity. A close link between liquidity and idiosyncratic risk has also been shown for other asset classes, e.g. private equity.[27]

We highlight evidence from two liquidity measures across MSAs in the U.S.: time on the market (TOM) and asking price discount. TOM measures the number of days between the original sale listing of a house and its actual sale. The asking price discount measures the difference between the original asking price and the final transaction price. Intuitively, in more liquid markets sellers will have to wait less time to sell (low TOM) and will be able to sell their properties for a price closer to the original asking price (low discount).

We use data from the online real estate marketplace *zillow.com* on median *time on zillow* and median *price cut* for 277 American MSAs for the last decade. Tables 1.10 and 1.11 compare both measures of liquidity in the 5% largest MSAs with the other 95% and the smallest 5% MSAs. In the largest MSAs, sellers take significantly less time to sell on average. Table 1.10 states that the difference between the largest and the smallest MSAs is around 30 days, compared to an overall mean of 100 days. Not only is mean TOM significantly lower in large cities, but it also fluctuates

27. For other illiquid assets, like private equity, Robinson and Sensoy (2016) show that most of the variation in cash-flows is idiosyncratic and Sorensen, Wang, and Yang (2014) demonstrate that idiosyncratic risk (non-systematic risk) faced by private equity investors arises due to its illiquidity. Furthermore, Mueller (2010) and Ewens, Jones, and Rhodes-Kropf (2013) provide empirical evidence that private equity funds with higher idiosyncratic risk also have higher expected returns.

significantly less over time. Results for the full MSA distribution can be found in Appendix 1.J.1.

Table 1.10. Differences in mean and standard deviation of TOM in days, US, 2012-2020

Sample	Mean	S.d. across time	N
Large vs rest	-10.90*(6.184)	-4.34***(0.904)	26869
Large vs small	-29.67***(9.918)	-9.89***(1.782)	2716

Notes: Data on the median number of days on zillow from Zillow.com for 277 MSAs for the period between 2012 and 2020.

Additionally, sellers have to decrease their initial asking price less in large cities compared to smaller cities, as can be seen in Table 1.11. We estimate that on average sellers have to decrease the asking price by 1.5 p.p. less in the largest cities compared to the smallest cities to sell their properties. The asking price discount also fluctuates significantly less over time in the largest cities.

Table 1.11. Differences in mean and standard deviation of asking price discount in p.p., US, 2012-2020

Sample	Mean	S.d. across time	N
Large vs rest	-0.87***(0.096)	-0.36***(0.016)	62688
Large vs small	-1.50***(0.184)	-0.75***(0.052)	6336

Notes: Data on the average discount to the asking price from Zillow.com for 277 MSAs for the period between 2012 and 2020.

The connection between idiosyncratic risk and housing market liquidity also implies that city-wide shocks – such as the often-cited decline of the car industry in Detroit – influence the distribution of sales-specific shocks. Van Dijk (2019) shows that housing liquidity dries up in declining housing markets. For instance, we also see that idiosyncratic risk in Detroit is far above other MSAs of similar size.[28] Moving beyond U.S. data, in Appendix 1.J.2 we also analyze differences in liquidity of the German housing market using data from an online real estate marketplace. For Germany too, the results show that, on a per capita basis, there are more potential sales in larger cities and more potential buyers per sale.

Summing up, housing markets appear considerably more liquid in large cities, both in the U.S. as well as in Germany. Differences in housing market liquidity across the city-size distribution appear to be a second central source of housing risk.

28. The MSA *Detroit-Warren-Livonia* has an average annualized standard deviation of 13.30 percentage points, by far the largest in the largest size bin, which has an average standard deviation of only 8.35 percentage points and also far above Boston-Cambridge-Quincy (7.40) and Washington-Arlington-Alexandria (6.08), which had a comparable MSA size.

1.6 Conclusion

This paper presented a novel data set to study the spatial distribution of return premia in residential real estate markets in the long run. The analysis covered long-run house prices, rent and housing return series for 27 large cities and their respective economies. Our key finding is that housing risk is larger outside the big and diversified cities. The latter tend to under-perform the rest of the country in terms of total returns. The data show that housing risk decreases with city size, driven by the co-variance between local housing returns and local income growth as well as by idiosyncratic risk, with the latter being closely associated with housing market liquidity. Large cities have more liquid housing markets and returns are less correlated with income risk. Both factors make superstar real estate a safer investment and investors willing to accept lower returns.

Our study marks a first step towards a better understanding of spatial risk and return patterns in housing markets over the long run. We anticipate that the new data will allow researchers to gain more insights into the spatial heterogeneity of housing risk and the effects of agglomeration on asset returns.

Appendix 1.A Additional data analyses

1.A.1 Market segmentation

For Cologne, we construct hedonic sub-indices using detailed micro-data between 1989 and 2019. All indices show similar trends. Average yearly house price appreciation differs by 0.217 log points between the complete value-weighted series and the series for single-family houses only. As our rent series might be biased towards apartments in the city center, it is reassuring to see that the value weighted series and the series only for apartments differ only by an average house price appreciation of 0.056 log points between 1989 and 2019.

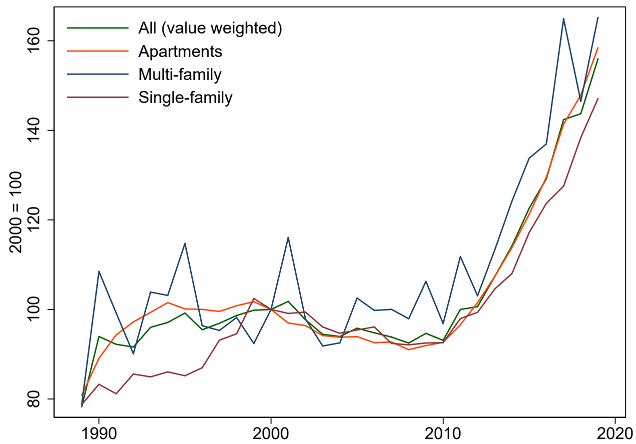

Figure 1.A.1. Cologne house price indices for different market segments

We next use these house price indices for different market segments to calculate housing returns for Cologne.

Table 1.A.1. Yearly housing returns (log points) for Cologne using different hp series, 1990 -2018

Market segment	Capital gain	Rent return	Total return
All (value weighted)	2.30	4.13	6.16
Apartments	2.24	4.24	6.25
Multi-family	2.47	4.02	6.12
Single-family	2.08	4.16	6.04

1.A.2 Log returns

Throughout the paper and in contrast to Jordà, Knoll, et al. (2019), we measure housing returns and their components in log points instead of simple (percentage) returns. This means we measure housing total returns as:

$$\text{Total return}_t = ln\left(\frac{P_t + R_t}{P_{t-1}}\right), \qquad (1.A.1)$$

and their components accordingly. This is a commonly used procedure in finance literature and frequently preferred to simple (percentage) returns for a variety of reasons.[29] In the following, we will discuss why we decided to use log returns throughout this paper, although this might complicate comparison to some other studies of housing returns.

Although simple returns and log returns are approximately equal for small numbers,[30] they have significantly different features. The most important one for our application is that simple returns aggregate linearly across securities, whereas log returns aggregate linearly across time (Meucci, 2010). Throughout this study, we mainly aggregate returns for various housing portfolios over time and compare these aggregates across space. Therefore, time additivity of returns is the more relevant feature in our application.

This feature is especially important when comparing average returns between city-level and national housing portfolios. Time additivity in this case implies that differences in the variance of returns across time do not bias our comparison. To see the contrast to simple returns, consider the following example: In city A, house prices increased by 50% in period 1 and fell by 1/3 in period 2, in city B house prices stayed constant. Using simple returns, average capital gains in city A are approximately 8.3% per year, but zero in city B. In fact, after two periods, prices in both cities are the same as in the very beginning and an investor holding a house for both periods realized a capital gain of zero. Using log returns, average capital gains for both example cities are indeed zero. As national housing portfolios are more diversified compared to city-level portfolios, their variance will typically be lower. Therefore, using simple returns would bias the comparison of city-level and national portfolios towards finding higher returns for city-level portfolios, although the returns over longer periods might not be favorable, just because we measure returns yearly and average them over time. The same bias might occur when comparing large to smaller cities. To be able to make unbiased comparisons, log returns are crucial in our study.

Apart from time additivity, log returns have other preferable features. First, log returns of securities are assumed to be normally distributed. This is true if security prices follow geometric Brownian motion, which is the stochastic process usually assumed for stock prices and the basis of the Black-Scholes-Merton model (Hull (2019), p. 316). Figure 1.5 suggests that log total housing returns are indeed close to be normally distributed. Even if the assumption of normally distributed log returns

29. See Hudson and Gregoriou (2015).

30. For returns that are smaller than 0.15, log and simple returns are very similar in size (Hudson and Gregoriou, 2015).

is violated, time additivity of log returns together with the central limit theorem ensure that compounded log returns converge to normality. Normal distribution of log returns is an important assumption for the estimation techniques used throughout our paper.

Other arguments for using log returns incorporate numerical stability and reduction of algorithmic complexity.[31] But there are also disadvantages of using log returns instead of simple returns.[32] In our application, using log returns implies that total returns are not equal to the simple sum of capital gains and rent returns. Moreover, and as stated above, log returns do not aggregate linearly across securities. Therefore, in the occasions in which we need to aggregate security returns, for example to calculate rest of country returns, we use simple returns and transform them to log returns only afterwards, but before time aggregation.

31. For a good summary please refer to https://quantivity.wordpress.com/2011/02/21/why-log-returns/.

32. For a more critical view on using log returns, refer for example to Hudson and Gregoriou (2015).

1.A.3 Summary statistics

Table 1.A.2. Summary statistics on city-level housing returns (log points)

	Full sample			Post 1950		
City	Capital gain	Rent return	Total return	Capital gain	Rent return	Total return
London	1.50 (9.65)	2.52 (0.87)	3.99 (9.54)	3.21 (10.61)	2.14 (0.71)	5.27 (10.66)
New York	1.45 (12.25)	3.52 (0.98)	4.93 (12.08)	1.39 (12.36)	3.06 (0.55)	4.41 (12.19)
Paris	0.62 (11.20)	4.12 (0.98)	4.73 (10.95)	4.85 (9.14)	3.66 (1.12)	8.33 (9.17)
Berlin	1.08 (18.53)	4.77 (2.27)	5.78 (12.00)	3.51 (10.44)	5.68 (2.14)	9.00 (10.26)
Tokyo	2.01 (16.51)	5.24 (2.01)	7.17 (15.95)	2.01 (16.51)	5.24 (2.01)	7.17 (15.95)
Hamburg	1.09 (24.73)	4.29 (1.46)	5.32 (10.22)	2.12 (6.47)	3.45 (0.80)	5.52 (6.17)
Naples	1.35 (9.02)	3.28 (1.08)	4.58 (8.99)	1.35 (9.09)	3.32 (1.05)	4.62 (9.06)
Barcelona	1.74 (15.27)	3.91 (1.32)	5.58 (15.04)	1.74 (15.27)	3.91 (1.32)	5.58 (15.04)
Madrid	1.76 (16.86)	3.68 (1.06)	5.37 (16.63)	1.76 (16.86)	3.68 (1.06)	5.37 (16.63)
Amsterdam	1.10 (7.73)	5.96 (1.41)	7.02 (7.36)	2.80 (9.46)	5.65 (1.77)	8.32 (9.19)
Milan	3.77 (13.59)	1.85 (0.81)	5.53 (13.62)	3.44 (13.41)	1.83 (0.81)	5.19 (13.43)
Melbourne	2.11 (10.67)	4.33 (2.34)	6.39 (10.45)	2.52 (7.93)	2.54 (0.98)	5.00 (7.85)
Sydney	2.18 (9.91)	4.93 (2.52)	7.04 (9.69)	2.87 (8.22)	2.93 (1.03)	5.72 (8.15)
Copenhagen	2.59 (8.99)	2.28 (1.11)	4.82 (8.92)	2.86 (8.80)	1.92 (0.67)	4.72 (8.87)
Rome	1.64 (8.70)	1.10 (0.38)	2.73 (8.63)	1.22 (8.03)	1.11 (0.38)	2.32 (8.00)
Cologne	0.14 (32.82)	3.43 (1.13)	3.56 (15.32)	2.93 (10.66)	3.86 (0.76)	6.68 (10.50)
Frankfurt	0.21 (23.04)	5.16 (2.92)	5.38 (16.70)	3.65 (13.88)	4.46 (1.97)	7.93 (13.85)
Turin	1.00 (7.08)	2.78 (1.15)	3.74 (7.13)	0.98 (7.13)	2.81 (1.12)	3.76 (7.18)
Stockholm	0.93 (8.67)	3.60 (1.03)	4.50 (8.48)	1.93 (8.48)	3.90 (1.08)	5.76 (8.25)
Oslo	0.90 (13.35)	2.97 (0.74)	3.84 (13.18)	2.21 (10.14)	3.28 (0.81)	5.42 (9.98)
Toronto	1.67 (8.69)	5.53 (2.31)	7.10 (8.84)	1.82 (8.06)	4.18 (0.69)	5.92 (8.08)
Zurich	1.71 (12.17)	4.01 (1.32)	5.65 (12.10)	2.35 (12.22)	3.77 (0.77)	6.05 (11.93)
Gothenburg	1.33 (9.67)	6.29 (1.62)	7.55 (9.47)	2.12 (9.37)	5.91 (1.58)	7.93 (8.98)
Basel	1.67 (11.30)	4.04 (0.57)	5.65 (11.09)	2.67 (10.60)	3.96 (0.57)	6.53 (10.37)
Helsinki	3.26 (10.64)	4.17 (3.02)	7.29 (10.97)	3.59 (10.58)	3.62 (2.03)	7.04 (11.04)
Vancouver	2.80 (11.37)	3.95 (0.81)	6.62 (11.38)	2.80 (11.37)	3.95 (0.81)	6.62 (11.38)
Bern	0.98 (13.63)	4.70 (1.19)	5.65 (13.33)	1.31 (13.80)	3.97 (0.57)	5.23 (13.54)
Global mean	1.45 (14.85)	4.07 (1.97)	5.47 (11.61)	2.44 (11.00)	3.62 (1.64)	5.98 (10.94)

Notes: The table shows arithmetic means of log returns for every city in our sample. Standard deviations are in parentheses. Returns are split up into capital gains and rent returns, log returns are calculated for each category separately. The full sample time period is city specific and refers to the minimum coverage of price and rent data by city depicted in Table 1.1. The post-1950 period covers the same time period per city including return data from 1951 to 2018, except for some German cities, for which the first years after World War II are missing due to data availability.

Table 1.A.3. Summary statistics on city-level simple housing returns (percentage points)

City	Full sample: Capital gain	Full sample: Rent return	Full sample: Total return	Post 1950: Capital gain	Post 1950: Rent return	Post 1950: Total return
London	2.22 (9.71)	2.54 (0.88)	4.76 (9.80)	3.84 (11.07)	2.16 (0.72)	6.00 (11.32)
New York	2.21 (12.43)	3.59 (1.01)	5.80 (12.64)	2.16 (12.71)	3.11 (0.56)	5.27 (12.85)
Paris	1.24 (10.95)	4.22 (1.02)	5.45 (11.18)	5.41 (9.93)	3.74 (1.17)	9.15 (10.37)
Berlin	1.79 (12.00)	4.91 (2.41)	6.70 (12.70)	4.12 (10.84)	5.87 (2.29)	9.99 (11.32)
Tokyo	3.36 (16.60)	5.40 (2.14)	8.76 (17.04)	3.36 (16.60)	5.40 (2.14)	8.76 (17.04)
Hamburg	1.62 (10.55)	4.39 (1.53)	6.01 (11.06)	2.36 (6.78)	3.51 (0.83)	5.87 (6.66)
Naples	1.78 (9.63)	3.34 (1.12)	5.12 (9.91)	1.78 (9.70)	3.38 (1.09)	5.16 (9.98)
Barcelona	2.90 (15.49)	3.99 (1.37)	6.90 (15.83)	2.90 (15.49)	3.99 (1.37)	6.90 (15.83)
Madrid	3.23 (17.87)	3.75 (1.11)	6.98 (18.32)	3.23 (17.87)	3.75 (1.11)	6.98 (18.32)
Amsterdam	1.40 (7.75)	6.15 (1.49)	7.55 (7.84)	3.28 (9.41)	5.83 (1.87)	9.12 (9.73)
Milan	4.82 (14.97)	1.87 (0.83)	6.68 (15.26)	4.45 (14.76)	1.85 (0.82)	6.30 (15.04)
Melbourne	2.82 (14.55)	4.46 (2.45)	7.28 (14.69)	2.87 (8.13)	2.57 (1.00)	5.44 (8.23)
Sydney	2.74 (11.62)	5.09 (2.65)	7.83 (11.87)	3.25 (8.58)	2.98 (1.06)	6.23 (8.75)
Copenhagen	3.04 (9.15)	2.31 (1.14)	5.35 (9.28)	3.29 (9.02)	1.94 (0.68)	5.23 (9.22)
Rome	2.05 (9.29)	1.11 (0.38)	3.16 (9.31)	1.56 (8.44)	1.12 (0.38)	2.68 (8.49)
Cologne	1.27 (14.62)	3.50 (1.17)	4.77 (15.02)	3.57 (11.52)	3.94 (0.78)	7.51 (11.70)
Frankfurt	1.56 (15.93)	5.34 (3.12)	6.90 (16.52)	4.67 (14.08)	4.58 (2.11)	9.25 (14.64)
Turin	1.26 (7.40)	2.82 (1.18)	4.08 (7.63)	1.25 (7.45)	2.86 (1.16)	4.10 (7.68)
Stockholm	1.31 (8.54)	3.67 (1.08)	4.97 (8.72)	2.30 (8.53)	3.98 (1.13)	6.29 (8.68)
Oslo	1.77 (13.07)	3.01 (0.77)	4.79 (13.27)	2.74 (10.17)	3.34 (0.84)	6.08 (10.34)
Toronto	2.07 (9.22)	5.71 (2.48)	7.78 (9.92)	2.17 (8.52)	4.27 (0.72)	6.44 (8.87)
Zurich	2.47 (12.32)	4.10 (1.37)	6.57 (12.65)	3.12 (12.36)	3.85 (0.80)	6.97 (12.53)
Gothenburg	1.79 (9.28)	6.51 (1.73)	8.30 (9.72)	2.56 (8.92)	6.10 (1.69)	8.66 (9.16)
Basel	2.32 (11.49)	4.13 (0.59)	6.45 (11.69)	3.27 (10.96)	4.04 (0.59)	7.31 (11.08)
Helsinki	3.34 (12.12)	4.45 (3.46)	7.80 (12.60)	4.23 (11.19)	3.71 (2.15)	7.95 (12.26)
Vancouver	3.50 (12.02)	4.03 (0.84)	7.53 (12.45)	3.50 (12.02)	4.03 (0.84)	7.53 (12.45)
Bern	1.91 (13.62)	4.82 (1.25)	6.73 (13.85)	2.24 (13.52)	4.05 (0.60)	6.29 (13.69)
Global mean	2.15 (12.02)	4.18 (2.09)	6.32 (12.39)	3.09 (11.37)	3.70 (1.72)	6.80 (11.74)

Notes: The table shows arithmetic means of simple (percentage point) returns for every city in our sample. Standard deviations are in parentheses. Returns are split up into capital gains and rent returns, simple returns are calculated for each category separately. The full sample time period is city specific and refers to the minimum coverage of price and rent data by city depicted in Table 1.1. The post-1950 period covers the same time period per city including return data from 1951 to 2018, except for some German cities, for which the first years after World War II are missing due to data availability.

Appendix 1.B Additional results for city vs national comparison

1.B.1 National housing data

Table 1.B.1 shows the geographical coverage of the national house price series used by Jordà, Knoll, et al. (2019) and constructed by Knoll, Schularick, and Steger (2017), except for two adaptions (cf. see below). For recent years, the series for most countries have nationwide coverage or cover at least the majority of urban areas. Going further back, however, geographical coverage becomes somewhat narrower and is even reduced to one or two large cities for some countries. Therefore, in our main analysis we only use the national series post-1950.

Table 1.B.1. Coverage of national house price series

Country	Period	Coverage
Australia	1870 - 1899	Melbourne
Australia	1900 - 2002	6 capital cities
Australia	2003 - 2018	8 capital cities
Canada	1921 - 1981	nationwide
Canada	1981 - 2018	27 metropolitan areas
Switzerland	1901 - 1929	Zurich
Switzerland	1930 - 1969	urban areas
Switzerland	1970 - 2018	nationwide
Germany	1870 - 1902	Berlin
Germany	1903 - 1923	Hamburg
Germany	1924 - 1938	10 cities
Germany	1939 - 1970	nationwide (Western Germany)
Germany	1971 - 2012	urban areas (Western Germany)
Germany	2013 - 2018	nationwide
Finland	1905 - 1969	Helsinki
Finland	1970 - 2018	nationwide
France	1870 - 1935	Paris
France	1936 - 2018	nationwide
United Kingdom	1899 - 1929	3 cities
United Kingdom	1930 - 1995	nationwide
United Kingdom	1995 - 2012	nationwide (England and Wales)
United Kingdom	2013 - 2018	nationwide
Italy	1927 - 1941	nationwide
Italy	1942 - 1966	8 cities
Italy	1966 - 1997	provincial capitals
Italy	1998 - 2018	nationwide
Japan	1913 - 1930	Tokyo
Japan	1931 - 1935	Kanto district
Japan	1936 - 2007	urban areas
Japan	2008 - 2018	nationwide
Netherlands	1870 - 1969	Amsterdam
Netherlands	1970 - 2018	nationwide
Norway	1870 - 2012	4 cities
Norway	2013 - 2018	nationwide
Sweden	1875 - 1952	Stockholm and Gothenburg
Sweden	1952 - 2018	nationwide
United States	1890 - 1928	22 cities
United States	1929 - 1940	106 cities
United States	1941 - 1952	5 cities
United States	1953 - 2018	nationwide

National rent series from Jordà, Knoll, et al. (2019) typically have a broad coverage, as they are taken from national CPIs, which are constructed to be representative on a national level. For cases when nationwide coverage was not possible, the authors tried to match geographical coverage of the house price series. For details please refer to Jordà, Knoll, et al. (2019).

We adapted the housing series of Jordà, Knoll, et al. (2019) only in two cases. First, we replaced the house price series for Japan from 2008 onward, because a series with a broader coverage and preferable methodology became available. The national house price series we use is produced by the *Ministry of Land, Infrastructure, Transport and Tourism* of Japan (https://www.mlit.go.jp/en/) using individual transaction-level data on detached houses and condominiums from the Land Registry of Japan. It covers all of Japan and uses the hedonic time-dummy variable approach. For more detail, please refer to the given source.

Second, we adapt the national house price series for Sweden between 1952 and 2018, because the series used in Jordà, Knoll, et al. (2019) had limited geographical coverage. We use three different sources, which are all in turn based on Statistics Sweden and very similar for overlapping periods. For the period after 1970, we rely on the nominal national house price index in the OECD analytical house prices indicators database. Between 1957 and 1970, we use the national series in Edvinsson, Blöndal, and Söderberg (2014) and before, we use the series kindly provided directly by Statistics Sweden. The OECD, in turn, uses the index of "Residential property prices, all owner-occupied houses, per dwelling, NSA" from Statistics Sweden from 1985 onward. Before this, all of our sources use the indices on "owner-occupied one- and two-dwelling buildings", also constructed by Statistics Sweden. All series are constructed using the SPAR-method and cover almost the entire universe of real estate transactions in Sweden. They are based on all transfers of real-estate properties that are registered in the Land Survey of Property Prices (LSPP).

As we replace the national house price indices for both countries, we also recalculate rental yields, rent returns, capital gains and housing returns using the methodology of Jordà, Knoll, et al. (2019).

Additionally, we added a new national housing return series for Canada from 1956 to 2018. House prices are taken from the Canadian Real Estate Association between 1956 and 1981. The series contains annual data on the average value and the number of transactions recorded in the Canadian Multiple Listing System (MLS) for all properties, i.e. it includes both residential and non-residential real estate, therefore has nationwide coverage, and is also used in Knoll, Schularick, and Steger (2017) between 1956 and 1974. Afterwards, we deviate from the aforementioned authors and use a house price series from Statistics Canada between 1981 and 2018. The index is computed from sales prices of new real estate constructed by contractors based on a survey that is conducted in 27 metropolitan areas with the number of builders in the sample representing at least 15 percent of the total building permit value of the respective city and year. The construction firms covered mainly develop single-unit houses. The index is a matched-model index, i.e. a constant-quality index in the sense that the characteristics of the structures and the lots are identical between successive periods. For details, please refer to Statistics Canada. We prefer the index to the one used in Knoll, Schularick, and Steger (2017), because it has wider geographical coverage. For rents, we entirely rely on the rent component of the national CPI constructed by Statistics Canada.

As stated in the paper, we also updated the series from Jordà, Knoll, et al. (2019) to 2018. To update house price series, we solely relied on the nominal national house price indices in the OECD analytical house prices indicators database. To update rental series, we mainly relied on the respective national statistical agencies and used nominal national rent indices mostly constructed as part of the CPI series. Exceptions are Portugal and the U.S, for which we got the same kind of data from the FRED database. Many of these sources are already used in Jordà, Knoll, et al. (2019) for recent years. We calculate real series using CPI indices in the JST-database updated with series from the IMF World Economic Outlook database or national statistical agencies. With these series at hand, we calculate returns forward using the approaches described by the aforementioned authors. For rental yields, we use the rent-price approach to calculate rental yields forward coming from the series of Jordà, Knoll, et al. (2019).

1.B.2 Splitting the sample into Europe and the rest of the world

Table 1.B.2. City-level and national yearly housing returns (log points), 1950-2018

	Europe				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	2.34	1.86	0.48* (0.27)	1.66	0.68** (0.29)
Rent return	3.54	4.90	-1.36*** (0.05)	5.14	-1.60*** (0.05)
Total return	5.79	6.67	-0.87*** (0.27)	6.72	-0.93*** (0.29)
N	1380				
	Rest of the world				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	1.94	1.70	0.23 (0.49)	1.54	0.39 (0.59)
Rent return	3.60	5.09	-1.49*** (0.11)	5.44	-1.84*** (0.12)
Total return	5.47	6.71	-1.24** (0.49)	6.91	-1.43** (0.59)
N	387				

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. $*$: $p < 0.1$; $**$: $p < 0.05$; $***$: $p < 0.01$.

In this section, we perform our main analysis from section 1.3.1, but we split our sample into a European sample and a non-European sample. Since our sample has a disproportionate amount of European cities we do this analysis to show that our results are not being driven solely by the European cities in our sample. In practice, this means that the non-European sample includes the United States, Canada, Australia and Japan. We report the results for both samples on Table 1.B.2. The Table shows that our results are both present in Europe as well as outside Europe.

1.B.3 Long-run comparison between large cities and national housing portfolios

In this section, we repeat our main analysis from section 1.3.1, but extend the series for selected cities and countries backwards. We select all cities, for which we have long-run series and where the national housing series have a wide geographical coverage, even before 1950. The period before 1950 was characterized by large shocks such as wars and the Great Depression as well as fundamentally different housing policies, which were changing more rapidly and drastically compared to the postwar period. Although this describes a fundamentally different setting compared to today, we want to demonstrate that our results are robust even when including this time period.

A severe problem for this analysis is that, for many countries, the geographical coverage of the housing series in Jordà, Knoll, et al. (2019) is limited before World War II. As national statistical agencies were not in existence for most countries, the authors had to rely on housing series from other sources, which often only covered some or even just one large city. As our aim is not to compare our large cities to other (or in fact often the same) large cities, we exclude all countries before 1950 that have a geographical coverage of house price or rent series of only a very small number of large cities. After matching with our city-level data, this leaves us, before 1950, with Germany starting 1925,[33] Norway starting 1891, the United Kingdom starting 1930[34] and the United States starting 1920.[35]

The results adding the large cities within these countries before 1950 are depicted in Table 1.B.3. For the sample of all 27 large cities, the results become, if anything, even stronger than when only including the data post 1950 in section 1.3.1. For the sample of only the largest city per country, the results stay virtually

33. We start in 1925 to exclude the period of German hyperinflation, for which measurement of real house price and rent development is subject to very high uncertainty and data is missing for some cities. Moreover, national data for Germany is missing during and in the aftermath of World War II (1939-1962).

34. We have to exclude World War II (1939-1946) because national data is missing.

35. As can be seen from Table 1.B.1, we needed to exclude a considerable number of countries because of narrow geographical house price coverage. From the remaining countries we exclude Italy, France and Switzerland, because the rent series before World War II only cover Milan, Paris and Zurich, respectively. Additionally, we exclude Australia because rent return series for the national Australian portfolio are subject to significant uncertainty before 1950, as can be seen in the Online Appendix of Jordà, Knoll, et al. (2019), and are moreover implausible compared to the housing series for Sydney and Melbourne from Stapledon (2007, 2012), which we use in our main data set. Housing return series start one year later, such that we are able to calculate capital gains with the wide coverage for all included countries.

Table 1.B.3. City-level and national yearly housing returns (log points), long-run

	27 large cities			Only largest city/country		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.15	1.72	0.43* (0.24)	2.39	1.98	0.41 (0.30)
Rent return	3.61	5.11	-1.50*** (0.04)	3.69	5.39	-1.70*** (0.06)
Total return	5.67	6.75	-1.08*** (0.24)	5.98	7.27	-1.29*** (0.30)
N	1920			1039		

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. All 27 large cities are included after 1950. Before 1950, we add Berlin, Hamburg, Cologne, Frankfurt (all after 1925), Oslo (after 1891), London (after 1930) and New York (after 1920). The left-hand panel shows the results averaged over all 27 large cities in our main data set. The right-hand panel shows the results only for the cities that had the largest population in their respective countries in 1950 in our data. $*: p < 0.1$; $**: p < 0.05$; $***: p < 0.01$.

unchanged. This demonstrates that our results are not dependent on starting in 1950 and excluding the period featuring larger shocks to the housing market. Of course, as we still include the full sample after 1950, the weight on the observations before 1950 is small. However, if we instead include only the cities within countries with data coverage before 1950, the absolute differences in total housing returns stays virtually unchanged, but is less precisely measured.[36]

All in all, our main results do not depend on starting our comparison in 1950. Instead, the results become somewhat stronger when we include the time period before 1950 for countries with wider geographical coverage. As the data quality is, however, in general not as good as for the post-war period and large shocks like wars are a source of strong measurement error, we prefer the specification shown in the main text.

36. The difference in total returns is -0.98** for all large cities in the respective countries and -1.14** for only the largest city per country. As the number of observations is considerably smaller in this specification, the results are, however, less precisely measured. The full results for this comparison are available on request.

1.B.4 Additional results for comparison of city-level and national housing portfolios.

Table 1.B.4. Difference in yearly housing returns (log points) by cities, 1950-2018

City	Capital gain	Rent return	Total return	N
London	0.83 (0.81)	-1.78*** (0.17)	-0.95 (0.83)	68
New York	0.60 (1.45)	-1.96*** (0.08)	-1.36 (1.43)	68
Paris	0.38 (0.79)	-1.17*** (0.11)	-0.74 (0.78)	68
Berlin	2.99** (1.15)	0.83*** (0.23)	3.65*** (1.16)	56
Tokyo	-1.99 (1.96)	0.77*** (0.24)	-1.10 (1.93)	59
Hamburg	0.21 (0.67)	-0.57*** (0.09)	-0.36 (0.67)	56
Naples	0.14 (1.11)	-0.73*** (0.08)	-0.59 (1.10)	68
Barcelona	-0.66 (1.97)	-0.75*** (0.15)	-1.38 (1.93)	68
Madrid	-0.63 (1.93)	-0.97*** (0.19)	-1.59 (1.90)	68
Amsterdam	0.26 (0.98)	-0.22 (0.15)	0.05 (0.95)	68
Milan	2.23 (1.62)	-2.21*** (0.10)	-0.01 (1.61)	68
Melbourne	0.05 (0.77)	-1.42*** (0.08)	-1.35* (0.76)	68
Sydney	0.39 (0.79)	-1.02*** (0.08)	-0.63 (0.77)	68
Copenhagen	0.88** (0.44)	-3.14*** (0.18)	-2.22*** (0.49)	68
Rome	0.01 (1.15)	-2.93*** (0.08)	-2.88** (1.14)	68
Cologne	0.22 (1.43)	-0.26** (0.11)	-0.05 (1.42)	56
Frankfurt	0.09 (1.65)	-0.25* (0.13)	-0.16 (1.63)	56
Turin	-0.23 (1.09)	-1.23*** (0.07)	-1.44 (1.07)	68
Stockholm	0.04 (0.98)	-2.84*** (0.20)	-2.77*** (0.99)	68
Oslo	-0.11 (0.72)	-3.13*** (0.18)	-3.18*** (0.75)	68
Toronto	0.64 (0.75)	-2.51*** (0.34)	-1.86** (0.85)	62
Zurich	1.19 (1.47)	-0.59*** (0.07)	0.57 (1.44)	68
Gothenburg	0.23 (0.14)	-0.83*** (0.13)	-0.61*** (0.18)	68
Basel	1.51 (1.33)	-0.40*** (0.07)	1.06 (1.32)	68
Helsinki	0.63*** (0.24)	-4.04*** (0.29)	-3.39*** (0.34)	68
Vancouver	1.56 (1.20)	-2.68*** (0.36)	-1.15 (1.26)	62
Bern	0.15 (1.71)	-0.40*** (0.09)	-0.25 (1.68)	68

Notes: The table shows the mean difference between city-level and national log housing returns, log capital gains and log rent returns by city. Standard errors (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

1.B.5 Alternative rental yield benchmarks

Table 1.B.5. Summary statistics on returns in log points

	Full sample			Post 1950		
City	Capital gain	Rent return	Total return	Capital gain	Rent return	Total return
London	1.50 (9.65)	2.50 (0.90)	3.97 (9.56)	3.21 (10.61)	2.11 (0.75)	5.23 (10.70)
New York	1.45 (12.25)	3.20 (1.16)	4.62 (12.07)	1.39 (12.36)	2.60 (0.60)	3.96 (12.16)
Paris	0.62 (11.20)	3.91 (1.17)	4.52 (10.91)	4.85 (9.14)	3.20 (1.27)	7.89 (9.27)
Berlin	1.08 (18.53)	4.28 (2.04)	5.29 (11.97)	3.51 (10.44)	5.09 (1.93)	8.44 (10.26)
Tokyo	2.01 (16.51)	4.69 (2.10)	6.62 (16.06)	2.01 (16.51)	4.69 (2.10)	6.62 (16.06)
Hamburg	1.09 (24.73)	4.28 (1.47)	5.31 (10.22)	2.12 (6.47)	3.43 (0.79)	5.49 (6.16)
Naples	1.35 (9.02)	3.28 (1.08)	4.58 (8.99)	1.35 (9.09)	3.32 (1.05)	4.62 (9.06)
Barcelona	1.74 (15.27)	3.58 (1.40)	5.26 (15.05)	1.74 (15.27)	3.58 (1.40)	5.26 (15.05)
Madrid	1.76 (16.86)	3.61 (1.09)	5.30 (16.64)	1.76 (16.86)	3.61 (1.09)	5.30 (16.64)
Amsterdam	1.10 (7.73)	6.10 (1.30)	7.14 (7.41)	2.80 (9.46)	5.95 (1.61)	8.60 (9.23)
Milan	3.77 (13.59)	3.11 (1.36)	6.74 (13.66)	3.44 (13.41)	3.09 (1.35)	6.40 (13.46)
Melbourne	2.11 (10.67)	4.33 (2.34)	6.39 (10.45)	2.52 (7.93)	2.54 (0.98)	5.00 (7.85)
Sydney	2.18 (9.91)	4.93 (2.52)	7.04 (9.69)	2.87 (8.22)	2.93 (1.03)	5.72 (8.15)
Copenhagen	2.59 (8.99)	2.90 (0.97)	5.42 (8.95)	2.86 (8.80)	2.65 (0.75)	5.42 (8.91)
Rome	1.64 (8.70)	2.27 (0.77)	3.88 (8.58)	1.22 (8.03)	2.29 (0.77)	3.49 (7.97)
Cologne	0.14 (32.82)	2.72 (0.90)	2.85 (15.35)	2.93 (10.66)	3.05 (0.60)	5.90 (10.53)
Frankfurt	0.21 (23.04)	4.58 (2.60)	4.80 (16.71)	3.65 (13.88)	3.95 (1.76)	7.45 (13.84)
Turin	1.00 (7.08)	2.78 (1.15)	3.74 (7.13)	0.98 (7.13)	2.81 (1.12)	3.76 (7.18)
Stockholm	0.93 (8.67)	2.69 (0.78)	3.60 (8.52)	1.93 (8.48)	2.91 (0.81)	4.79 (8.29)
Oslo	0.90 (13.35)	2.97 (0.74)	3.84 (13.18)	2.21 (10.14)	3.28 (0.81)	5.42 (9.98)
Toronto	1.82 (8.06)	3.56 (0.59)	5.32 (8.08)	1.82 (8.06)	3.56 (0.59)	5.32 (8.08)
Zurich	1.71 (12.17)	3.93 (1.37)	5.58 (12.07)	2.35 (12.22)	3.65 (0.88)	5.93 (11.89)
Gothenburg	1.33 (9.67)	4.03 (1.05)	5.31 (9.51)	2.12 (9.37)	3.78 (1.02)	5.84 (9.08)
Basel	1.67 (11.30)	3.52 (0.71)	5.13 (11.10)	2.67 (10.60)	3.15 (0.48)	5.73 (10.44)
Helsinki	3.26 (10.64)	4.17 (3.02)	7.29 (10.97)	3.59 (10.58)	3.62 (2.03)	7.04 (11.04)
Vancouver	2.80 (11.37)	3.27 (0.67)	5.96 (11.38)	2.80 (11.37)	3.27 (0.67)	5.96 (11.38)
Bern	0.98 (13.63)	4.18 (1.54)	5.14 (13.37)	1.31 (13.80)	3.15 (0.61)	4.42 (13.57)
Global mean	1.45 (14.89)	3.78 (1.76)	5.17 (11.62)	2.44 (11.00)	3.38 (1.41)	5.74 (10.93)

Notes: The table shows arithmetic means of log returns for every city in our sample. Standard deviations are in parenthesis. Returns are split up into capital gains and rent returns, log returns are calculated for each category separately. The full sample time period is city specific and refers to the minimum coverage of price and rent data by city depicted in Table 1. The post 1950 period covers the same time period per city from 1950-2018. In the data we use right now, some years are still interpolated (esp. Germany). Returns from interpolated series are included here. This table uses alternative benchmarks for current rental yields.

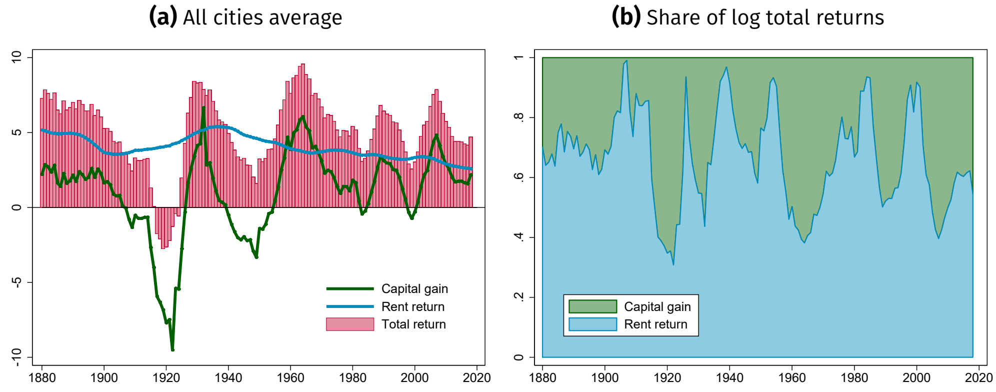

Figure 1.B.1. Log city-level returns and their decomposition, 1870-2018

Notes: The figure shows averages of city-level log housing returns and its components over time. All cities get an equal weight. The displayed series are 10-year lagged moving averages to display the trend component of housing returns. Panel (a) shows the absolute height of log housing return components over time. Panel (b) shows the share of log capital gains and log rent returns in the sum of both. In the few cases when moving average log capital gains have been negative, we take the absolute value of the moving average log capital gains instead. This figure uses alternative benchmarks for current rental yields, MSCI benchmarks are used in the main text Figure 1.6.

1.B.6 Results for different sub-periods

Table 1.B.6. Yearly housing returns (log points) for largest cities per country until and post 1990

	Until 1990			Post 1990		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.97	2.50	0.47 (0.440)	2.03	1.61	0.42 (0.328)
Rent return	3.76	5.75	-1.99*** (0.102)	3.18	4.44	-1.26*** (0.059)
Total return	6.61	8.11	-1.50*** (0.445)	5.15	5.98	-0.84** (0.328)
N	584			420		

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. The left-hand side shows the results for the years from 1950 to 1990. The right-hand side shows the results for the years from 1991 to 2018. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

To demonstrate that our main result is not driven by specific time periods, we depict the difference between city-level and national housing portfolios over time.

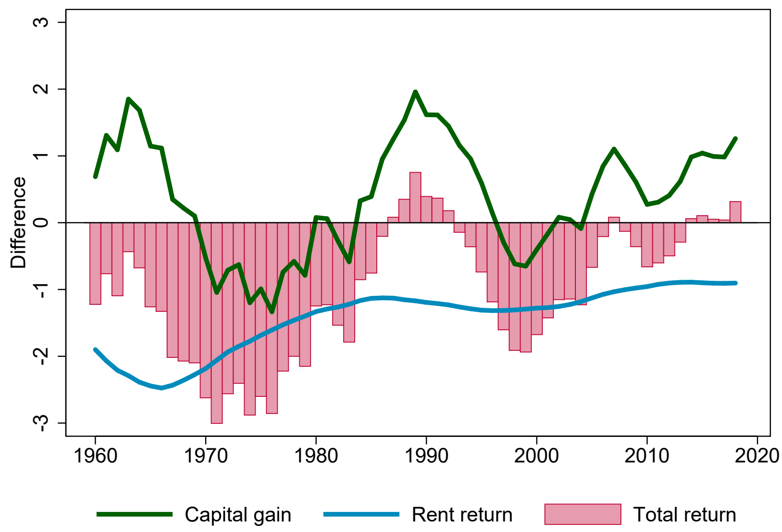

Figure 1.B.2. Average differences in city-level and national returns (log points) over time, 1950-2018

Notes: This graph shows 10 year lagged moving averages of the mean difference in log capital gains, log rent returns and log total returns between the city-level and the respective national housing portfolios. The return period covered is 1951 to 2018, such that the moving averages start in 1960, except for the German cities, Tokyo and Toronto, because the national data starts later for these cities.

As we want to minimize the effect of housing cycles, we compute 10 year lagged moving averages of this average difference.[37]

The outcomes are plotted in Figure 1.B.2. It shows that the main result is prevalent over time. The difference in rent returns is stable and negative over the entire time period. The difference in capital gains, in contrast, is more volatile and it is still possible to spot the influence of housing cycles. In consequence, the difference in total returns is also volatile, but negative during most periods.

Appendix 1.C Within country comparison - Data and further results

1.C.1 US data set

The within country US data set covers 316 MSAs on decadal frequency between 1950 and 2010 and additionally the year 2018. The core of this data set is the data constructed by Gyourko, Mayer, and Sinai (2013) for the decades from 1950

37. Again we rely on the results of Bracke (2013), who shows that the mean duration of complete housing cycles in 19 OECD countries between 1970 and 2010 was around 10 years.

to 2000. It is built using data from the US *Census on Housing and Population*. The authors aggregate the data such that MSA borders are constant over time. For details please refer to the cited paper. In Figure 1.C.1 we show a map with the location of the MSAs in our sample.

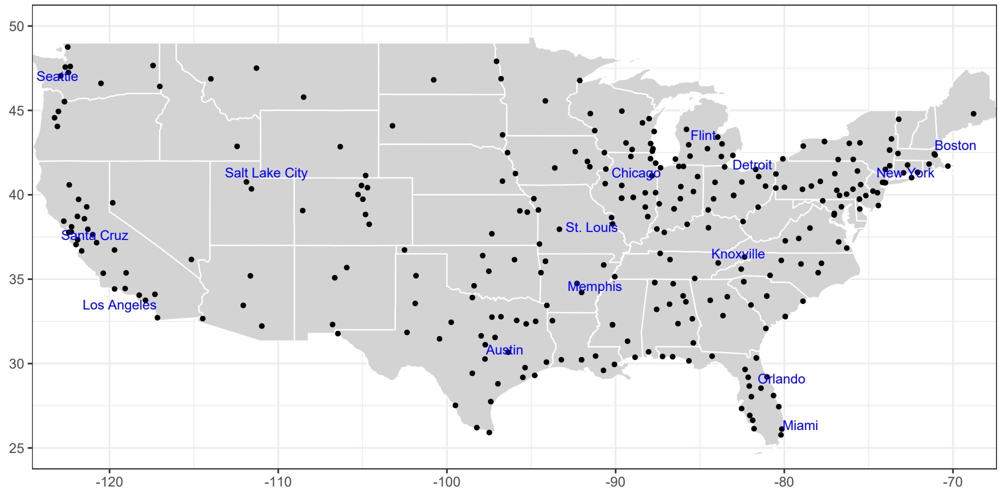

Figure 1.C.1. Geographical distribution of the American MSA sample

Notes: Latitude and longitude are given on the y- and x-axis, respectively. The map was built using the shape file in Bureau (2018).

We extended this data set to also cover the years 2010 and 2018 using data from the *American Community Survey* (ACS).[38] This nationwide survey is the replacement of the long form of the former US census after 2000 and is also conducted by the U.S. Census Bureau. It includes over 3.5 million households every year and asks detailed questions on population and housing characteristics. We use information on aggregated housing value and aggregated rents from the tables B25082, B25075, B25065 and B25063. The main drawback of this source is that only a limited number of geographies is published. We use the one-year estimates, which only include data on counties with more than 65,000 inhabitants.[39]

To construct the data set for the years 2010 and 2018, we use county-level data and merge counties to MSAs following the replication files by Gyourko, Mayer, and Sinai (2013). As the ACS does not cover all counties, for 161 of the total 316 MSAs at least one county is missing or has missing price or rent data in at least one of the years 2010 and 2018. We assume that housing returns in missing counties have been equal to the average of the counties covered within each MSA. As we are still able

38. Unfortunately, the 2010 and 2020 Census did not include questions on housing anymore.

39. The 1-year supplemental estimates do not publish information on aggregated rents and housing values. 5-year estimates cannot be used due to the varying time the data was surveyed, which might induce a considerable bias.

to cover the largest counties within each MSA, the resulting bias is probably small. Most importantly, our main results are robust to restricting our sample to only the 155 MSAs with full data coverage in 2010 and 2018.

To construct housing returns, we approximate capital gains and rental yields based on aggregated housing values and rents. First, we assume constant yearly house price growth within MSAs between the (decadal) data points, such that we compute yearly capital gains from the total capital gain between the respective and the previous data point. Second, gross rental yields are constructed as the inverse of the price-rent ratios calculated by Gyourko, Mayer, and Sinai (2013) and adjusted downwards for maintenance costs and depreciation. Following Jordà, Knoll, et al. (2019), we assume that one third of gross rents is spent on these costs.[40] For the return comparisons, we average rental yields between the respective and the previous data point within each MSA, such that the time coverage of capital gains and rental yields is the same.[41] This way, each data point of both return components can be interpreted as decadal averages within MSAs over the preceding decade. Total housing returns are calculated as the simple sum of these capital gains and rental yields. We are not able to use rent returns because of the decadal frequency of the data. Decadal rental yields are, however, a decent approximation of yearly rent returns, because yearly capital gains are small, such that the difference between rental yields and rent returns is negligible.

Summary statistics of the final housing returns data set can be found in Table 1.C.1.

1.C.2 German data set

We built a data set for German cities using data from a German real estate agents organization. The final data set covers 42 medium-sized and large German cities for the period between 1974 and 2018 (long data set) and as many as 127 West German cities from 1992 until 2018 (wide data set). In Figure 1.C.2 we show a map with the geographical distribution of the cities. The black dots indicate the cities in the long-run data set, while the grey dots indicate the cities in the short-run data set. The data is taken from yearly reports of the largest real estate agents association in

40. This assumption potentially neglects cross-sectional differences in maintenance costs and depreciation as share of gross rents. Any resulting bias will, however, work against us for two reasons: First, for similar properties, rents will be significantly higher in the larger cities, but cross-sectional differences in maintenance costs and depreciation will be low. Second, the share of land value in total housing value will also be higher in large cities, reducing the share of maintenance costs and depreciation in housing value mechanically.

41. This procedure is the same as a linear interpolation of rental yields. The way we approximate rental yields for each data point does not influence our main results. All results look very similar if we use beginning or end of period rental yields. Pairwise correlations of rental yields between MSAs of two subsequent data years are between 0.60 and 0.86 and highly significant.

Table 1.C.1. Summary statistics of US MSA-level log housing returns

	Mean	StdDev	Min	Max
Population 1950	340075.53	748199.80	4286.00	8627356.00
Capital gain 1960	2.20	1.10	-0.22	7.06
Rental yield 1960	4.36	0.58	2.81	7.17
Total return 1960	6.47	1.28	3.17	13.62
Capital gain 1970	0.85	0.87	-1.53	3.24
Rental yield 1970	4.58	0.52	3.21	6.22
Total return 1970	5.39	0.95	2.76	8.26
Capital gain 1980	2.93	1.53	-0.75	7.62
Rental yield 1980	4.19	0.51	2.76	5.69
Total return 1980	7.00	1.41	3.22	10.48
Capital gain 1990	0.37	2.53	-7.03	8.22
Rental yield 1990	3.89	0.64	1.80	5.32
Total return 1990	4.26	2.17	-2.93	11.71
Capital gain 2000	1.65	1.73	-3.45	5.89
Rental yield 2000	3.80	0.68	1.65	5.66
Total return 2000	5.38	1.95	-0.63	9.13
Capital gain 2010	1.92	1.37	-2.92	6.19
Rental yield 2010	3.37	0.65	1.47	5.44
Total return 2010	5.23	1.28	0.72	9.25
Capital gain 2018	0.69	1.68	-3.94	7.43
Rental yield 2018	3.25	0.71	1.11	5.54
Total return 2018	3.92	1.59	-1.37	8.46
Observations	316			

Notes: The table contains summary statistics for the U.S. MSA-level data set. All return variables are measured in log points. The data is constructed using the data from Gyourko, Mayer, and Sinai (2013) (1950-2000) and extended using the ACS (2010, 2018).

Germany.[42] These include data on apartment prices, apartment rents and price-rent ratios for a varying sample of cities. In the long data set, we include all cities that have price and rent data starting in 1974 and including 2018 and have coverage for prices and rents for a minimum of 35 years in-between. In the wide data set, we include all cities that have price and rent data starting in 1992 and including 2018 and have coverage for prices and rents for a minimum of 20 years. Price and rent data for missing years is linearly interpolated.

To construct the yearly reports, the real estate agents association collected data from members located in each specific city relying on their local expertise. Prices and rents are given as mode values within each city. Rents are given for three construc-

42. The *Immobilienverband Deutschland (IVD)* and one of its predecessors, the *Ring deutscher Makler* (RDM).

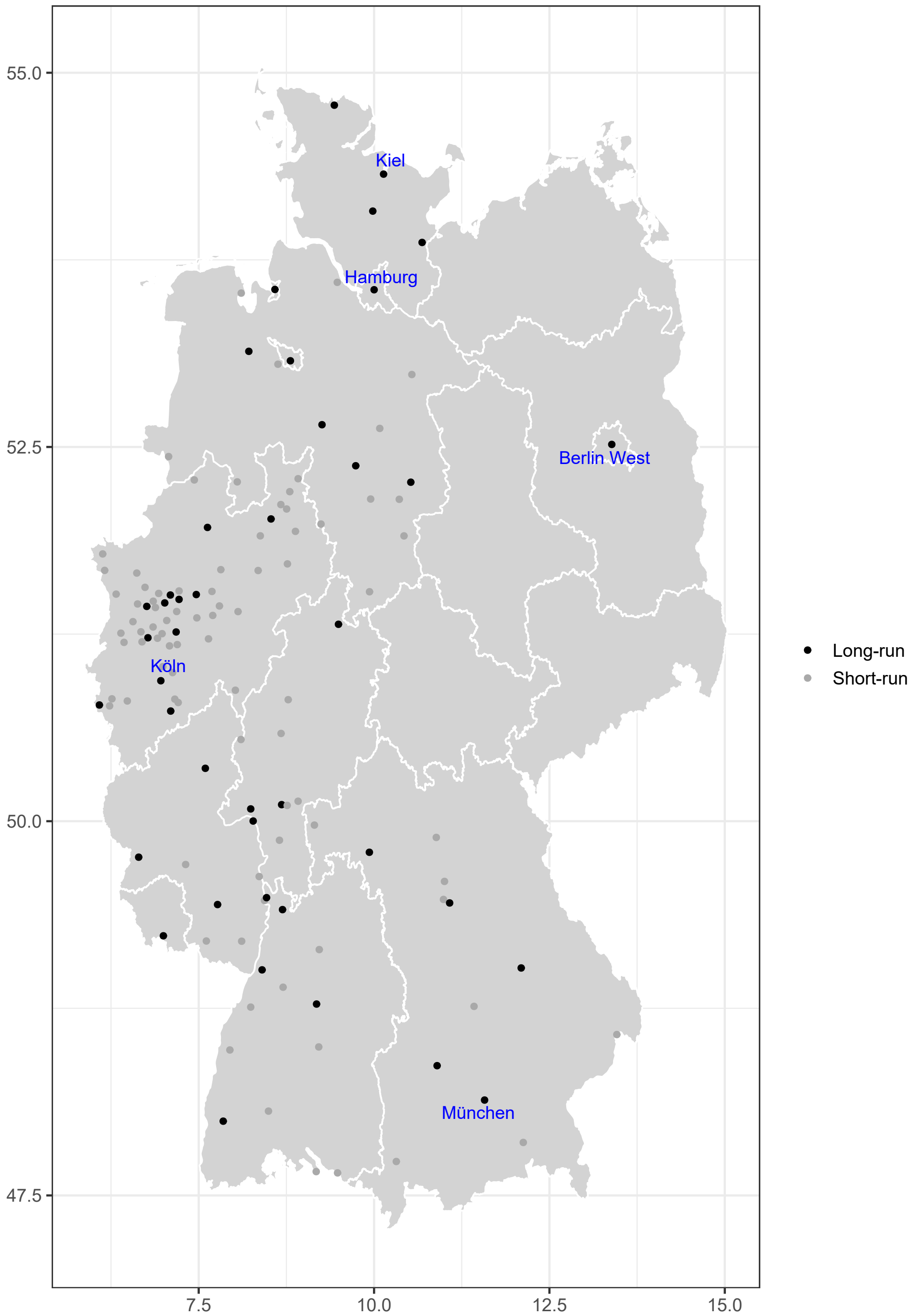

Figure 1.C.2. Geographical distribution of the German city sample

Notes: Latitude and longitude are given on the y- and x-axis, respectively. The map was built using the shape file in Hub (2019).

tion categories, until 1948, after 1948 and for new construction in the respective year, and are, for each category, additionally separated in three different quality bins. Flat prices are separated into four different quality bins and from 2005 onward additionally into new and existing construction. To get a constant quality index, we exclude new construction and build a price and a rent index using a chained matched model approach and simple averages over the non-missing category and quality bins.[43]

Additionally, the data source also provides mode price-rent ratios for residential investment buildings for two construction periods, before and after 1948, from 1989 onward. We calculate gross rental yields as the inverse of these mode price-rent ratios and afterwards take a simple average over the two construction periods. The stated price-rent ratios are already net of running costs and vacancy rates. To calculate net rental yields, following Jordà, Knoll, et al. (2019), we assume that one third of gross rents is used for maintenance and depreciation.[44] For the years prior to 1989 and the missing years in-between,[45] we use the rent-price approach also used to extrapolate rental yields in our main data set. Out of these net rental yield estimates we calculate rent returns using the city-level apartment price indices.

We merge the price and rent indices with CPI data from the JST database until 2013 and IMF for 2014 until 2018 to calculate real price and rent series. We use these real price series to calculate yearly capital gains. We add up these with the rent return estimates to get total housing returns for each city and year. Finally, we take logs of all our return series. We also merge our data to population data for German municipalities (*Gemeinden*) from the statistical office of Germany.[46] We take end of year population for 1975 and 1989, such that we are able to use population at the beginning of our sample period, respectively, and, therefore, our analysis does not suffer from any selection or survivorship bias. In Germany, municipalities cover the complete city, but exclude the hinterlands.[47] Therefore, municipalities are the preferred administrative unit to compare city size. Moreover, the data from the IVD also used municipalities as administrative regions for their city samples.

Summary statistics for both German data sets can be found in Table 1.C.2.

43. We use a simple average, as data on the distribution of the different bins within the housing stock is not available. Using simple averages has the advantage that the weighting of the various bins is the same for every city, such that differences between cities cannot be due to differences within the quality of the housing stock.

44. As already stated above, this assumption neglects cross-sectional differences in these costs, but any resulting bias will work against us.

45. Price-rent ratios are missing for approximately 9.4% of city-year pairs from 1989 onward.

46. Data is taken from the *Gemeindeverzeichnis* from the *Statistisches Bundesamt*.

47. In contrast to counties.

Table 1.C.2. Summary statistics of German city-level log housing returns

	Long data set				Wide data set			
	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max
Population 1975	417029.48	413298.12	30978.00	1984837.00	197296.69	286531.28	21896.00	1984837.00
Population 1998	400584.93	410918.74	30290.00	2130525.00	191571.90	281125.67	21221.00	2130525.00
Capital gain	-0.20	8.69	-59.92	42.70	-0.56	6.61	-42.93	39.21
Rent return	4.79	1.21	1.61	12.91	5.55	1.12	2.04	12.91
Total return	4.60	8.61	-53.77	47.34	5.03	6.44	-37.04	44.34
Observations	1848				3302			

Notes: The table contains summary statistics for both German city-level data sets. The long data set covers housing returns between 1975 and 2018 for 42 cities and the wide data set between 1993 and 2018 for 127 cities. All return variables are measured in log points.

1.C.3 Additional results

Table 1.C.3. Distribution of housing returns (log points) by size of city, US 1950-2018

	1a	1b	2	3	4	5	6	7	8	9	10a	10b
Total return	5.75	5.65	5.46	5.44	5.55	5.47	5.41	5.26	5.37	5.25	5.19	4.93
Rental yield	3.96	3.94	3.93	3.93	4.19	4.13	3.96	3.98	3.99	3.83	3.48	3.32
Capital gain	1.87	1.78	1.59	1.57	1.42	1.39	1.5	1.33	1.43	1.47	1.77	1.66
N	16	16	32	31	32	31	32	32	31	32	16	15

Notes: All returns are log returns. Cities are divided into bins based on the size of MSA population in 1950. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins (1a, 1b, 10a, and 10b) split the smallest and largest deciles in half. As the data for American MSAs only exist in decadal steps, we are not able to construct rent returns. Rental yields are, however, a decent approximation of rent returns.

Table 1.C.4. Distribution of housing returns by size of city, Germany 1993-2018

	1a	1b	2	3	4	5	6	7	8	9	10a	10b
Total return	6.01	5.00	4.89	4.80	4.84	5.15	4.66	4.82	5.39	5.04	5.39	4.88
Rent return	6.57	5.99	5.56	5.86	5.82	5.55	5.34	5.47	5.59	5.13	5.08	4.62
Capital gain	-0.61	-1.06	-0.72	-1.11	-1.04	-0.42	-0.72	-0.69	-0.22	-0.11	0.32	0.25
N	7	6	13	13	12	13	13	12	13	13	6	6

Notes: All returns are log returns. Cities are divided into bins based on the size of city population in 1989. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins (1a, 1b, 10a, and 10b) split the smallest and largest deciles in half.

Appendix 1.D Taxes

Real estate ownership is subject to various taxes: capital gains tax, tax on rent and imputed rent and direct property taxes. These taxes have a direct impact on the returns to housing and it is, therefore, important to take them into account when comparing returns across cities. To make this point clearer, consider the housing return equation, where we specifically account for taxes:

$$\text{Total return}_t = \frac{(P_t - P_{t-1})(1 - \tau_t^{capital})}{P_{t-1}} + \frac{R_t^{gross}(1 - \tau_t^{income} - \tau_t^{property})}{P_{t-1}}, \quad (1.D.1)$$

where $\tau_t^{capital}$ is the tax rate on capital gains, τ_t^{income} is tax rate on rental income, $\tau_t^{property}$ is the property tax rate paid by the owner and R_t^{gross} is the rent net of utility and maintenance costs, but not taxes.

The tax incidence differs geographically and could distort post-tax total returns. If the tax incidence is systematically lower in smaller cities, this - rather than higher pre-tax returns - could explain why we do not find a premium for large cities. For this to be the case, the small-city tax advantage would need to exceed the size of the small city premium.

As mentioned in Section 1.2 we used data on net operating income yields from MSCI to benchmark our rent return series following the same procedure as in Jordà, Knoll, et al. (2019). MSCI defines the net operating income as being net of property taxes. Therefore, our results with the main data set are not driven by differences in property taxes between large and small cities. Nevertheless, we do not take into account capital gains and rental income taxes in the construction of our series for the main data set. Additionally, we also do not explicitly take into account taxes in the construction of the series in the US and in the German data sets.

In this section of the appendix we provide suggestive evidence that this omission in the construction of our series is not driving our main results.

1.D.1 Rental income & capital gains taxes

From Sections 1.3.1 and 1.4 we know that the largest cities have higher capital gains, but lower rental returns than the small cities. Therefore, if rental income is taxed considerably more than capital gains, then, post-taxes, the large city negative premium could disappear. Unfortunately, a precise measurement of the effective tax rates is extremely complicated, since these tax classes are often associated with partial or even full exemptions.[48] Nevertheless, we can still explore the fact that in the post-World War II period a great number of the countries in our sample tried to promote home ownership by reducing the tax burden on homeowners. Through the introduction of mortgage interest deduction and the abolition, or considerable decrease, of capital gains and imputed rents taxes, governments tried to incentivize home ownership. Since, throughout this period, rental income continued, in most cases, to be

48. For example, landlords can deduct a substantial amount of property maintenance costs from the rental income taxes in the US and other countries in our sample. In Germany homeowners are exempted from capital gains taxes if they have owned the property for more than 10 years.

taxed as normal income, this could lead to an effective higher tax burden on rental incomes as compared to capital gains. To test whether this was actually the case we used the series constructed in Kholodilin, Kohl, Korzhenevych, and Pfeiffer (2021) to identify the combinations of countries and periods in which capital gains taxes, mortgage interest deductability or imputed rents taxes were effective. We then divided our sample into different sub-samples depending on the degree to which the tax system was effectively incentivizing home ownership or not. More precisely, we created the following three sub-samples: (i) "not pro homeowner" where only one of the three instruments was in place, (ii) "medium pro homeowner" where two of the instruments were in place and (iii) "strong pro homeowner" where all three instruments were in place. We then compared the return differences between the cities in our sample and the respective countries. The results can be seen in Table 1.D.1.

Table 1.D.1. Difference in yearly housing returns (log points), 1950-2018

Sample	Capital gain	Rent return	Total return	N
Not pro homeowner	0.03 (0.40)	-1.13*** (0.07)	-1.09*** (0.40)	859
Medium pro homeowner	0.90*** (0.31)	-1.66*** (0.06)	-0.76** (0.31)	683
Strong pro homeowner	0.84*** (0.26)	-1.74*** (0.06)	-0.90*** (0.26)	840

Notes: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance stars are calculated using paired t-tests to test equal means of city-level and national return variables. The left-hand side shows the results averaged over all cities in our main data set. The right-hand side shows the results for the cities, which had the largest population in their respective countries in 1950. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

1.D.2 Property taxes in the US data set

In the United States, the American Community Survey (ACS) provides detailed information on aggregate tax income generated by property taxes and the estimated tax values of homes on the county or even Census tract level. Contrary to other countries whose tax assessment values are far from market values, the US property tax is levied on a regularly assessed value of the underlying property and is thus partially a capital gains tax imposed every year. The average effective tax rate expresses the tax expenses as percentage of the average home value which can differ widely even within counties.

Figure 1.D.1 shows for tax data from the pooled 2010-2014 surveys that larger counties and larger MSAs have slightly larger effective tax rates. This suggests that returns in the largest MSAs in our US data set are disproportionately affected by taxes, with the difference in post-tax returns between large and small MSAs becoming even bigger than the difference in pre-tax returns.

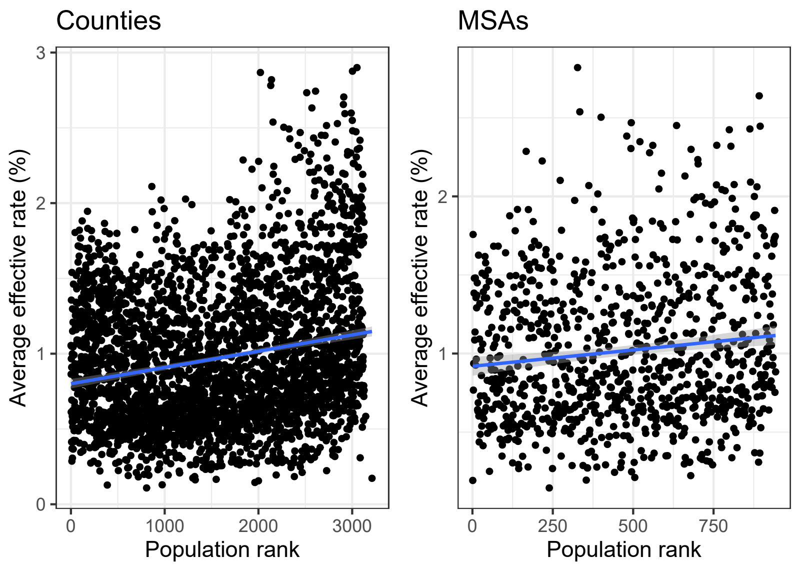

Figure 1.D.1. Effective property tax rates (percent) in counties and MSAs, 2010-2014

Notes: The figure plots the relation between the average effective rate (in percent) for the period between 2010 and 2014 for the universe of U.S. counties (left) and U.S. MSAs (right). The sources of the are described in the text.

Appendix 1.E Housing return expectations

The theory of diagnostic beliefs, as described in Gennaioli and Shleifer (2018), provides a unifying framework, which accounts for the different behavioral biases, i.e. deviations from rational expectations theory, that were documented in the finance and economics literature. It states that people form expectations by extrapolating from past experiences and by overweighting specific representative patterns in the data they observe. Representativeness is defined in the sense of Tversky and Kahneman (1983): "an attribute is representative of a class ... if the relative frequency of this attribute is much higher in that class than in a relevant reference class". In other words, some patterns in the data are more salient than others and, therefore, their importance is overvalued. This theory has found empirical support not only in stock return expectations (Bordalo, Gennaioli, Porta, et al., 2019), but also in macroeconomic expectations, such as for consumption or investment (Bordalo, Gennaioli, Ma, and Shleifer, 2020). In these cases, forecasters are shown to extrapolate from past trends in the data and to overreact to macroeconomic news. There has not been an explicit attempt to study housing markets from the lens of diagnostic beliefs, but most studies investigating behavioral biases in house price or return expectations find evidence for extrapolation. Expectations of future house price growth are strongly correlated with recent house price appreciation (see e.g. Kuchler and Zafar (2019), De Stefani (2020) or Case, Shiller, and Thompson (2014)), and expectations causally affect future housing investment decisions (see Armona, Fuster, and Zafar (2018) or Bailey, Cao, Kuchler, and Stroebel (2018)). Therefore, we will use this framework to organize our discussion on potential biases in housing return expectations.

The housing literature (e.g. Gyourko, Mayer, and Sinai (2013)) and section 1.3.1 have shown that large "superstar" cities have outperformed the rest of their countries in terms of house price appreciation. Moreover, media coverage and the public debate in recent years seem to have focused on the strong house price growth in specific cities, for example concerned about the resulting affordability problems. Recent research by De Stefani (2020) shows peoples' perceptions about the local house price evolution depend on past local price growth. This could potentially explain why homebuyers are more optimistic about the future of the housing markets in large cities than in smaller cities or rural areas and, therefore, willing to pay a higher house price today. In addition, it might be plausible that homebuyers overweight the capital gains component of total returns over the rent return component. We know from section 1.3 that rent returns represent the majority of housing returns, still most news about the housing market focuses exclusively on the evolution of house prices and not on rent returns.[49]

From the perspective of diagnostic beliefs, capital gains are a good candidate for being a representative heuristic of total housing returns, since they are more salient than rent returns. Combining extrapolation of past house price growth and overweighting of the capital gains component has the potential to explain why housing return expectations could be differentially biased between large cities and the rest of the country. If this bias is persistent over time, this could, in turn, explain why house prices in large cities are elevated and, consequently, housing returns are smaller than in other cities as observed in the data.[50]

For illustration, we take the extreme assumption that discount rates are non-stochastic and equal between cities, such that we can drop them from equation 1.4. Next, we assume that expectations are formed using past average capital gains and rent returns, but placing a different weight on the capital gain component, such that we can rewrite the equation as:[51]

49. One reason might be the fact that house price data over time is more readily available than rent data.

50. There is, however, evidence that the effect of expectations on house prices depends on the level of interest rates (Adam, Pfäuti, and Reinelt, 2020) and might, therefore, not be persistent over time. Periods of low interest rates can lead to larger fluctuations in expectations-driven house price dynamics.

51. Here we also make the assumption that extrapolation of past house price growth is constant across cities. There is evidence that sentiment plays a larger role is local housing markets with a higher share of less-informed buyers (Soo, 2018). Nevertheless, there is no clear evidence on the relation between sentiment and expectations.

$$w^P * \overline{cap\ gain}^A + \overline{rent\ return}^A \quad = \quad w^P * \overline{cap\ gain}^B + \overline{rent\ return}^B, \qquad (1.E.1)$$

where w^P is the subjective weight that homebuyers attach to capital gains. We know that capital gains in the large city A have been higher on average than in the small city B, $\overline{cap\ gain}^A > \overline{cap\ gain}^B$. If $w^P > 1$, then the expected returns would increase relatively more in the large city A compared to B. As a result, the expected discounted returns in city A and B could equalize holding discount rates constant across both cities.

Unfortunately, to the best of our knowledge, data on housing return expectations is scarce, let alone on a regional level. Existing surveys mostly focus on house price developments only and are only representative on the national level.[52] Therefore, we are not aware of a direct way to test this hypothesis. However, with a back of the envelope calculation, we are able to approximate the subjective capital gain weight (w^P) that would be necessary for equation 1.E.1 to hold in equilibrium over our long-run data. In the comparison between large cities and national housing portfolios in section 1.3.1, the resulting weight on capital gains would approximately need to be 2.35.[53] This implies that home-buyers would need to attach more than double the weight (or attention) to capital gains than to rent returns, when forming their expectations about future housing returns. Consequently, a substantial behavioral bias would be necessary to explain spatial differences in housing returns without any differences in discount rates.

For homebuyers planning to become owner-occupiers a considerable bias in housing return expectations might, however, be probable. These types of buyers might neither have a reliable estimate of the rent a potential property would be able to earn nor pay much attention to future rent growth. For large-scale (e.g. institutional) real estate investors, in turn, who buy houses or apartments to rent them out, a large behavioral bias seems to be less realistic. Due to their investment strategy, these types of investors can be assumed to take rent returns into account and not overweight capital gains to a large extent. Still, we observe that large real estate investors are concentrated in the largest cities, although housing returns have been lower in these cities on average. Preqin data show that city size is an important predictor for how many real estate deals and residential housing value changed hands in

52. Although there are some more detailed surveys on housing, e.g. the National Housing Survey from Fannie Mae or the Michigan Survey of Consumers, which contain questions on price and rent expectations, these neither allow approximating rent return expectations directly, as price-rent ratios are missing and questions are not very specific, nor do they feature enough observations to reliably approximate expectations on a city-/MSA-level.

53. To calculate the weight on capital gains we first transform the log returns from Table 1.3 into percentage returns, because log returns do not aggregate linearly across return components. By assuming that capital gains weights are constant across cities and countries, we can then simply calculate the necessary weight for the differences to be equal to 0. For our main specification (*Cities vs National*) we calculate a subjective capital gains weight of 2.35.

big deals among institutional investors in Europe in the 2010s (see appendix 1.J.3). At least for these expert homebuyers, a rational explanation seems to be more likely.

Our main results focus on the mean differences in housing returns between large and small cities over a long time period. Deviations from rational expectations in housing markets found in the literature, e.g. extrapolative expectations, have been established over the housing cycle. In that sense, the theory of diagnostic beliefs is more appropriate to explain the cyclical behavior in housing markets. Since we would expect the biases in beliefs to correct over a sufficiently long time period, we propose an alternative rational explanation for the mean differences in returns. In the next subsection, we will test for differences in housing risk between cities, which would lead to locally different discount rates and thereby be able to rationalize differences in expected housing returns between cites.

Appendix 1.F Corelogic deed data set

This section describes in detail the steps that were taken to treat the raw transaction data from the Corelogic deed data set. Our main goal was to remove all data entries corresponding to non-normal sales, i.e. sales which do not correspond to normal market real estate transactions. In the rest of this section we make the concept of market sales clearer, by explaining the steps we took to remove all transactions that did not correspond to this definition. When organizing the data set we took the following steps:

(1) We first exclude all transactions where there was evidence that the contractual parties did not act independently of each other, i.e. where the buyer or the seller was significantly influenced in the process. Typically, these kinds of transactions take place between family members or companies with the same shareholders. Using the *Primary Category Code* from Corelogic we exclude all transaction that are considered to be non-arm's length.

(2) We then exclude all transactions, for which the following is true:
  - The date of the transaction is missing.
  - The transaction amount was wrongly typed, i.e. it contains letters, or it is missing.
  - The transaction amount is smaller than $2000 at the time of purchase
  - The transaction took place before 1990.
  - The zip code or county FIPS code or the house number field is missing.
  - The number of buildings involved in the transaction is larger than one.
  - The transaction is considered a partial sale or a lease by Corelogic.
  - The transaction is based on a quit claim deed.
  - The transaction of a house which has been substantially renovated after 1996.

- The transaction is identified as being part of a multiple sale, i.e. a sale in which different properties are assigned to the same deed.

(3) In a next step, we identify and eliminate duplicates. We first identify complete duplicates, i.e. observations for which all fields are identical, and almost complete duplicates, i.e. observations which have the same internal id, sale date, zip code, house number and transaction amount. Whenever we identify duplicates we leave only one observation per group of duplicates.

(4) We then identify the repeat-sales using Corelogics' unique property identifier alongside the FIPS code, the zip code and the house number.

(5) To make sure we have sufficient observations per MSA, we then drop all MSAs, which have less than 3000 repeat sales in the period between 1990 and 2020.

(6) We also exclude all repeat-sales with holding period shorter than one year, as well as, all sales, which are considered by *Corelogic* to be associated with new construction.

(7) In a final step, we exclude all MSAs, for which the first recorded sale in the data set takes place after 1992. This way we ensure that more recent MSAs are not included in the final data set.

Appendix 1.G Method used to estimate idiosyncratic risk

In this section we describe in more detail the method we used to estimate idiosyncratic risk. Like we mentioned in section 1.I, we mostly follow the method employed by Giacoletti (2021). We measure idiosyncratic risk as the unexplained variation in house price returns after controlling for: (i) market-level fluctuations and (ii) common house and transaction characteristics. Here we explain in more detail all the steps.

Before analyzing the results, it is important to note that our estimation differs from the one in Giacoletti (2021) in two ways. First, we are not able to explicitly take remodeling expenses into account, as the necessary data is missing. However, as shown by Giacoletti (2021), remodeling expenses mainly affect the mean and not the standard deviation of the sales specific shock, which is our variable of interest. Secondly, we do not explicitly control for physical characteristics of housing, since these are absent from the data we use. Nevertheless, our estimates of idiosyncratic risk for the MSAs in California are very similar to the ones in Giacoletti (2021). Therefore, we do not think that these limitations influence our city-level comparisons.

We define the local market at the county level. To measure house prices at the county level, we build new house price indices from January 1990 to December 2020 combining repeat-sales indices from FHFA, which cover the period between 1990 and 1996, and price indices from Zillow.com, which cover the period after 1996. The FHFA indices are built based on single-family transactions covered by mortgages guaranteed by Fannie Mae or Freddie Mac. More details regarding the methodology used to produce the series are described in Bogin, Doerner, and Larson (2018). The Zillow Home Value Index is based on *zestimates* for single-family houses. *Zestimates* are quality-adjusted house price estimates, constructed using proprietary algorithms that incorporate data on sales and listings prices and other home and transaction characteristics from a variety of sources.[54]. We then aggregate the county level indices to the msa-level using repeat sales transaction weights from the Corelogic data set.

Following Giacoletti (2021) we combine the county level series with the corelogic transaction level data to construct the Local Market Equivalents (LME). LMEs measure the extent to which a specific house re-sale deviates from the value fluctuation of the median house in the same county. They are computed as follows:

$$LME_t = \frac{P^{loc}_{i,t_i} - P_{i,t_i}}{P_{i,t_i}} \tag{1.G.1}$$

$$P^{loc}_{i,t_i} = \frac{P_{i,T_i}}{R^{loc}_{t_i,T_i}}, \tag{1.G.2}$$

where P_{i,T_i} is the nominal price at which the house was sold, P_{i,t_i} is the price at which the house was initially bought and R^{loc} is the gross capital gain on the local County price index, i.e. $R^{loc}_{t_i,T_i} = \frac{Index_{county_i,T_i}}{Index_{county_i,t_i}}$. P^{loc}_{i,t_i} is then the market-adjusted buying value of the house.

Although the LMEs do not measure the full extent of idiosyncratic shocks, they already provide a good measure of the extent to which the individual house returns deviate from the market value changes. We computed the standard deviation of the distribution of log LMEs ($lme = log(1 + LME)$) by MSA, and then aggregated the MSAs into population size groups for the period between 1990 and 2020. The results can be seen in 1.G.1. In the first row we present the results for holding period log LMEs and in the second row for annual log LMEs. In both cases we can clearly see a differences across locations. The smallest MSAs have substantially higher LMEs than the largest MSAs.

Table 1.G.1. Log LME across MSAs size bins, 1990-2020

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Holding Period	44.32	40.38	39.44	41.80	36.58	35.65	34.33	32.84	32.78	31.03	31.00	30.50
Annual	17.18	15.96	15.61	16.18	14.48	13.96	13.80	13.17	13.09	12.57	12.54	12.91
N	13.00	12.00	25.00	25.00	24.00	25.00	25.00	24.00	25.00	25.00	12.00	12.00

Notes: The Table shows the standard deviation of the log LME estimates for 248 MSAs by size decile group. The first row shows the estimates for holding period log LME (lme_i) and the second row for annual log LME ($\tilde{lme}_i$), which is explained in the text.

54. More details about the data and methodology can be found in *www.zillow.com*

Overall, we can already see that the individual housing returns fluctuate less in larger MSAs. Nevertheless, the changes in individual house values can also stem from transaction and house characteristics, which are more prevalent in specific MSAs. Therefore, in a second step, we remove the additional return variation determined by common house and transaction characteristics from the individual house resale value fluctuations. For that purpose we run the following regression:

$$\begin{aligned} \tilde{lme}_i =& \alpha_{s,y} + \alpha_{e,y} + \alpha_{s,m} + \alpha_{e,m} && (1.G.3) \\ &+\alpha_{zip} + \beta_P log(P_{i,t_i}) + BX_i + u_i, && (1.G.4) \end{aligned}$$

where $\tilde{lme}_i = \frac{lme_i}{\sqrt{hp_i}}$ and hp_i is the holding period in years. The rescaling by holding periods follows Sagi (2021) and deals with potential collinearity arising from differences in holding periods across resales. $\alpha_{s,y}$ and $\alpha_{e,y}$ are fixed effects for the year in which the house was bought and sold,$\alpha_{s,m}$ and $\alpha_{e,m}$ are fixed effects for the month in which the house was bought and sold and α_{zip} is a zip-code fixed effect. $log(P_{i,ti})$ is the log of the price at which the house was bought, which is also a control for other unobservable persistent characteristics. BX_i is a vector of additional transaction characteristics. The vector X_i contains dummies for different holding periods (less than 2 years, between 2 and 3 years, between 3 and 5 years, between 6 and 8 years, between 8 and 10 years and longer than 10 years), it also contains dummies for sales or resales which fit the following descriptions: short sales, bought solely with cash, foreclosures, and bought or sold by institutional investors or real estate developers. For a full description of the methodology please refer to Giacoletti (2021).

The residuals u_i then capture the unexplained component of returns, which is controlled for systemic price fluctuations and common house and transaction characteristics. We then measure annual idiosyncratic risk as the standard deviation of the residuals within a specific MSA. The standard deviation in measured in terms the original price's %. Since the dependent variable of the regression is scaled by the square root of the holding period we need to rescale the residual as $\hat{e}_i = \hat{u}_i \sqrt{hp_i}$ in order to have the residual associated with the holding period.

We also do a comparison of the standard deviation of the residuals across MSAs. The results can be seen in Table 1.G.2. As predicted, the idiosyncratic risk estimates are now overall lower, as when compared to the lmes, but the pattern is the same: larger MSAs have a lower idiosyncratic risk than smaller MSAs.

Table 1.G.2. Idiosyncratic risk across MSAs, 1990-2020

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Holding Period	31.30	30.46	28.81	29.13	27.80	26.07	25.85	23.85	23.58	23.23	22.69	22.40
Annual	12.32	12.00	11.36	11.49	10.93	10.28	10.37	9.55	9.35	9.41	9.05	9.29
N	13.00	12.00	25.00	25.00	24.00	25.00	25.00	24.00	25.00	25.00	12.00	12.00

Notes: The Table shows the standard deviation of holding period idiosyncratic risk ($\hat{u}_i$) and annual idiosyncratic risk ($\hat{e}_i$) for 248 MSAs by size decile group.

Appendix 1.H Housing risk distribution

1.H.1 Co-variance risk across the distribution

In this subsection of the appendix we show that the co-variance between excess housing returns and income growth decreases almost monotonically across the city-size distribution. In Figure 1.H.1, we plot the average co-variance between excess housing returns and income growth by MSA-size group for the period between 1950 and 2018. We can see that the co-variance is significantly positive for the smallest MSAs, and decreases almost monotonically with MSA-size. For the largest MSAs the estimated co-variance is not significantly different from zero.

1.H.2 MSA-level housing betas

A key question for the empirical analysis of the Consumption CAPM (CCAPM) is how to measure consumption. Reliable local consumption data do not exist for a longer time period over a large cross-section of cities. To circumvent this problem, we proxy consumption growth with income growth.

In the CCAPM setting only the exposure to aggregate income should be compensated with higher returns, because households can hold assets to diversify away idiosyncratic income risk. However, most households are highly dependent on local economic conditions, both because all household members earn their wages in the local labor market and because being a homeowner implies that a large share of their asset portfolio has a local component.[55] We, therefore, use income aggregated at the MSA-level. This is in line with the regional economics literature (e.g. Blanchard and Katz (1992), Moretti (2013)) that treats local labor markets as sub-economies for which we can observe market equilibrium outcomes. It also relaxes the assumption that households fully ensure against idiosyncratic income risk, but it still assumes that households can insure against most idiosyncratic income shocks.

To calculate betas between MSA-level income and MSA-level housing returns, we rely on the US Census data from Gyourko, Mayer, and Sinai (2013), which we

55. In the case of housing, most households own only one house. In the case of equity there is also evidence that households hold under-diversified portfolios (Goetzmann and Kumar, 2008).

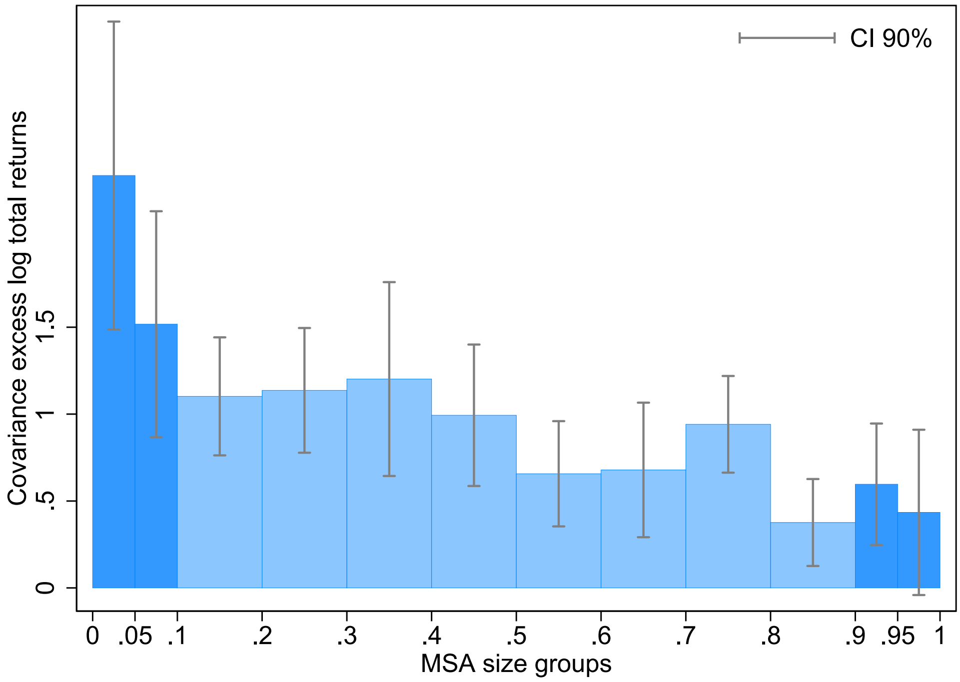

Figure 1.H.1. Co-variance between log excess total housing returns and log income growth by MSA size, 1950-2018

Notes: The figure shows the co-variances for different MSA size groups for the period between 1950 and 2018. MSAs are divided into bins based on the size of MSA population in 1950. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins split the smallest and largest deciles in half.

updated to 2018. This data provides both a measure of total housing returns as well as mean family income at the MSA-level, which we use to measure the growth of income over time.

We calculate MSA-specific co-variances as:

$$\beta_s = \frac{Cov(R_s - R_f, y_s)}{Var(y_s)},$$

where R_s is total real log housing return for MSA s, R_f is total real log return on short-term US t-bills and y_s is average real log income growth in MSA s. We calculate income betas for the period between 1950 and 2018.[56] We then test whether income betas are smaller in large MSAs. The results are depicted in Table 1.H.1 column 3. It shows that income betas of total housing returns are indeed significantly smaller in large MSAs compared to the rest. The difference becomes larger when we compare

56. Note that given the decadal frequency of the data, we have overall 7 data points for each variable MSA combination.

the largest MSAs to only the smallest ones. Appendix 1.H.2 shows results for the entire distribution of MSAs.

Table 1.H.1. Differences in income betas by city size, US, 1950-2018

Sample	Capital gain	Rental Yield	Total return	N
Large vs rest	-0.23*** (0.036)	-0.24*** (0.018)	-0.29*** (0.033)	2212
Large vs small	-0.57*** (0.079)	-0.35*** (0.032)	-0.66*** (0.073)	217

Notes: The table shows differences in income betas for log excess total returns, log excess capital gains and log excess rental yields between large MSAs and the rest of the sample or small MSAs. Differences are measured as coefficients in a cross-sectional regression of the dependent variable (income beta) on a large MSA dummy. Robust standard errors in parenthesis. Large MSAs are defined as being at or above the 95th percentile of the MSA population distribution in 1950. The second row shows the same, but comparing large MSAs only to small MSAs, which are defined as being at or below the 5th percentile of the MSA population distribution in 1950. Overall, we use estimates for 316 MSAs between 1950 and 2018. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

We do the same analysis for the two components of log total returns: log capital gains and log rental yields. We calculate the income betas for each one of the components separately. The results can be found in Table 1.H.1 columns 1 and 2, which also show that betas for both components are smaller in the largest cities.

Figure 1.H.2 plots income betas for total housing returns by MSA-size bins. Section 1.5.2 describes how betas are calculated. It shows that betas are decreasing with MSA-size and that the difference is especially strong for the tails of the distribution, mirroring the picture for housing returns.

Table 1.H.2 shows results from a regression of MSA-level income betas of log MSA population size in 1950. It shows that betas are significantly decreasing with MSA-size.

Table 1.H.2. Regression results of income betas on city size

	Returns	Cap. Gains	Rental Yields
Log Population 1950	-0.154***	-0.131***	-0.101***
	(0.0278)	(0.0290)	(0.0139)
Constant	2.129***	1.674***	1.280***
	(0.345)	(0.360)	(0.174)
Observations	316	316	316
R^2	0.098	0.065	0.143

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows differences in income betas for log excess total returns, log excess capital gains and log excess rental yields by log population size in 1950. Overall, we use estimates for 316 MSAs between 1950 and 2018. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

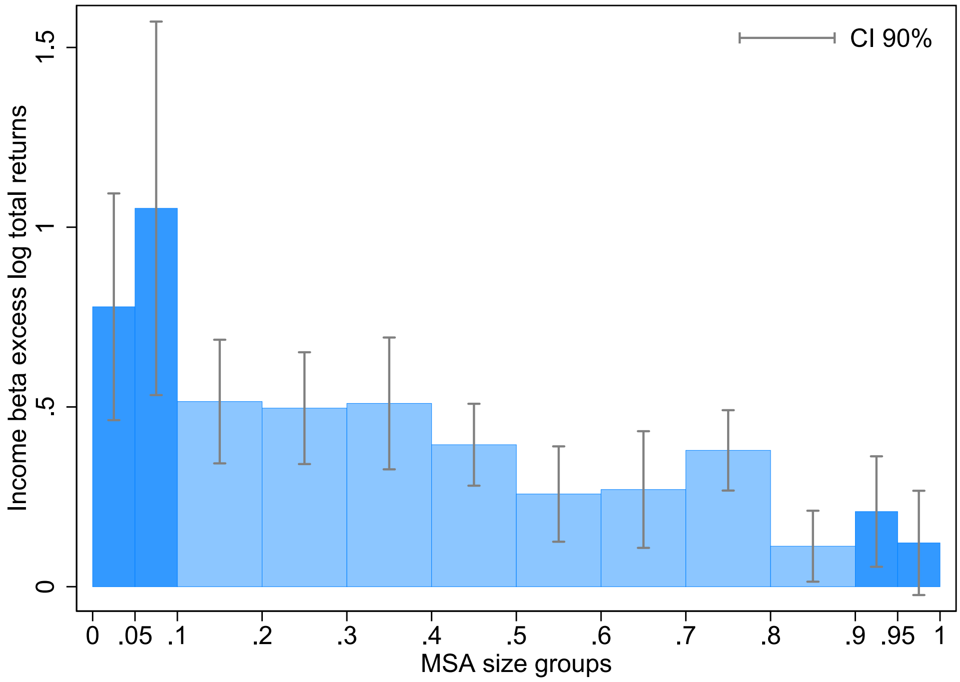

Figure 1.H.2. Income betas on log excess total housing returns by MSA size, 1950-2018

Notes: The figure shows income betas for different MSA size groups for the period between 1950 and 2018. MSAs are divided into bins based on the size of MSA population in 1950. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins split the smallest and largest deciles in half.

1.H.3 Distribution of house price growth variation

Table 1.H.3. Total house price growth variation and its decomposition by MSA size, 1990-2020

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Local risk	4.96	5.02	5.65	5.14	5.19	5.16	5.46	5.79	5.51	5.09	6.58	6.63
Idiosyncratic risk	12.32	12.00	11.36	11.49	10.93	10.28	10.37	9.55	9.35	9.41	9.05	9.29
Share of idios. risk	0.81	0.83	0.76	0.80	0.78	0.76	0.76	0.72	0.73	0.75	0.64	0.65
Total risk	13.61	13.17	13.11	12.80	12.47	11.86	12.18	11.64	11.18	10.96	11.54	11.63
# Repeat sales	139369	126863	266186	359600	406745	616257	732470	1038956	1532555	3003841	3190250	4779689
# MSAs	13	12	25	25	24	25	25	24	25	25	12	12

Notes: All risk measures are yearly and in percentage points of initial prices. MSAs are divided into bins based on the size of MSA population in 1990. The bins go from the smallest MSAs (bin 1A) to the largest MSAs (bin 10B). The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins split the smallest and largest deciles in half.

Table 1.H.3 shows annual total house price growth variation and its decomposition across the MSA-size distribution for the period between 1990 and 2020. Following Giacoletti (2021), we define total house price growth variation as the sum of idiosyncratic risk and local house price risk. We measure local house price risk as the standard deviation of the yearly growth of the local house price index. We divide

the 248 MSAs into increasing size bins according to their population in 1990. The first row shows that local risk increases slightly with MSA size. This finding might seem counter-intuitive at first glance,[57] but can be explained by the observation that large urban centers tend to have tighter housing supply constraints,[58] which amplify shocks to house prices leading to higher house price index volatility.[59] Conversely, idiosyncratic risk is substantially smaller in the largest cities and clearly decreases with MSA-size.

Next, we look at total house price risk. As idiosyncratic house price risk represents the major share of total house price risk across the entire MSA-size distribution (Row 3), the pattern of idiosyncratic risk across MSAs is reflected in the distribution of total risk. Consequently, Row 4 of Table 1.H.3 reveals that total risk also decreases with MSA-size. While the smallest MSAs had on average an annual total house price risk of 13.61% of the sales price of a house between 1990 and 2020, the largest MSAs had a considerably lower total risk of 11.63% relative to the sales price.

Appendix 1.I Rental yield risk and city size

In this section, we provide evidence on spatial differences in rental yield volatility. Rental yields at the property level are defined as the rental income of a property divided by its potential sales price. Consequently, volatility in rental yields can have two possible sources: changes in rental income or changes in the sales price. Changes in rental yields driven by changes in the sales price are negatively related to changes in capital gains. To see why, consider the following simplified example: Assume a property at time t has a rental yield of 5%. At time $t+1$, its price doubles, but the rental income stays constant. This leads to a capital gain of 100 percentage points in t+1, but its rental yield is reduced to 2.5%, such that total returns only change by 97.5 percentage points. The negative co-variance between rental yields and capital gains at the property level attenuates capital gain volatility, but only to a small extent.[60]

The other source of rental yield volatility are changes in the rental income of a property. As done in section 1.5.3 for capital gains, we can also decompose volatility in rents in a location-wide and an idiosyncratic component. The problem is that, to the best of our knowledge, no data set exists that covers rental income at the property level over a long-enough time period for a cross-section of cities. Still, in the

57. This result is, however, not new, but has already been shown for example in Bogin, Doerner, and Larson (2018).

58. See, for example, Saiz (2010).

59. See Paciorek (2013) for a theoretical and empirical explanation of the relation between housing supply constraints and house price index volatility.

60. Eichholtz, Korevaar, Lindenthal, and Tallec (2020) also find a negative co-variance of rental yields and capital gains empirically.

remainder of this section we show empirical evidence that suggests that, if anything, both components of rental income risk are lower in large cities.

First, we analyze location-wide rent risk. Unfortunately, there does not exist a data set with long-run annual rent data on city- or MSA-level for the U.S. However, the German data set we constructed and use in section 1.4.2 does feature rent indices for a large cross-section of German cities.[61] We use these data to calculate location-wide rent volatility on city level. Figure 1.I.1 plots volatility in annual rent growth by city size. For both samples, one of 42 cities for the period between 1975 and 2018 (left hand side) and the other of 127 cities between 1993 and 2018 (right hand side), rent growth volatility is smaller in larger cities.

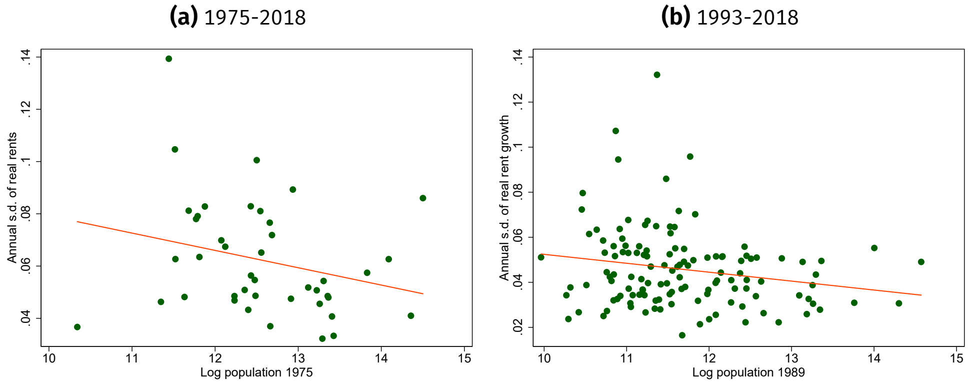

Figure 1.I.1. Real rent growth volatility and population, Germany

Notes: Standard deviation of real rent growth for 42 German cities between 1975 and 2018 (Panel (a)) and for 127 German cities between 1993 and 2018 (Panel (b)). More details on the data sources can be found in section 1.4.2 of the paper and Appendix 1.C.2.

Next to changes in the location-wide rent level, changes in rental vacancies also induce volatility in rental income of a property. One the one hand, for a large-scale investor with a high number of rental units within a city, volatility of city-level vacancy rates add to location-wide rental income risk. On the other hand, for a small property owner with only one rental unit, a higher city-level vacancy rate induces a higher idiosyncratic risk, because it increases the probability that his one unit is vacant. We use data from the American Housing Survey from the period between 1985 and 2020 for 49 MSAs to compare vacancy rates between large and smaller MSAs. The results can be found in Table 1.I.1. It shows that the mean as well as the standard deviation of annual rental vacancies is lower in large cities.

Both pieces of evidence suggest that location-wide risk in rental income is smaller in large cities. Regarding idiosyncratic risk, as stated above, there does not exist any data set we know of that would enable us to compare this risk component

61. For details on this data set please refer to Appendix 1.C.2.

Table 1.I.1. Differences in mean and standard deviation of rental vacancies in p.p., US, 1985-2020

Sample	Mean	N	S.d.	N
Large vs rest	-2.06*(1.093)	1372	-0.73***(0.169)	1372
Large vs small	-1.25(1.415)	168	-1.06***(0.274)	168

Notes: The Table shows the difference in rental vacancy rates between the 5% largest MSAs in terms of 1970 population relative to the other MSAs in the sample (Row 1) and to the 5% smallest MSAs (Row 2). The data covers 49 MSAs for the period between 1985 and 2020 and is collected from the American Housing Survey.

between cities. However, as we argue in section and is shown by Giacoletti (2021), Sagi (2021) and Kotova and Zhang (2019), idiosyncratic risk in capital gains is mainly driven by liquidity in the housing market. As the rental market is not fundamentally different from the house sales market, we also expect liquidity to play a considerable role for idiosyncratic risk of rental income. Both information as well as location-wide demand and supply for rental units will determine rental income risk of an individual property to a large extent.

Unfortunately, we cannot use the liquidity measures for the US for the rental market that we use for the house sales market. However, the mean difference in rental vacancy rates already points at higher liquidity in the rental market in large US cities. Additionally, we can replicate the two measures we use for liquidity in Germany also for the rental housing market. Figure 1.I.2 shows the results, which are, if anything, even stronger then for the house sales market and highly significant. This result shows that liquidity is larger in large cities also in the rental market. This strengthens the assumption that idiosyncratic rental income risk is, if anything, smaller in large cities.

To summarize, the evidence presented in this section is only suggestive, because we cannot calculate rental yield volatility at the property level for a cross-section of cities. However, each piece of evidence points at a lower rental yield volatility in large cities compared to smaller ones. As capital gain volatility represents the larger share of total housing returns volatility, this evidence suggests, that, if anything, including rental yields volatility would reduce the difference in Sharpe ratios between large and small cities even further.

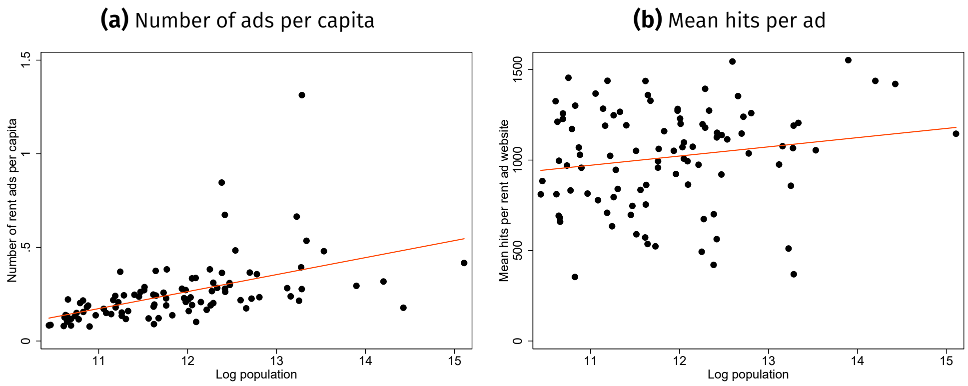

Figure 1.I.2. Thickness of the rental market by city size, Germany

Notes: The figure shows the number of rental real estate advertisements per capita (Panel (a)) and the median clicks per rent advertisement (Panel (b)) on city level for 98 German independent city counties (*kreisfreie Städte*) between 2007 and 2019 by population size in 2015. All data is from the largest German listing website for real estate *ImmoScout24*. In a regression including year fixed effects, log population is significant at the 1%-level for both panels. For details about the data source please refer to Klick and Schaffner (2020).

1.I.1 Risk-adjusted housing returns

In the last sections we showed evidence that housing risk is lower in large cities compared to smaller ones and that differences in housing market liquidity might explain this finding. The question arises whether these risk differences are able to explain the return differences we found in the first part of the paper. To make a first step to answer this question we test whether there are systematic differences in risk-adjusted returns across cities in the remainder of this paper. The simple asset pricing framework we outlined at the beginning of section 1.I, predicts returns to converge across city size bins after properly adjusting for risk. Nevertheless, assuming risk-averse investors, we would also expect there to remain some differences across cities.

We use Sharpe ratios as a simple measure of risk adjusted returns. To approximate housing Sharpe ratios at the MSA level, we merge our total risk estimates with total return data from the US Census and the American Community Survey for the period between 1990 and 2018.[62] This data is already used to extend the US data set used in section 1.4.1 and described in more detail in Appendix 1.C.1. Additionally, we also use data on the real returns of US T-bills from Jordà, Knoll, et al. (2019) to construct excess total real housing returns. We calculate Sharpe ratios as:

62. Theoretically, it is also possible to construct total return estimates using solely Zillow.com data. However, Zillow only has rent estimates for 100 large MSAs. This would severely restrict the sample of our analysis.

$$\textit{Sharpe ratio}_m = \frac{\textit{total return}_m - \textit{return}^{tbill}}{\sigma_{m,total}},$$

where $\textit{total return}_m$ is the average of the log sum of housing capital gain and rental yield for MSA m and $\textit{return}^{tbill}$ is the average log US T-bill rate over the same time period.

Table 1.I.2. Housing Sharpe ratio and its decomposition by MSA size, 1990-2020

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Sharpe ratio	0.40	0.35	0.38	0.39	0.38	0.38	0.37	0.40	0.34	0.42	0.38	0.36
Total excess return	5.01	4.44	4.68	4.61	4.37	4.24	4.31	4.43	3.63	4.46	4.25	4.09
Total risk	13.61	13.17	13.11	12.80	12.47	11.86	12.18	11.64	11.18	10.96	11.54	11.63
Number of MSAs	13	12	25	25	24	25	25	24	25	25	12	12

Notes: MSAs are divided into bins based on the size of MSA population in 1990. The bins go from the smallest MSAs (bin 1A) to the largest MSAs (bin 10B). The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins split the smallest and largest deciles in half.

Approximated Sharpe ratios as well as excess total housing returns and total risk by MSA size bins can be found in Table 1.I.2.[63] In contrast to total housing returns, Sharpe ratios do not seem to differ systematically over the city size distribution. Not surprisingly, the remaining differences in Sharpe ratios between large cities and smaller ones are not significant anymore, as can be seen in Table 1.I.3. This is in line with the theoretical prediction discussed above that higher housing returns in smaller cities are a compensation for higher housing risk.

Table 1.I.3. Differences in Sharpe ratio, 1990-2020

Sample	Sharpe ratio	N
Large vs rest	-0.02(0.025)	247
Large vs small	-0.05(0.046)	25

Notes: The table shows differences in average Sharpe ratio between large MSAs and the rest of the sample or small MSAs. Differences are measured as coefficients in a cross-sectional regression of the dependent variable (Sharpe ratio) on a large MSA dummy. Standard errors (in parenthesis) are clustered at the MSA-level. Large MSAs are defined as being at or above the 95th percentile of the MSA population distribution in 1990. The second row shows the same, but comparing large MSAs only to small MSAs, which are defined as being at or below the 5th percentile of the MSA population distribution in 1990. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$

This first-order approximation of Sharpe ratios still has some drawbacks. While we take rental yields into account for the construction of housing return estimates, we do neglect rental yields when measuring housing risk. Unfortunately, as discussed

63. To calculate housing returns here, we use the same sample and time period as for our housing risk estimates. Therefore, the housing returns differ from long-run housing returns in section 1.4.1. Still, the relation between housing returns and city size also shows up for the smaller sample and shorter time period.

above, due to data limitations we are not able to estimate the risk associated with rental yields on property level. Under the assumption that most of the variation in returns actually comes from capital gains variation, we think that our approach is still a reasonable approximation of the actual Sharpe ratios. In appendix 1.I we additionally show evidence that suggests that including volatility of rental yields would reduce risk in large cities relative to small ones even further. This way, the remaining gap in Sharpe ratios between the largest cities and the rest might reduce further. A complete housing risk profile including rental yields is left for future research.

Appendix 1.J Additional results on housing liquidity

1.J.1 Housing liquidity over the MSA-size distribution in the US

Table 1.J.1. Cross-sectional differences of time on the market for 277 MSAs, 2012-2020

	1	2	3	4	5	6	7	8	9	10
mean	114.92	97.31	107.26	98.96	107.84	101.26	93.60	99.61	89.69	85.56
sd	39.37	27.72	29.42	30.54	32.51	27.81	26.55	26.98	24.86	25.69

Notes: MSAs are divided into decile bins based on the size of MSA population in 2010. Decile represents the 10% smallest MSAs. Each bin contains between 27 and 28 MSAs. Data on the median number of days on Zillow from Zillow.com for 277 MSAs for the period between 2012 and 2020.

Table 1.J.2. Cross-sectional differences of asking price discount in p.p. for 277 MSAs, 2012-2020

	1	2	3	4	5	6	7	8	9	10
mean	114.92	97.31	107.26	98.96	107.84	101.26	93.60	99.61	89.69	85.56
sd	39.37	27.72	29.42	30.54	32.51	27.81	26.55	26.98	24.86	25.69

Notes: MSAs are divided into decile bins based on the size of MSA population in 2010. Decile represents the 10% smallest MSAs. Each bin contains between 27 and 28 MSAs. Data on the average discount to the asking price from Zillow.com for 277 MSAs for the period between 2012 and 2020.

1.J.2 House sale liquidity in Germany

We analyze two liquidity measures for Germany, which are connected to the thickness of the housing market. Using data from the online real estate marketplace *immobilienscout24.de*, we test whether large cities in Germany have a stronger supply and demand for housing. We first look at the supply side by analyzing the number of sales ads posted per capita in each city. The results can be found in panel (a) of Figure 1.J.1. It shows that in larger cities there are significantly more ads posted per capita. This indicates that even on a per capita basis, housing supply is larger in large cities.

We next quantify demand for housing. To do so, we look at the number of hits per sales ad by city. Figure 1.J.1 panel (b) shows that in large cities housing ads receive substantially and significantly more hits, and therefore have more potential buyers, than in small cities. This indicates that, even relative to a higher supply, demand per supplied unit is substantially larger in large cities.

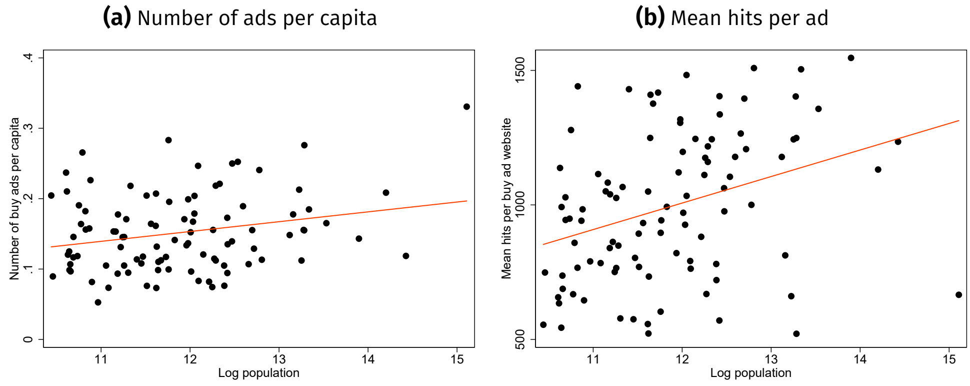

Figure 1.J.1. Thickness of the housing market by city size, Germany

Notes: The figure shows (a) the number of real estate sales ads per capita and (b) the median clicks per sales ad on city level for 98 German independent city counties (*kreisfreie Städte*) between 2007 and 2019 by population size in 2015. All data is from the largest German listing website for real estate *ImmoScout24*. In a regression including year fixed effects, log population is significant at the 1%-level for both panels. For details about the data source please refer to Klick and Schaffner (2020).

The results based on German data are very insightful because they measure liquidity on a per sale or per capita basis. The fact that there are mechanically more sales and inhabitants in larger cities amplifies the effect. Other local housing market characteristics might additionally reinforce the link between larger liquidity and lower risk in large cities. For example, large cities might have more institutionalized housing markets, which further reduce matching frictions and can make better use of the more abundant information from comparison prices.[64]

1.J.3 Real estate liquidity of institutional portfolios in European cities

Finally, we document how the big real estate transactions, residential and total, as recorded by Preqin rather take place in cities of bigger size in European cities of the 2010s.

64. For example in Germany, the *Gutachterausschüsse* in larger cities publish shadow-prices for housing characteristics. The quality of such estimates and the level of detail possible increases noticeably with sample size and these estimates help to get a more accurate approximation of the value of a specific building.

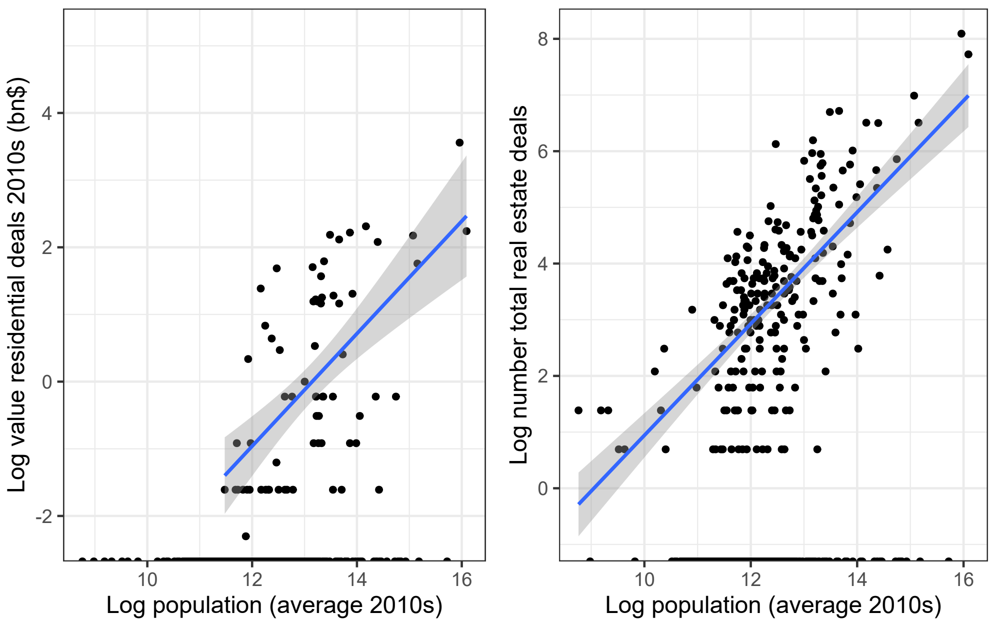

Figure 1.J.2. Liquidity of housing markets in European cities

Notes: Preqin data for big deals of institutional investors, total sum since 2011 and Eurostat population data averaged for the 2010s.

Appendix 1.K Data appendix for 27 large cities

In this appendix, we describe the methods and sources used to build our new long-run city-level housing return data set. In the following, there is a subsection for every city, which are each divided into three parts. We first describe how we built the price series, then the rent series and, finally, the rental yield series. The focus of this section is on the sources and methods used in our final series, but we also provide an overview of alternative sources, whenever we know of their existence. As we mentioned in the text, we used the rent-price approach to build the rental yield series. This approach can lead to the accumulation of measurement errors over time. For this reason, we also show when and how we used historical benchmarks to correct our rental yield series.

We would like to thank the following persons and agencies for their very valuable contributions to the data collection process, without which this project would not have been possible:

Australia: Dimitrios Kanelis for his outstanding help in the data collection process.

Denmark: David Gosselin and Karl Schweizer for assisting in making data from *boligportal* and the *Berlingske* accessible.

Finland: Statistics Finland for providing us with long-run rent series for Helsinki.

Germany: Werner Brinkmann and Ursula Zimmermann (*Rheinische Immobilienbörse e.V.*) for assistance with the data in Cologne; Werner Brinkmann and Julian Götting (*Immobilienverband Deutschland* (IVD)) for the assistance with the data from the IVD and RDM we used to construct the German data set used in the main paper; Susanne Düwel and the *Gutachterausschuss für Grundstückswerte in Berlin* for enabling us to use their transaction-level data for Berlin; Almut Pantke, Dieter Hagemann and the *Gutachterausschuss für Grundstückswerte in der Stadt Köln* for enabling us to use their transaction-level data as well as providing maps of archived transaction data for Cologne; Michael Debus, Christine Helbach, Sandra Schilling, Stefanie Schäfer and the *Gutachterausschuss für Immobilienwerte* in Frankfurt for enabling us to use their transaction-level data as well as granting us access to their archives for Frankfurt.

Italy: Anita Braga and Andrea Balestracci for assisting us in the construction of house price and rent series as well as for the collection of newspaper data from Milan.

Japan: Masashi Tanigaki for his great work on the data collection for Tokyo.

Netherlands: Matthijs Korevaar for valuable help with the data sources.

Norway: Liudmila Kiseleva for extracting data from Oslo's newspapers and Are Oust for consultation.

Spain: Andrea Balestracci and Marta Barceló for their great work on the data collection of Barcelona and Madrid.

Sweden: Wolfram März and Tina Bergsten for accessing the data from Göteborgposten and Roger Gustafsson from Boverket for making regional rent survey data accessible.

UK: Matthijs Korevaar and Avner Offer for valuable help with the data sources.

US: Jason Barr for providing us with data on New York and extremely helpful insights.

1.K.1 Australia

The six largest cities in Australia in 1900 were in descending order: Melbourne, Sydney, Adelaide, Brisbane, Newcastle and Ballarat. All of them encompassed more than one percent of the country population in 1900.[65] According to the algorithm described in the main paper, we took the two largest cities, Melbourne and Sydney, and then stopped, as we already covered more than 25 percent of the country's population in 1900. Long-run housing data are available for both cities. For the other cities, in contrast, long-run housing data are sparse.

1.K.1.1 Melbourne

House Price Series. Stapledon (2012) builds a long-run house price index for Melbourne for the period between 1880 and 2011. His index is based on real estate ads from the newspaper *Melbourne Age* until 1970. For the sub-period 1880–1943, the index is computed from the median asking price for all types of residential buildings, indiscriminate of their characteristics; for 1943–1949, Stapledon (2012) estimates a fixed price;[66] for 1950–1970, he uses the median sales prices from housing auctions, which were also reported in the newspaper *Melbourne Age* for this period.

For the post-1970 period, Stapledon (2012) uses both Abelson and Chung (2005) as well as Australian Bureau of Statistics (2020), which we explain in more detail below. Stapledon (2012) also discusses an existing historical asking price series for Melbourne from Butlin, which covers the period between 1861 and 1891 and that is consistent with his own series.

Abelson and Chung (2005) build a median house price index for Melbourne for the period between 1970 and 2003. For this series they rely on: (i) a 1991 study by *Applied Economics and Travers Morgan*, which draws on sales price data from the *Land Title Offices*, for the period between 1970 and 1979, and (ii) on sales price data from the Department of Housing, i.e. the *Victorian Valuer-General Office*, for the period between 1980 and 2003.

65. City-level population data are taken from Reba, Reitsma, and Seto (2016) and country-level population data from Jordà, Schularick, and Taylor (2017).

66. A 1943 law fixed nominal house prices in Australia to 1942 levels during this period. The house price freeze ended in 1949, which explains the strong jump in the series between 1949 and 1950.

The *Australian Bureau of Statistics* (ABS) published a quarterly index for Melbourne for the period between 1986 and 2005 for i) established detached residential dwellings and ii) *project homes*, i.e. dwellings available for construction on a client's block, excluding land. The indices are calculated using a stratification method, where the transaction prices are stratified by geographical area and physical characteristics of the dwelling (Australian Bureau of Statistics, 1993).

Since 2004, ABS also publishes quarterly indices for Melbourne for i) established detached dwellings and ii) attached homes. They additionally publish a third index, which is an aggregation of the first two, which we use to construct our final Melbourne house price index. The indices are calculated using a stratification method. Locational, structural and neighborhood characteristics are used to mix-adjust the index, i.e. to control for compositional change in the sample of houses. Each quarter, the strata are re-valued by applying a price relative (i.e. the current period median price of the stratum compared to the previous period median price of the same stratum) to the value of the dwelling stock for that stratum to produce a current period stratum value. The series are constructed as Laspeyres-type indices. Sales price data are taken from the State Valuer-General Offices and is supplemented by data on property loan applications from major mortgage lenders (Australian Bureau of Statistics, 2018).

Table 1.K.1 summarizes the main components of our final house price index. From 1880 to 1970 we rely on the index by Stapledon (2012), because it is the only long-run index and, as mentioned above, correlates very strongly with the other existing indices for the overlapping periods. We then splice the index from Stapledon (2012) with the one from Abelson and Chung (2005) for the period between 1970 and 1986. Abelson and Chung (2005) discuss all existing indices for this period and their choice of final index is based on the representativeness of the underlying data. From 1986 onward, we use the indices produced by the Australian Bureau of Statistics, because these are the only available indices, which adjust for quality and compositional changes by applying a stratification index methodology. This is the only period where our series differs from the one in Stapledon (2012).

Rent Series. Stapledon (2007) builds a rent index for Melbourne between 1901 and 1954 using adjusted rent estimates for census years (1911, 1921, 1933 and 1947). The estimates are the weighted average of gross rents of tenant-occupied dwellings and the imputed rents of owner-occupied dwellings. The imputed rents are calculated using a 20% premium. These estimates are interpolated using the national CPI rent component from ABS. Whenever the average change between the estimated census years is higher or lower than the change in the national rent index, the author adjusts his results proportional to the difference between rent index rate of change and average change between census data points.

To the best of our knowledge there does not exist a rent index for Melbourne for the period between 1955 and 1972. To fill this gap, we collected data on average

Table 1.K.1. Final house price index for Melbourne

PERIOD	SOURCE	DESCRIPTION
1880-1970	Stapledon (2012)	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Newspaper ads; *Method*: Median asking prices until 1943 and median sales prices after 1950.
1971-1985	Abelson and Chung (2005)	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Transaction data from Land Title Offices, Productivity Commission and Valuer-General Offices; *Method*: Median prices.
1986-2003	ABS	*Type(s) of dwellings*: New and existing detached houses; *Type of data*: Transaction data from real estate organizations and government agencies; *Method*: Stratification index, where the transactions are stratified by geographical area and physical characteristics of the dwelling.
2004-2018	ABS	*Type(s) of dwellings*: New and existing detached and attached dwellings; *Type of data*: Data obtained from state and territory land title offices or Valuer-General offices, and real estate agents' data provided by *CoreLogic*; *Method*: Stratification index, where the transactions are stratified by dwelling type, long-term median price and a Socio-economic Index for Areas score.

weekly rent for tenant-occupied houses and apartments for the city of Melbourne from the Local Government Areas editions of the Australian Census of Population and Housing for the years 1954, 1961, 1966, 1971 and 1976. The city of Melbourne was already to a great extent urbanized by 1954, in contrast to the region covered by the Local Government Area of Melbourne. Therefore, our estimates should not be influenced by changes in the urban/rural mix. Like Stapledon (2007), we then interpolated the census years using the national CPI rent component series from ABS.

Since 1972, ABS provides a decomposition of the national CPI rent component for eight capital cities, including Melbourne, on a quarterly frequency. These indices are built using the same methodology as for the national series. Rental prices are obtained from real estate agents and territory housing authorities under a matched sample approach, i.e. rents are collected for the same sample of tenant-occupied dwellings every quarter. The samples are stratified according to location, dwelling type, and size of dwelling based on the most recent Census of Population and Housing (Australian Bureau of Statistics (2018)).

Rental Yield Series. For 2018, CoreLogic (2018) reports a gross residential rent-price ratio of 2.8%, which results from a price and rent estimate for the average

Table 1.K.2. Final rent index for Melbourne

PERIOD	SOURCE	DESCRIPTION
1901-1954	Stapledon (2007)	*Type(s) of dwellings*: All kinds of tenant and owner-occupied residential dwellings; *Type of data*: Census data; *Method*: Weighted-average gross rents.
1955-1972	Own compilation	*Type(s) of dwellings*: Tenant-occupied residential houses and apartments; *Type of data*: Census data; *Method*: Average weekly gross rent.
1973-2018	ABS	*Type(s) of dwellings*: All kinds of rented residential dwellings; *Type of data*: Rent data are collected from real estate agents and state and territory housing authorities; *Method*: Matched sample approach.

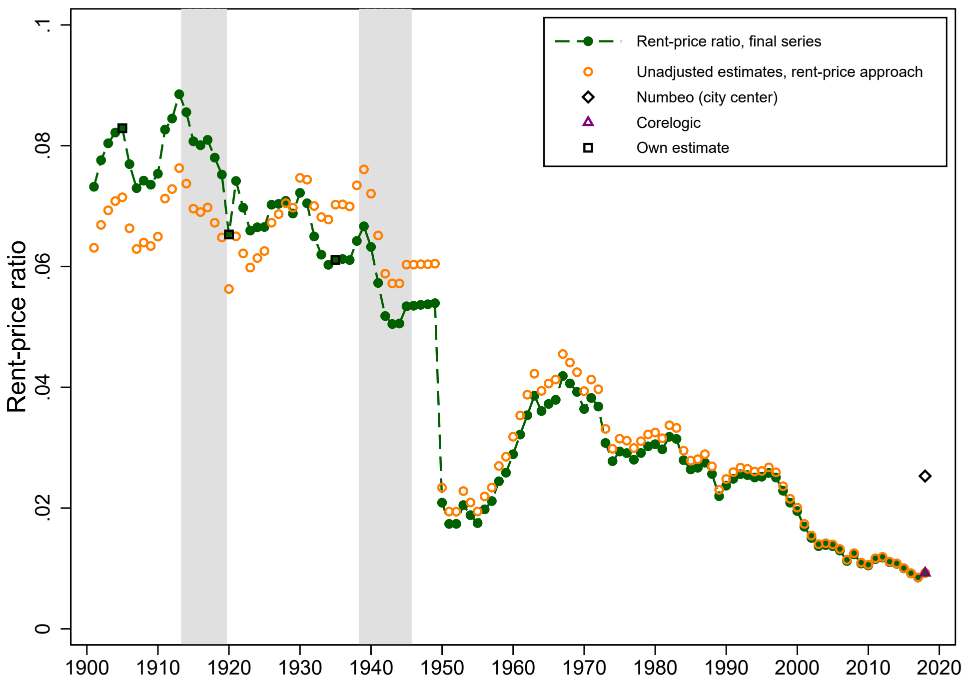

Figure 1.K.1. Melbourne: plausibility of rental yields

house in Melbourne. Fox and Tulip (2014) estimate that running costs and depreciation costs amounted to 2.2% of house prices in 2014. We update their estimate to 2018 using price and rent inflation in Melbourne and estimate that in 2018 the costs amount to 1.87% of house prices. Thus, we estimate the net rent-price ratio to be 0.93% in 2018. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in Figure 1.K.1. We make some adjustments to these series to correct for possible mismeasurement of rental growth when the wartime price controls were lifted in 1949/50 (see below for details). This gives us the adjusted final rent-price ratio series—the green-circled line in Figure 1.K.1.

We collected an additional rent-price ratio estimate from *Numbeo.com* for 2018, which is also shown in Figure 1.K.1. *Numbeo.com* reports a gross rent-price ratio of 4.27% for 2018. If we apply the same cost estimate described above, we get to a net estimate of 2.3%, which is higher than the CoreLogic (2018) estimate. Since the *CoreLogic* estimate is derived using a hedonic approach, we decided to choose it over the estimate from *Numbeo.com*, which is a simple average. Furthermore, the *CoreLogic* estimates are constructed using actual price and rent data from land titles and notaries, while the *Numbeo.com* estimates are made using the price and rent estimates given by the website users.

As is clearly visible from Figure 1.K.1, the long-run rent-price ratio shows a structural break in 1949/1950 caused by a surge in house prices after the lifting of wartime price controls in 1949 (price controls for houses and land were introduced in 1942). While it is clear that the end of price controls increased house prices, it is harder to say what effects it had on the relationship between prices and rents. Therefore, we decided to use additional historical benchmarks to adjust our series. Unlike in the case of Sydney, we could not find historical benchmarks for Melbourne. Therefore, we created our own benchmarks using newspaper ads containing asking prices and rents for the exact same residential properties in the city center of Melbourne, i.e. we did not include observations from the suburbs of Melbourne. We arrived at the following estimates of gross rental yields: 11.47% in 1905, 9.77% in 1920 and 8.7% in 1935. We then use the estimates of maintenance costs and depreciation, reported in Table A.22 in Stapledon (2007), to which we add insurance and commissions and deduct property taxes, reported by the *Australian Bureau of Statistics* for census years, which can also be found in Stapledon (2007). Since we only get cost estimates for the years 1900, 1910, 1920, 1932 and 1938, we linearly interpolate the costs for our benchmark years. As a result, we estimate that net yields are 8.29% in 1905, 6.53% in 1920 and 6.11% in 1935. The estimates are shown in the black boxes in Figure 1.K.1.

1.K.1.2 Sydney

House Price Series. Stapledon (2012) builds a long-run house price index for Sydney for the period between 1880 and 2011. His index is based on real estate ads from the newspaper *Sydney Morning Herald* until 1970. For the sub-period 1880–1943, the index is computed from the median asking price for all types of residential buildings, indiscriminate of their characteristics; for 1943–1949, Stapledon (2012) estimates a fixed price;[67] for 1950–1970, he uses the median sales price, which was also reported in the newspaper *Sydney Morning Herald* for this period. For the post-1970 period, Stapledon (2012) uses both Abelson and Chung (2005) as well as Australian Bureau of Statistics (2020), which we explain in more detail below. Stapledon (2012) also discusses other existing historical price series for Sydney, from Abelson (1985) and Neutze (1972) for the pre-1970 period, and shows that they are strongly correlated with his own series.[68]

Abelson and Chung (2005) build a median house price index for Sydney for the period between 1970 and 2003. For this series they rely on: (i) a 1991 study by *Applied Economics and Travers Morgan*, which draws on sales price data from the *Land Title Offices*, for the period between 1970 and 1979, and (ii) on sales price data from the Department of Housing, i.e. the *North South Wales Valuer-General Office*, for the period between 1980 and 2003.

The *Australian Bureau of Statistics* (ABS) published a quarterly index for Sydney for the period between 1986 and 2005 for i) established detached residential dwellings and ii) *project homes*, i.e. dwellings available for construction on a client's block, excluding land. The indices are calculated using a stratification method, where the transaction prices are stratified by geographical area and physical characteristics of the dwelling (Australian Bureau of Statistics, 1993).

Since 2004, ABS also publishes quarterly indices for Sydney for i) established detached dwellings and ii) attached homes. They additionally publish a third index, which is an aggregation of the first two, which we use to construct our final Sydney house price index. The indices are calculated using a stratification method. Locational, structural and neighborhood characteristics are used to mix-adjust the index, i.e. to control for compositional change in the sample of houses. Each quarter, the strata are re-valued by applying a price relative (i.e. the current period median price of the stratum compared to the previous period median price of the same stratum) to the value of the dwelling stock for that stratum to produce a current period stratum value. The series are constructed as Laspeyres-type indices. Sales price data are taken from the State Valuer-General Offices and are supplemented by data on property loan applications from major mortgage lenders (Australian Bureau of Statistics, 2018).

Table 1.K.3 summarizes the main components of our final house price index. From 1880 to 1970 we rely on the index by Stapledon (2012), because it is the only long-run index and, as mentioned above, correlates very strongly with the other existing indices for the overlapping periods. We then splice the index from Stapledon

67. A 1943 law fixed nominal house prices in Australia to 1942 levels during this period. The house price freeze ended in 1949, which explains the strong jump in the series between 1949 and 1950.

68. For a clear graphical comparison of the available series for Sydney see Figure 20 of the Online Appendix of Knoll, Schularick, and Steger (2017)

(2012) with the one from Abelson and Chung (2005) for the period between 1970 and 1986. Abelson and Chung (2005) discuss all existing indices for this period and their choice of final index is based on the representativeness of the underlying data. From 1986 onward, we use the indices produced by the Australian Bureau of Statistics, because these are the only available indices which adjust for quality and compositional changes by applying a stratification index methodology. This is the only period where our series differs from the one in Stapledon (2012).

Table 1.K.3. Final house price index for Sydney

PERIOD	SOURCE	DESCRIPTION
1880-1970	Stapledon (2012)	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Newspaper ads; *Method*: Median asking prices until 1943 and median sales prices after 1950.
1971-1985	Abelson and Chung (2005)	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Transaction data from Land Title Offices, Productivity Commission and Valuer-General Offices; *Method*: Median prices.
1986-2003	ABS	*Type(s) of dwellings*: New and existing detached houses; *Type of data*: Transaction data from real estate organisations and government agencies; *Method*: Stratification index, where the transactions are stratified by geographical area and physical characteristics of the dwelling.
2004 - 2018	ABS	*Type(s) of dwellings*: New and existing detached and attached dwellings; *Type of data*: Data obtained from state and territory land titles offices or Valuer-General offices, and real estate agents' data provided by *CoreLogic*; *Method*: Stratification index, where the transactions are stratified by dwelling type, long-term median price and a Socio-economic Index for Areas score.

Rent Series. Stapledon (2007) builds a rent index for Sydney between 1901 and 1954 using adjusted rent estimates for census years (1911, 1921, 1933 and 1947). The estimates are the weighted average of gross rents of tenant-occupied dwellings and the imputed rents of owner-occupied dwellings. The imputed rents are calculated using a 20% premium. These estimates are interpolated using the national CPI rent component from ABS. Whenever the average change between the estimated census years is higher or lower than the change in the national rent index, the author adjusts his results proportional to the difference between rent index rate of change and average change between census data points.

To the best of our knowledge there does not exist a rent index for Sydney for the period between 1955 and 1972. To fill this gap, we collected data on average weekly rent for tenant-occupied houses and apartments for the city of Sydney from the Local Government Areas editions of the Australian Census of Population and Housing for the years 1954, 1961, 1966, 1971 and 1976. Since the region covered by the city of Sydney was already fully urbanized by 1954, in contrast to the region covered by the Local Government Area of Sydney, our estimates are not influenced by changes in the urban/rural mix. Like Stapledon (2007), we then interpolated the census years using the national CPI rent component series from ABS.

Since 1972, ABS provides a decomposition of the national CPI rent component for eight capital cities, including Sydney, on a quarterly frequency. These indices are built using the same methodology as for the national series. Rental prices are obtained from real estate agents and territory housing authorities under a matched sample approach, i.e. rents are collected for the same sample of tenant-occupied dwellings every quarter. The samples are stratified according to location, dwelling type and size of dwelling based on the most recent Census of Population and Housing (Australian Bureau of Statistics (2018)).

Table 1.K.4. Final rent index for Sydney

PERIOD	SOURCE	DESCRIPTION
1901 - 1954	Stapledon (2007)	*Type(s) of dwellings*: All kinds of tenant and owner-occupied residential dwellings; *Type of data*: Census data; *Method*: Weighted-average gross rents.
1955 - 1972	Own compilation	*Type(s) of dwellings*: Tenant-occupied residential houses and apartments; *Type of data*: Census data; *Method*: Average weekly gross rent.
1973 - 2018	ABS	*Type(s) of dwellings*: All kinds of rented residential dwellings; *Type of data*: Rent data are collected from real estate agents and state and territory housing authorities; *Method*: Matched sample approach.

Rental Yield Series. For 2018, CoreLogic (2018) reports a gross residential rent-price ratio of 3.1%, which results from a price and rent estimate for the average house in Sydney. Fox and Tulip (2014) estimate that running costs and depreciation costs amounted to 2.2% of house prices in 2014. We update their estimate to 2018 using price and rent inflation in Sydney and estimate that in 2018 the costs amount to 1.95% of house prices. Thus, we estimate the net rent-price ratio to be 1.15% in 2018. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rent-price ratio series depicted as orange circles in Figure 1.K.2. We make some adjustments to these series to correct for possible mismeasurement of rental growth when the wartime price controls were lifted in 1949/50 (see below

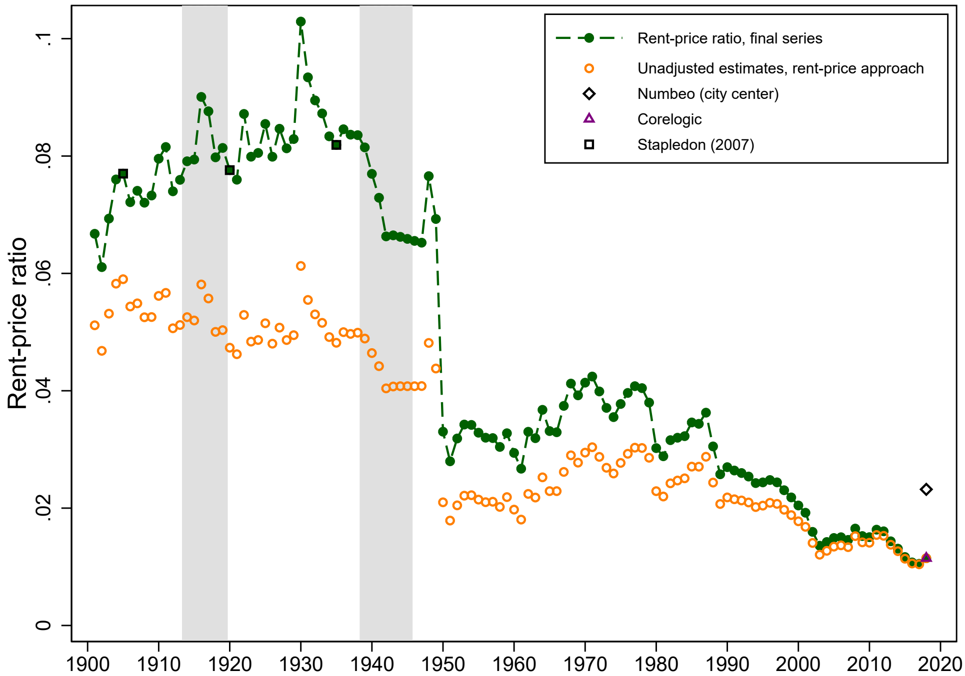

Figure 1.K.2. Sydney: plausibility of rental yields

for details). This gives us the adjusted final rent-price ratio series—the green-circled line in Figure 1.K.2.

We collected an additional rent-price ratio estimate from *Numbeo.com* for 2018, which is also shown in Figure 1.K.2. *Numbeo.com* reports a gross rent-price ratio of 4.27% for 2018. If we apply the same cost estimate described above, we get to a net estimate of 2.3%, which is higher than the CoreLogic (2018) estimate. Since the *CoreLogic* estimate is derived using a hedonic approach, we decided to choose it over the estimate from *Numbeo.com*, which is a simple average. Furthermore, the *CoreLogic* estimates are constructed using actual price and rent data from land titles and notaries, while the *Numbeo.com* estimates are made using the price and rent estimates given by the website users.

As is clearly visible from Figure 1.K.2, the long-run rent-price ratio shows a structural break in 1949/1950 caused by a surge in house prices after the lifting of wartime price controls in 1949 (price controls for houses and land were introduced in 1942). While it is clear that the end of price controls increased house prices, it is harder to say what effects it had on the relationship between prices and rents. Therefore, we decided to use additional historical benchmarks from Stapledon (2007) to adjust our series. The estimates are shown in the black boxes in Figure 1.K.2. The gross rent-price ratio estimates for 1905, 1920 and 1935 were built from newspaper ads. We built net rent-price ratios using the estimates on running costs and depreciation combined with tax data also reported in Stapledon (2007). We estimate that costs and taxes represented around 27.8% of prices in 1900, 33.2% in 1920 and 30.4% in 1938. We assume linear changes in costs relative to house prices over time to estimate the net rent-price ratios for the benchmark years.

1.K.2 Canada

According to the Canadian census the five largest cities in Canada in 1901 were in descending order: Montreal, Toronto, Vancouver, Quebec, Ottawa. All of them encompassed more than one percent of the country population in 1901.[69] It is important to note that the estimate for Vancouver is taken as the sum of Burrard and Vancouver city from the Canadian population census from 1901. Burrard officially became part of the city of Vancouver in 1904.

Together, Toronto and Vancouver represented less than 10% of Canada's population in 1901. Unfortunately, we are not aware of historical house price and rent series for other cities in the country. As such, we limit ourselves to these two cities.

Generally speaking there is very limited work on historical housing series for Canadian cities. As shown below, most of the price and rent indices for Toronto and Vancouver are composed of new indices.

For the last few decades, particularly since the 1980s, both Statistics Canada as well as the Canada Mortgage and Housing Corporation have started publishing housing data for different Canadian cities. More recently, the Canadian Real Estate Board started publishing a high-quality house price index at the regional level, which we describe in more detail below.

1.K.2.1 Toronto

House Price Series. Firestone (1951) uses data from the value of real estate transfers in Toronto based on tax records and registry entries to construct a series of total value of residential real estate in Toronto for the period between 1921 and 1949.

Morrison (1978) uses data on single-family and two-family houses to build a repeat-sales index for the city center of Toronto for the period between 1951 and 1973.[70]

The University of British Columbia (UBC) published a house price index for Toronto, which covers the period between 1975 and 2012. The index is based on prices for existing bungalows and two-story executive detached houses in Toronto (Urban Economics and Real Estate, 2013). The index is built by using a population

69. City-level population data are taken from the Canadian Population Census of 1901 (*Fourth census of Canada, 1901* 1901) and country-level population data from Jordà, Schularick, and Taylor (2017).

70. For more details please refer to Revilla (2021)

weighted average of the price change in each neighborhood for which data are available. Subsequently, the index is weighted on changes in the price level of different housing types, i.e. detached bungalows and executive detached houses, according to their share in total units sold. Data are drawn from the house price survey of the real estate company *Royal LePage*.

The *Toronto Real Estate Board* has published historical data from average prices that were listed in the Multiple Listings Service (MLS) since 1975; however, these estimates also include commercial real estate (Board, 2021).

Statistics Canada publishes a house price index based on the price of new dwellings since 1981. The price data on single-family homes, semi-detached homes and townhomes (row or garden homes) come from a survey conducted by *Statistics Canada* on real estate contractors, which covers at least 15% of the total building permit value in Toronto in a given year. Since the survey also includes questions on the characteristics of the properties, *Statistics Canada* is able to construct a matched-model index, in which only similar properties are compared over time (Canada, 2021b).

Since a continuous quality-adjusted long-run index was missing for Toronto, Amaral, Dohmen, Lyons, and Revilla (2021) built a hedonic house price index for the period between 1900 and 1990 for a companion project on long-run regional housing prices in Canada.[71] To construct the series the authors collected asking prices on residential dwellings from the real estate advertisement section of the local newspaper *Toronto Star*. On average, 200 observations per year are used to estimate a house price index based on a hedonic time-dummy adjacent period approach. More precisely, the authors regress the log asking price on time dummies and on the following house characteristics: type of property (house, detached house or apartment), size (number of rooms), location (city center or suburbs), whether the dwelling is new or not and other features (whether it has a garage, swimming pool, air conditioning).[72]

Revilla (2021) compares the long-run index in Amaral et al. (2021) with the existing indices for Toronto. When comparing to the index by Firestone (1951) the long-run index shows less volatility and price appreciation between 1921 and 1949, which we would expect since the series by Firestone do not adjust for quality changes in the sample. With respect to the index by Morrison (1978) and the more recent indices from MLS, UBC and *Statistics Canada* the long-run index shows the exact same trends and a very high level of correlation across time.

Table 1.K.5 summarizes the components of our final house price index. We decided to use series from Amaral et al. (2021) for the period between 1900 and 1990 mainly for two reasons. First, because it is the only continuous long-run house price index for Toronto and starts earlier than all existing series. Second, because it cor-

71. For more details about this project as well as Canadian housing series please refer to Amaral et al. (2021).

72. For more details about the construction of the index please refer to Revilla (2021).

relates very strongly with existing quality-adjusted series for the last decades and, therefore, we preferred to keep the same series over time. For the period after 1990, we opted for the index from Statistics Canada, since it is the only quality-adjusted index which covers the complete period until 2018.

Table 1.K.5. Final house price index for Toronto

PERIOD	SOURCE	DESCRIPTION
1900-1990	Amaral et al. (2021)	*Type(s) of dwellings*: Houses, detached houses and apartments; *Type of data*: Asking prices from the *Toronto Star*; *Method*: Hedonic time-dummy adjacent period index.
1991-2018	Canada (2021b)	*Type(s) of dwellings*: New owner-occupied dwellings; *Type of data*: Transaction prices from housing survey ; *Method*: Matched-model price index.

Rent Series. Amaral et al. (2021) also built a rent price index using real estate ads from the *Toronto Star*. Due to the lack of ads in the first two decades of the 20th century, the rent index starts in 1921. The same methodology as in the house price index is used, i.e. the index is built using a hedonic time-dummy adjacent period method, which controls for the same set of characteristics as listed above in the description of the long-run house price index. Since the number of observations is very low for the period between 1943 and 1950, we considered the index not to be of sufficient quality to be published. As a result the final rent index has a gap between 1943 and 1950.

Statistics Canada publishes a rented accommodation index for the census metropolitan area of Toronto since 1971 as part of the Toronto consumer price index (CPI). The index is based on data from the Labour Force Survey (LFS), which is conducted on a monthly basis for a rotating representative sample, such that the rent of the same dwelling is always recorded for six successive months (Claveau, Lothian, and Gauthier, 2009). This allows *Statistics Canada* to build the rental index by matching the rents for the same dwellings across time, thereby constructing a quality-adjusted rental index. The sampling method used by *Statistics Canada* also implies that only one sixth of the dwellings sampled each period are new in the sample. This means that, by construction, the index mostly tracks the price evolution of existing rental contracts.

Statistics Canada also publishes the average rental value for dwellings of different sizes for the Toronto metropolitan area since 1987. The data come from a yearly survey conducted by the *Canada Mortgage and Housing Corporation* (CMHC) in different Canadian cities including Toronto. The average rental values are built each year for apartments of different sizes in buildings of different sizes. However, the only series which starts already in 1987 is the series for the rental value of apartments in buildings with six or more apartments.[73] We build an average rent index by taking a yearly unweighted average of the rental values for apartments of different sizes in buildings with six or more apartments.[74]

Revilla (2021) builds a decadal rent index for Toronto using census data on average rent paid for the city of Toronto for the period between 1921 and 1981. The author shows that the long-run rental index in Amaral et al. (2021) and the index built using census data follow the same trend between 1920 and 1980.

Revilla (2021) shows that for the period between 1971 and 1990 the indices from Amaral et al. (2021) and from *Statistics Canada* show significantly different trends, with the series in Amaral et al. (2021) growing substantially more in this period. One major difference between the two indices relates to the fact that Amaral et al. (2021) use only data on new rental contracts, which often included rents on newly constructed buildings, while *Statistics Canada* mostly use data on existing rental contracts. This is especially important, since new buildings were not subject to rent controls, which were introduced in 1975 in the province of Ontario.[75] As a result, by construction, the index in Amaral et al. (2021) reflects to a much larger extent rental price evolution in the rental sector not exposed to the rental control laws. This explains the fact that the index grows more in this period than the index by *Statistics Canada*.

Comparing the CPI rent component index from *Statistics Canada* with the CMHC index for the period between 1990 and 2018 produces very similar results. While the CMHC index grows by a factor of approximately 2 in this period, the CPI rent component grows only by 50%. We know that by construction the CMHC index also takes into account rental contracts in newly constructed dwellings, which might explain these differences.

Table 1.K.6 summarizes the components of our final rent index. From 1921 to 1990 we use the index from Amaral et al. (2021). Since our long-run house price index only reflects the price evolution of newly constructed dwellings in Toronto after 1990, we decided to use the CMHC rent index for that period. Using the rent component of the CPI index would create a very large wedge in terms of the types of dwellings being covered by the rent and price series, since the price series only uses data on newly constructed dwellings for the period after 1990.

Rental Yield Series. Our main benchmark for Toronto is taken from *MSCI*, as described in the main paper. This benchmark is slightly above the benchmark we collected for 2018 from *Numbeo.com*. According to *Numbeo.com* the gross rental yield in the city center of Toronto was 4.5% in 2018, adjusting for one-third costs

73. Canada, 2021c.

74. To distinguish this index from the other one published by Statistics Canada, we will refer to this one as the CMHC index.

75. See Revilla (2021) for a detailed description of rent control laws in Canada throughout the twentieth century.

Table 1.K.6. Final rent index for Toronto

PERIOD	SOURCE	DESCRIPTION
1921-1990	Amaral et al. (2021)	*Type(s) of dwellings*: Houses, detached houses and apartments; *Type of data*: Asking rents from the newspaper *Toronto Star*; *Method*: Hedonic time-dummy adjacent period index.
1991-2018	Canada (2021c)	*Type(s) of dwellings*: Apartments in buildings with six or more apartments; *Type of data*: Rents from the Canada Mortgage and Housing Corporation Survey; *Method*: Average rental value.

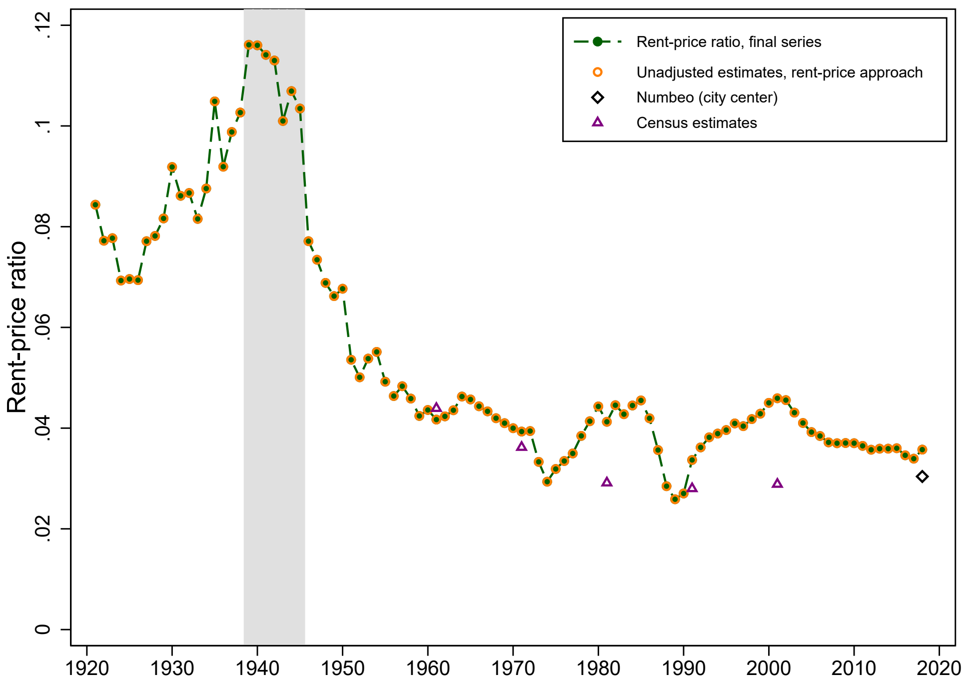

Figure 1.K.3. Toronto: plausibility of rental yields

we estimate a rental yield of 3% for 2018. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.3.

Using census household-level data based on a representative sample of the metropolitan area of Toronto we were able to build average gross yields for the years 1971, 1981, 1991 and 2001 using a matching approach.[76] We built average rental

76. The so-called public use microdata file (PUMF) can be accessed by researchers via *Statistics Canada*.

yields, by matching the average rent of renter-occupied dwellings with the average value of owner-occupied dwellings of the same type and with the same number of rooms within urban areas of the metropolitan area of Toronto.[77] We then aggregate the average gross yields using the number of observations of the respective dwelling type and number of rooms combinations as weights. For the year 1961 the household-level data are not available. Therefore, we build a benchmark for 1961, using the median value of owner-occupied single-detached non-farm houses for the city of Toronto and the average rent paid for renter-occupied non-farm dwellings for the city of Toronto from the national housing census in 1961 (*1961 Census of Canada : Housing Vol. II* 1961).[78] We then adjust the gross yield estimates for one-third costs and get the net yield estimates, which are depicted in Figure 1.K.3. While the census estimates for 1981 and 2001 lie slightly below our unadjusted rental yield series, in general the census estimates match our series quite accurately. As a result we do not adjust our rental yield series.

As is visible from Figure 1.K.3 there is a downward jump in our rental yield series from 1945 to 1946. As we mentioned above, our final rent index has a gap between 1943 and 1950. To build a continuous rental yield series we had to linearly interpolate the rent index for this period. As a result the yearly changes in rental yield series are mostly a product of the changes in the price series for this period. However, to the best of our knowledge, there do not exist rental yield benchmarks for the pre-1950 period which would allow us to better understand the actual rental yields in Toronto for this period.

1.K.2.2 Vancouver

House Price Series. The University of British Columbia (UBC) published a house price index for Vancouver, which covers the period between 1975 and 2012. The index is built using the exact same data sources and methods as the one for Toronto. For more details about the series please refer to the Toronto data appendix.

Amaral et al. (2021) built a hedonic house price index for the city of Vancouver for the period between 1950 and 1984 using asking prices from newspaper real estate ads from the *The Vancouver Sun*. On average, 250 observations per year are used to estimate a house price index based on a hedonic time-dummy adjacent period approach. More precisely, the authors regress the log asking price on time dummies and on the following house characteristics: type of property (house, bungalow or

77. The types of dwellings were the following: single-family houses, semi-detached, rowhouse, duplex, apartment and mobile.

78. Unfortunately, the censuses prior to 1961 do not contain information on the value of dwellings.

duplex), size (number of rooms, bathrooms and kitchen), location (neighborhood) and other features (whether it has a garage, garden or a basement).[79]

Statistics Canada publishes a house price index based on the price of new dwellings since 1981 for Vancouver. The methods and data sources are exactly the same as the ones for Toronto. For more details about the series please refer to the Toronto data appendix.

The *Canadian Real Estate Association* together with the regional real estate associations publishes the MLS HPI index since 2005 for different Canadian metropolitan areas. The index is built based on transaction data from the multiple listing service's database (MLS), which contains practically the whole universe of real estate transactions in Canada. The index is built separately for different types of residential dwellings using a hedonic approach.[80] The hedonic regressions are built on an extensive set of geographical, size of property and local amenities controls.[81] The hedonic regressions for the different residential dwelling types are used to impute the price of a dwelling whose attributes are typical of the dwellings traded in the area where it is located. These imputed prices are then aggregated to a composite price of the metropolitan area using the number of sales as weights. A Fisher chained index methodology is then used to link the imputed prices for the different time periods.

Table 1.K.7 summarizes the components of our final house price index. From 1950 to 1984 we use the hedonic index by Amaral et al. (2021). From 1985 to 2004 we use the price index from *Statistics Canada* on newly built dwellings. After 2005, we rely on the MLS hedonic index. The reason for this is that the index from *Statistics Canada* rises substantially less than the one from MLS after 2005. Since the data and methodologies used to build the MLS index are considerably better, we think that the MLS index is more reliable.

Rent Series. Amaral et al. (2021) built a rent price index for Vancouver for the period between 1950 and 1984 based primarily on real estate rental ads from the *Vancouver Sun*.[82] On average, 250 observations per year are used to estimate a rent price index based on a hedonic time-dummy adjacent period approach. More precisely, the authors regress the log asking yearly rent on time dummies and on the following

79. We have data on the following 22 neighborhoods in the city of Vancouver: Arbutus Ridge, Downtown, Dunbar-Southlands, Fairview, Grandview-Woodland, Hastings-Sunrise, Kensington-Cedar Cottage, Kerrisdale, Killarney, Kitsilano, Marpole, Mount Pleasant, Oakridge, Renfrew-Collingwood, Riley Park, Shaughnessy, South Cambie, Strathcona, Sunset, Victoria-Fraserview, West End and West Point Grey.

80. The different types of dwellings are the following: single-family houses, one-story house, two-story houses, townhouses and apartments.

81. For an exact description of the data and methods used to build the MLS HPI index please refer to Association (2021b).

82. For the year 1981, the authors collected additional observations from the newspaper *Vancouver Heights*.

Table 1.K.7. Final house price index for Vancouver

PERIOD	SOURCE	DESCRIPTION
1950-1984	Amaral et al. (2021)	*Type(s) of dwellings*: House, bungalow or duplex; *Type of data*: Asking prices from the *Vancouver Sun*; *Method*: Hedonic time-dummy adjacent period index.
1985-2004	Canada (2021b)	*Type(s) of dwellings*: New owner-occupied dwellings; *Type of data*: Transaction prices from housing survey; *Method*: Matched-model price index.
2005-2018	Association (2021a)	*Type(s) of dwellings*:Single-family house, townhouses and apartments; *Type of data*: Transaction prices from the MLS database; *Method*: Chained Fisher index based on imputed hedonic approach.

house characteristics: type of property (single-family house, bungalow, apartment or duplex), size (number of rooms, bathrooms and kitchen), location (neighborhood) and other features (whether it has a garage, garden or a basement).[83]

Statistics Canada publishes a rented accommodation index for the census metropolitan area of Vancouver since 1971 as part of the Vancouver consumer price index (CPI). Just like in the case of Toronto, the index is based on data from the Labour Force Survey (LFS). For more details about the series please refer to the Toronto data appendix.

Statistics Canada also publishes the average rental value for dwellings of different sizes for the Vancouver metropolitan area since 1987. As for Toronto, the data come from a yearly survey conducted by the *Canada Mortgage and Housing Corporation* (CMHC). We build an average rent index by taking a yearly unweighted average of the rental values for apartments of different sizes in buildings with six or more apartments.[84]

Table 1.K.8 summarizes the components of our final rent index. From 1950 to 1984 we rely on the hedonic rent index from Amaral et al. (2021), since this index is based on the same method of our house price index for Vancouver for the same period. Between 1984 and 1987 we use the rent component of the CPI by *Statistics Canada*. From 1988 onward we rely on the index on the CMHC average rent index

83. We have data on the following 22 neighborhoods in the city of Vancouver: Arbutus Ridge, Downtown, Dunbar-Southlands, Fairview, Grandview-Woodland, Hastings-Sunrise, Kensington-Cedar Cottage, Kerrisdale, Killarney, Kitsilano, Marpole, Mount Pleasant, Oakridge, Renfrew-Collingwood, Riley Park, Shaughnessy, South Cambie, Strathcona, Sunset, Victoria-Fraserview, West End and West Point Grey.

84. To distinguish this index from the other one published by *Statistics Canada*, we will refer to this one as the CMHC index.

for Vancouver. As in the case of Toronto, we think that this index better captures the rent price dynamics in the overall market, including newly constructed buildings. As a result it matches our final house price index better.

Table 1.K.8. Final rent index for Vancouver

PERIOD	SOURCE	DESCRIPTION
1950-1984	Amaral et al. (2021)	*Type(s) of dwellings*: Single-family house, bungalow, apartment and duplex; *Type of data*: Asking rents from the *Vancouver Sun* and the *Vancouver Heights*; *Method*: Hedonic time-dummy adjacent period index.
1985-1987	Canada (2021a)	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Rents from the Labour Force Survey by Statistics Canada; *Method*: Match-model approach.
1988-2018	Canada (2021c)	*Type(s) of dwellings*: Apartments in buildings with six or more apartments; *Type of data*: Rents from the Canada Mortgage and Housing Corporation Survey; *Method*: Average rental value.

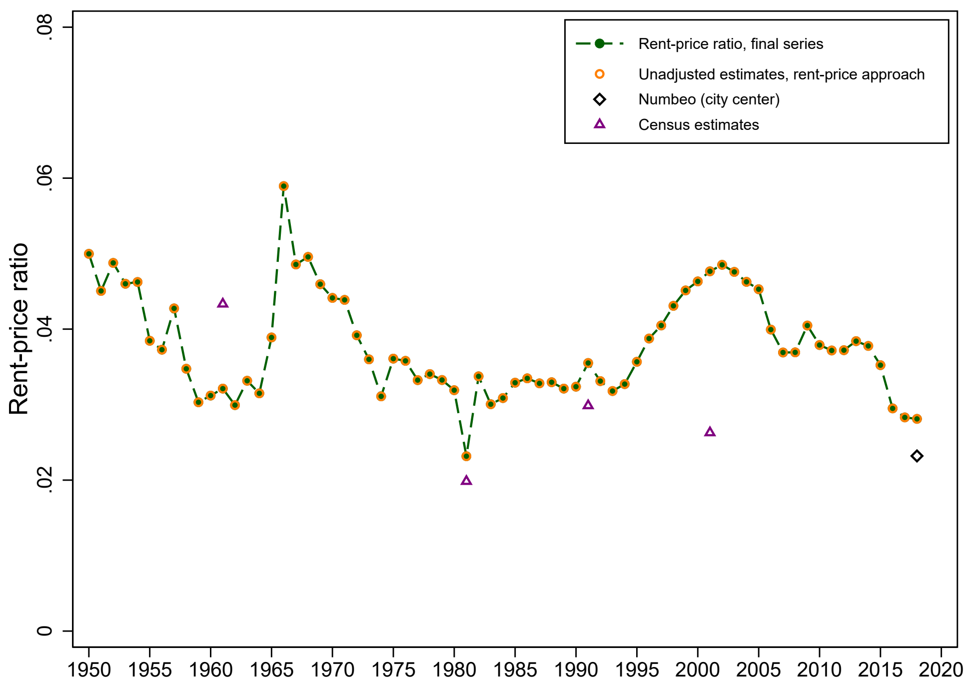

Figure 1.K.4. Vancouver: plausibility of rental yields

Rental Yield Series. Our main benchmark for Vancouver is taken from *MSCI*, as described in the main paper. This benchmark is slightly above the benchmark we collected for 2018 from *Numbeo.com*. According to *Numbeo.com* the gross rental yield in the city center of Vancouver was 3.48% in 2018, adjusting for one-third costs we estimate a net rental yield of 2.3% for 2018. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.4.

We use the same approach as in the case of Toronto to build historical rental yield benchmarks out of census data. Unfortunately, the household-level data from 1971 census does not contain data for Vancouver. As such, we were only able to construct benchmarks for the years 1981, 1991 and 2001. We use the exact method as the one described in the Toronto data appendix. Additionally, we also build a benchmark for 1961, using the median value of owner-occupied single-detached non-farm houses for the city of Vancouver and the average rent paid for renter-occupied non-farm dwellings for the city of Vancouver from the national housing census in 1961 (*1961 Census of Canada : Housing Vol. II* 1961). As such, the benchmark for 1961 is only a rough approximation to actual rental yield in Vancouver at the time. We then adjust the gross yield estimates for one-third costs and get the net yield estimates, which are depicted in Figure 1.K.4. While the census estimates for 1961 and 2001 lie slightly above and below our unadjusted rental yield series respectively, in general the census estimates match our series quite accurately. As a result we do not adjust our rental yield series.

1.K.3 Denmark

Copenhagen has been Denmark's largest city by far, with 462,000 inhabitants or eighteen percent of the total population in 1900 (followed by Aarhus, Odense, Aalborg and Vilborg with 51,000 or less). It is also the only city for which there are existing house price series and some available data on rent prices before 1970, where we had to construct new rent price series ever since.

1.K.3.1 Copenhagen

House Price Series. Prior to 1938, there is no existing house price index in Copenhagen (even the national-level data are based on farm prices). Since 1938, we draw on a a stratified Fisher index provided by Abildgren (2018) for the city of Copenhagen, based on transaction data from single-family houses by number of rooms. The strata are built using the housing stock for every five years. The source are official statistical publications from *Statistics Denmark*. From 1992 onward we use an index for single-family houses from Statistics Denmark (2021). The index is based on high-quality transaction-level data from the Ministry of Taxation which are collected weekly through an electronic land registration system. It uses the SPAR method and covers Greater Copenhagen. We use yearly averages of the quarterly index. The components of our final house price index are summarized in Table 1.K.9.

Table 1.K.9. Final house price index for Copenhagen

PERIOD	SOURCE	DESCRIPTION
1938-1992	Abildgren (2018)	*Type(s) of dwellings*: Single-family houses; *Type of data*: Transaction data; *Method*: Stratification index
1992-2018	Statistics Denmark and Tax Authorities	*Type(s) of dwellings*: Single-family houses; *Type of data*: Transaction level data from taxation records; *Method*: SPAR method.

Rent Series. Data for the rent series in Copenhagen come from a number of different sources. For the time period between 1885 and 1970, we draw on various years of the city yearbooks published by the *Copenhagen Statistical Office*. The yearbooks report every five years the annual rent for all rented dwellings in Copenhagen by number of rooms. Every five years we take a weighted average of the yearly average rent for dwellings with different numbers of rooms. The years in between are linearly interpolated.

From 1970 onward we constructed a new rent series. It is composed of different parts. From 1970 to 1990 we build a hedonic rent index using rental ads from a local newspaper, *Berlingske*, which publishes a small daily and a longer weekly section on real estate. We located and scanned the relevant pages from microfilm and extracted all ads containing at least information on rent, size and location. This produces a sample of 957 ads with complete information. The rental market sections mainly contained two segments, one on renting single-family houses also located in the Copenhagen region, one on rental apartments, predominantly in the city of Copenhagen. Our ads cover both market segments. We use a hedonic regression of the log rent levels on year dummies, controlling for a dummy for the market segments, city district dummies, a dummy for furnished units and housing size. We harmonized housing size by converting square meters into 0.023 rooms, based on our within-sample estimates. The explained variance is 0.852.

From 1990 to 2000 we use a weighted average of the house price indices published by the newspaper *The Economist*. The *Economist Intelligence Unit* publishes yearly data on living costs in global cities, containing information on rent levels by different room and quality bins. We average the data for all bins containing full information over time.

From 2000 onward, we accessed the archived real estate websites *boligportal*, *akutbolig* and *bolig-siden* and collected 728 rental ads on apartments for Greater Copenhagen containing full information on rents, size and location. We use a hedonic regression of the log rent levels on year dummies, including as control variables a dummy for the rental market segments, dummies for the city districts, a dummy for furnished units and housing size. We harmonized housing size by converting square meters into 0.023 rooms, based on our within-sample estimates. The explained variance is 0.721.

Table 1.K.10. Final rent index for Copenhagen

PERIOD	SOURCE	DESCRIPTION
1885-1970	City Yearbooks of Copenhagen (several issues)	*Type(s) of dwellings*: Apartments; *Type of data*: Municipal survey data; *Method*: Stratification index (interpolated).
1970-1990	Own series	*Type(s) of dwellings*: Single-family houses and apartments; *Type of data*: Newspaper ads from *Berlingske*; *Method*: Hedonic index.
1990-2000	The Economist	*Type(s) of dwellings*: Apartments with two to four bedrooms; *Type of data*: Estimates of the *Economist Intelligence Unit*; *Method*: Average over quality bins.
2000-2018	Own series	*Type(s) of dwellings*: Single-family houses and apartments; *Type of data*: Ads from real estate website *boligportal*; *Method*: Hedonic index.

Rental Yield Series. Our main benchmark for Copenhagen is taken from *MSCI*, as described in the main paper. Applying the rent-price approach to this benchmark results in the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.5.

We collected two additional benchmarks for 2018. First, we use the gross rental yield for the city-center of Copenhagen from *Numbeo.com*. We adjust this benchmark to capture the net rental yield by subtracting one-third following Jordà, Knoll, et al. (2019). Secondly, we collected a recent benchmark from *Collers Copenhagen*. Both benchmarks are somewhat above but reasonably close to our main benchmark from *MSCI*. We use the benchmark by *Numbeo.com* as our alternative benchmark in the robustness section of the main paper. We also collected additional rental yield benchmarks from *Numbeo.com* between 2010 and 2018. These benchmarks show a very similar pattern as our long-run net rental yield series and the value in 2010 is close to our long-run series.

To collect historical benchmarks, we draw on data from the 1950 publication of Statistics Denmark's *Vurderingen til Grundskyld og Ejendomsskyld*, an assessment of property values and rents for taxation purposes. The rental yields are computed as weighted average of the ratio of imputed rents to property values for residential real estate in Copenhagen, using the number of properties with different housing units as weights. The resulting benchmarks are clearly below our unadjusted series.

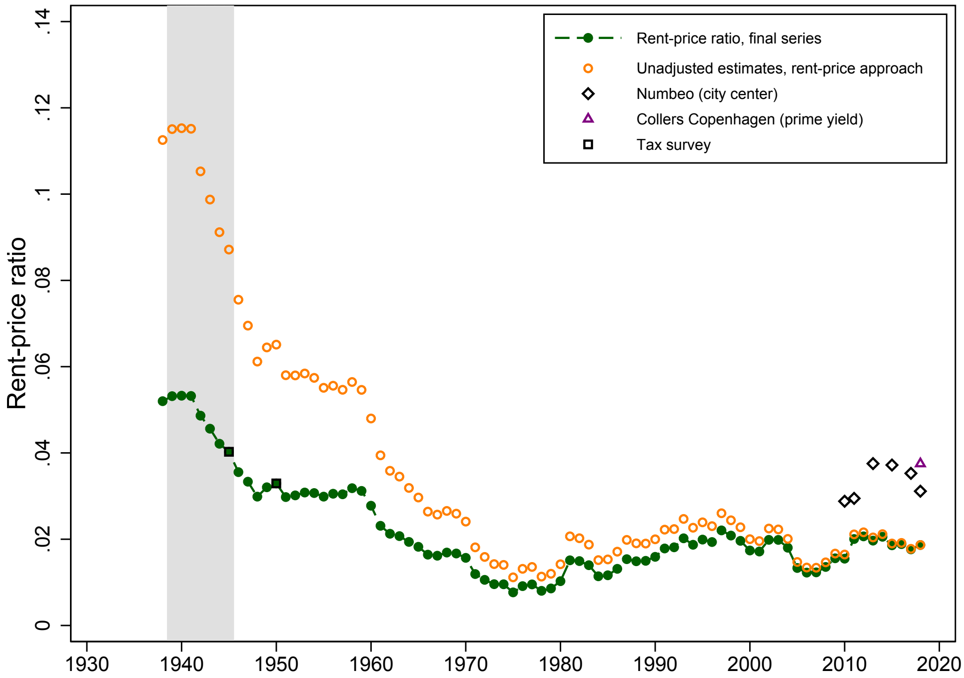

Figure 1.K.5. Copenhagen: plausibility of rental yields

As we have to rely on interpolation to construct our long-run rent series and the level of rent controls in Denmark were historically high, it could be the case that our long-run rental yield series are biased. Therefore we adjust our series to these benchmarks. The final rental yield series is plotted as the green-circled line in Figure 1.K.5.

1.K.4 Finland

In 1900 the three largest cities in Finland were in descending order: Helsinki, Tammerfors and Abo. All of these cities had more than 1% of the total national population in 1900. Unfortunately, we could only build long-run house and rent indices for Helsinki, since we could not find existing series or primary sources for the other two cities.

1.K.4.1 Helsinki

House Price Series. The long-run house price series for Helsinki covers the period between 1946 and 2018 and is composed of three different series. For the period between 1946 and 1970 we take the average price per square meter of dwellings in existing blocks of apartments in the center of Helsinki from *Statistics Finland*.[85] The transaction data were collected by *Statistics Finland* from different local real estate agencies.[86]

For the period between 1971 and 1988 we use a mix-adjusted hedonic house price index from *Statistics Finland* for the city of Helsinki. The index is based on transaction prices of dwellings in existing blocks of apartments and the data come from the *Finnish Tax Administration*. *Statistics Finland* uses housing stock weights to aggregate the hedonic indices for apartments with different numbers of rooms.[87]

For the period between 1988 and 2018 we use a mix-adjusted hedonic house price index from *Statistics Finland* for the city of Helsinki. The index is based on transaction prices of dwellings in existing terraced houses and existing blocks of apartments gathered by the *Finnish Tax Administration* for asset transfer tax calculation purposes, which are stratified by type of dwelling, number of rooms and location. The housing stocks of each strata are used as weights to build the final index.[88]

Statistics Finland also publishes a house price index of single-family houses for the Greater Helsinki area.[89] The index is a mix-adjusted hedonic index, where single-family houses are stratified by number of rooms and location. A hedonic regression is then applied to estimate the price index for each stratum. The strata are then combined using the housing stock as weights. The data on transaction prices come from the real estate register of the *National Land Survey of Finland* and they are combined with data from the real estate information system of the *Population Register Centre*.[90]

The index on single-family houses correlates very strongly with the index on existing dwellings.[91] Since the latter covers only the city of Helsinki, i.e. it excludes the the other cities in the Greater Helsinki area, we opted to choose it for our final series.

Table 1.K.11 summarizes the components of our final house price index.

Rent Series. Our long-run rent series for Helsinki also covers the period between 1946 and 2018 and is composed of two different series. To build the series we collected average rents of apartments in the city of Helsinki by number of rooms from the Helsinki statistical yearbook.[92] From 1946 to 1974 we build a stratified Fisher index using the stock of apartments by number of rooms as weights. On average, we

85. Existing blocks of apartments are buildings which were built at least two years before the publication of the series.
86. Statistics Finland, 2020a.
87. Statistics Finland, 2020a.
88. More information can be found in Statistics Finland (2020a).
89. Greater Helsinki includes the cities Helsinki, Espoo, Vantaa and Kauniainen.
90. Statistics Finland, 2020b.
91. Over the period between 1985 and 2018 the two indices have a correlation of 0.98.
92. *Statistical Yearbook of the City of Helsinki* various years.

Table 1.K.11. Final house price index for Helsinki

PERIOD	SOURCE	DESCRIPTION
1946-1970	Statistics Finland (2020a)	*Type(s) of dwellings*: Dwellings in existing blocks; *Type of data*: Average transaction price per square meter; *Method*: Average.
1970-1988	Statistics Finland (2020a)	*Type(s) of dwellings*: Dwellings in existing blocks; *Type of data*: Transaction prices; *Method*: Mix-adjusted hedonic price index.
1988-2018	Statistics Finland (2020a)	*Type(s) of dwellings*: Existing dwellings; *Type of data*: Transaction prices; *Method*: Mix-adjusted hedonic price index.

have the stock of apartments by number of rooms every five years. Since data are missing for the years 1948, 1949, 1951, 1952, 1953, 1955 and 1956 we interpolate the index for those years using the national rent index from Jordà, Knoll, et al. (2019). For the period after 1974 we construct a chained stratification Fisher index using the stock of apartments of the last available data point as weights.[93]

Table 1.K.12 summarizes the components of our final rent index.

Table 1.K.12. Final rent index for Helsinki

PERIOD	SOURCE	DESCRIPTION
1946-1975	Own compilation	*Type(s) of dwellings*: Apartments ; *Type of data*: Average rents; *Method*: Fisher stratified index using stocks as weights.
1975-2018	Own compilation	*Type(s) of dwellings*: Apartments ; *Type of data*: Average rents; *Method*: Chained fisher stratified index using stocks as weights.

Rental Yield Series. Unfortunately, *MSCI* does not provide an estimate for Helsinki. As a result we benchmark our series to the estimate from *Numbeo.com* for 2018 for Helsinki city center. According to *Numbeo.com* the gross rental yield was 3.06%, to which we subtract one-third costs. According to *KTI*, a Finnish property markets analysis firm, the gross yield on residential property in the city center of Helsinki was 3.7%.[94] Applying again one-third costs, we get a net-yield benchmark of 2.4% for 2018, which is very close to the estimate from *Numbeo.com*. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.6.

Unfortunately, we were not able to find any historical benchmarks for Helsinki. As a result our final adjusted rental yield series — the green-circled line in Figure

93. From 2005 onward data on the stock of apartments are available every year.
94. *The Finnish Property Market, 2019* 2019.

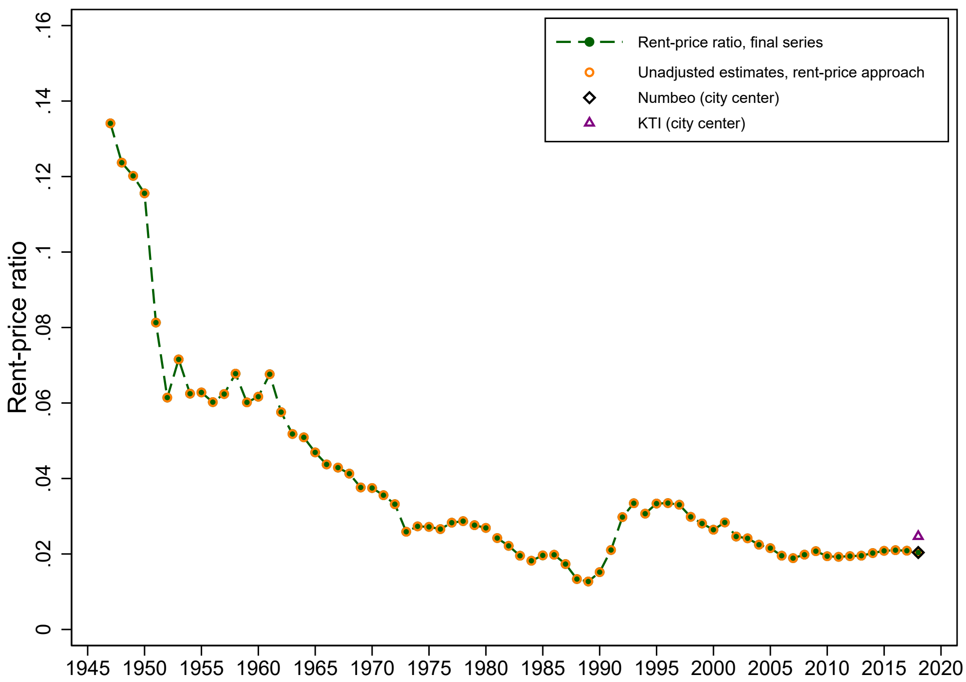

Figure 1.K.6. Helsinki: plausibility of rental yields

1.K.6 - matches the unadjusted series, which features very high values before 1950. The considerable drop around 1950 might be driven by the relaxation of wartime rent controls during this period. We can, however, not exclude that our long-run rental yield series is biased for the earlier period due to the fact that our house price index is based on a simple average and, thus, does not control for sample shifts and quality changes over time. This problem might be especially relevant in the direct aftermath of World War II, when the local housing market was still recovering from the damages of the war. Since our main analysis only starts in 1950, this bias is probably less relevant for our main results.

1.K.5 France

In 1900 the four largest cities in France were in descending order: Paris, Lyon, Marseille and Bordeaux. With the exception of Bordeaux, all of these represented more than 1% of the national population in 1900.[95]

To the best of our knowledge there do not exist continuous long-run series for Marseille and Lyon. As such, we only produced a long-run series for Paris. However, Bonneval and Robert (2010) build a housing return series for Lyon for the period

95. Reba, Reitsma, and Seto, 2016.

between 1890 and 1968 based on archival data from a local real estate company. Future work will, hopefully, complement this series to builds a representative housing return series of the city of Lyon. The scarcity of historical housing series for other French cities stands in contrast to the abundance of sources and work on the Parisian housing market. Nevertheless, in the last few decades, there has been a significant effort to extend the quality of available information on the housing market to the rest of France. As explained in more detail below, the *National Institute of Statistics and Economic Studies* (INSEE) has united forces with the notaries association in France to put to use the very extensive notaries' data set.

1.K.5.1 Paris

House Price Series. The *Conseil General de l'Environnement et du Developpement Durable* (CGEDD)[96] publishes a price index for residential property in the Greater Paris area. The index starts in 1200 and builds upon several different data sources. For the time period analyzed in this paper (1870-2018), the index is composed of three different series. The first part of the index (1870-1944) is based on a repeat sales index by Duon (1946) using data gathered from property registers of the local tax department. It covers apartment buildings (*maisons de rapport*) such that commercial properties, single-family houses, or apartments sold by the unit remain excluded.[97] The second part of the index (1944-1999) is based on price data for apartments sold by the unit compiled by CGEDD from the notaries' database and calculated using the repeat sales method.[98] Since data for the period before 1950 are very scarce, the gap between 1945 and 1949 was filled by linearly interpolating the data from Duon (1946) and the data for the period after 1950 (Friggit, 2002). For the post-1999 period, the index is again spliced with an index by the National Institute of Statistics and Economic Studies (INSEE) for existing apartments in Paris (Clarenc, Cote, David, Friggit, Gallot, et al., 2014).[99] This index is built using transaction data from the notaries' database and a mix-adjustment hedonic approach, i.e. a hedonic index is built for various subparts of Paris, which is then aggregated using stock weights. The hedonic indices include a rich set of apartment characteristics, which include a very precise location of the apartments as well as their sizes.[100]

As mentioned above the Paris index for the period between 1944 and 1999 is based on data from the notaries' database in France, which only actually started

96. Conseil General de l'Environment et du Developpement Durable, 2021.

97. Until World War I single apartments could not be sold alone.

98. In France there exist two main real estate transaction databases from the notaries: The *BIEN base*, which is mostly focused on the Paris region and is managed by *Chambre Interdépartmentale des Notaires de Paris*, and the *PERVAL France*, which covers the rest of France and is managed by the *Conseil Supérieur du Notariat*.

99. The index only focuses on existing apartments, i.e. new apartments are excluded from the sample.

100. For more details about the method please consult Clarenc et al. (2014).

recording transactions in 1990. However, when a transaction takes place the current price as well as the price of the previous sale are recorded in the database. This poses two problems. First, as acknowledged in Friggit (2008), the number of observations for the first years of the index is quite low, making the index more sensitive to outliers. Second, we know that the observations in the 1950s represent cases of extremely long holding periods (at least 35 years), which also introduces a strong sampling bias. Usually very long holding periods (over 30 years) are excluded from repeat sales samples, because these transactions are atypical and might signal that the properties are of poor quality or have lower than average demand (Eurostat, 2013).

The CGEDD also publishes a national house price index for France, which starts in 1936. For the 1950s the Paris index grows by more than double the national index.[101] As argued in Friggit (2008), we would expect prices in Paris to outpace the rest of the country as a result of the end of the rent control laws, which had disproportionately affected Paris. However, the difference in price growth rates still seems to be too large and, given the concerns about the quality of the Paris index in this period, we decided to construct our own house price index using an independent data source.

For the period between 1950 and 1958 we collected residential apartment sales ads from the newspaper *Le Figaro* and used them to construct a new house price series. In total we collected 1,595 observations, which contained information on asking price, size of the apartment, the *arrondissement* in which the apartment was located and further characteristics of the apartment, which are described in more detail below. We then used these data to build a hedonic house price index for Paris using the time dummy approach.[102] More precisely, we use the following model:

$$ln(p_{i,t}) \; = \; \alpha_0 + \delta_1 D_t + \sum_{k=1}^{21} \beta_{arrond_k} + \sum_{j=1}^{12} \beta_{rooms_j} + \alpha_{new} + \alpha_{bathroom} + \alpha_{kitchen} + u_i$$

where we regress the log asking price at time t of apartment i on a series of dummy variables indicating: the year in which the ad was posted (D_t), the *arrondissement* k, in which the apartment is located (β_{arrond_k}),[103] the number of rooms j in the apartment (β_{rooms_j}),[104] whether the apartment was new or had recently been fully

101. The national index is also built using the same method as the Paris index. However, the national nature of the index means that it is built based on a larger set of observations for this period and is therefore less prone to outliers.

102. This approach has been frequently been used in the literature to build house price indices using newspaper ads (see e.g. Lyons et al (2019).

103. Since we also collected data for the suburbs Levallois and Neuilly, we added them as the 21st *arrondissement*. The results almost do not change if we do not include them.

104. Here we considered number of rooms to be the sum of bedrooms (*chambres*), living rooms (*living*), dining rooms (*salle a manger*) and receptions (*entrée*).

renovated α_{new}, and whether the ad indicated it contained a bathroom ($\alpha_{bathroom}$) or a kitchen ($\alpha_{kitchen}$). We estimated this regression using OLS by pooling all years together, the "all-in-one" approach, and by using a rolling three-year window. From Figure 1.K.7 it becomes clear that both series are strongly correlated. In the end, we decided to use the "all-in-one" series, since it uses the largest number of observations for the regression and has, therefore, more precisely estimated coefficients. The regression results show that we explain around 73% of the variation in log prices with the set of independent variables detailed above.

In Figure 1.K.7 we compare our Paris index with the repeat sales indices for Paris from Friggit (2002). While the index from Friggit (2002) grows by more than a factor of 9, our index grows by a factor of approximately 3. As mentioned above, the small number of observations makes the index from CGEDD very sensitive to outliers for this period.[105] Therefore, we think that our series is more trustworthy for this period and decided to use it in our long-run series.

In Table 1.K.14 we summarize the series we used to build our long-run index. As mentioned above, the biggest difference with respect to the long-run series from CGEDD is that we use our own house price series for the period between 1950 and 1958, which we chain-linked to the series from CGEDD.

Rent Series. For 1870–1945, we use a rent index for Paris constructed by Marnata (1961). The index is based on a sample of 11,800 different rent contracts. Data come from lease management books from residential neighborhoods in Paris and mostly refer to dwellings of relatively high quality or more expensive housing.

For the years prior to 1949, data on rents are also available for Paris (1914–1962) from the yearbooks of the International Labour Organization (various years). As shown in Knoll (2017), the series by Marnata (1961) and the series published by the International Labour Organization (various years) are highly correlated for the years they overlap.

After 1946, we rely on the rent component of the CPI for Paris from various editions of the *Annuaire Statistique de la France* published by the INSEE and which were assembled by Jaques Friggit.[106] The index covers tenants' rents only, i.e. imputed rents of owner-occupiers are excluded. After 1989 we use the mean rent per square meter of apartments in the city of Paris, considering only the estimates for the inner city, i.e. the *arrondissements*, from the *Observatoire des Loyers de l'Agglomeration Parisienne* (OLAP). This series was also assembled by Jaques Friggit and the data can be found in the yearly reports of OLAP.[107]

105. The notaries' database is not generally available to researchers and, as such, we did not have access to the data. Therefore, we could not test whether our data concerns could actually be biasing the series from CGEDD.

106. More precisely, the component *loyers et charges* of the CPI was always published separately.

107. Observatoire des Loyers de l'Agglomeration Parisienne, various years.

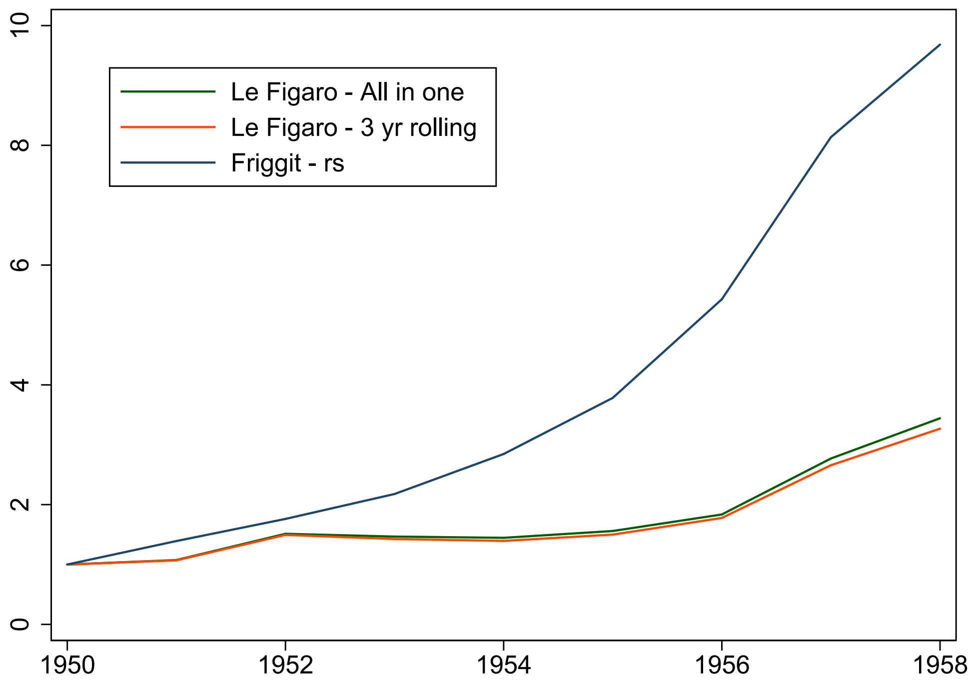

Figure 1.K.7. Nominal house price indices for Paris, 1950=1

For the period after 1989, the INSEE has also been publishing a rent index for the region of Paris. As shown in Friggit (2013), the OLAP and the INSEE series differ in this period, with the OLAP series growing slightly more than the one from INSEE. Since our house price series focuses on the city of Paris, i.e. it excludes the outskirts, we have decided to use the series from OLAP.

Rental Yield Series. Our main benchmark for Paris is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to the alternative benchmark we collected for 2018 from *Numbeo.com*. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.8.

As can be seen in Figure 1.K.8 our final series diverges substantially from the unadjusted series in the 1950s. As was mentioned above, in the 1950s the rent controls were gradually abolished in Paris. The rent CPI component that we are using for this period probably does not fully capture the effect of new rentals in the market, which were not capped by the law. As a result our rent series grows substantially less than the price series, which includes almost exclusively apartments, which were no longer affected by the rent freeze law of 1948. To correct the resulting bias on our rental yield series we collected several different historical benchmarks, which we describe below.

Table 1.K.13. Final house price index for Paris

PERIOD	SOURCE	DESCRIPTION
1870-1949	Conseil General de l'Environment et du Developpement Durable (2021)	*Type(s) of dwellings*: Apartment buildings; *Type of data*: Transaction prices from tax department; *Method*: Repeat sales method.
1950-1958	Own compilation	*Type(s) of dwellings*: Apartments; *Type of data*: Asking prices from *Le Figaro*; *Method*: Hedonic time-dummy index.
1959-1999	Conseil General de l'Environment et du Developpement Durable (2021)	*Type(s) of dwellings*: Apartments; *Type of data*: Transaction prices from the notaries' database; *Method*: Repeat sales.
2000 - 2018	Conseil General de l'Environment et du Developpement Durable (2021)	*Type(s) of dwellings*: Existing apartments; *Type of data*: Data from the notaries' database; *Method*: Mix-adjusted hedonic house price index.

Table 1.K.14. Final rent price index for Paris

PERIOD	SOURCE	DESCRIPTION
1870-1945	Marnata (1961)	*Type(s) of dwellings*: Apartments; *Type of data*: Rental data from lease management books; *Method*: Chain index of repeat rents.
1946-1988	INSEE (various years)	*Type(s) of dwellings*: All types of residential dwellings; *Type of data*: Rental data from survey; *Method*: Average based on repeated rental contracts.
1989-2018	Observatoire des Loyers de l'Agglomeration Parisienne (various years)	*Type(s) of dwellings*: Apartments; *Type of data*: Rental data from survey *Method*: Average rent per square meter.

For the year 1894 Saint-Genix et al. (1895) estimates that the total value of rents is 819 million francs while the market value of the housing stock is 13000

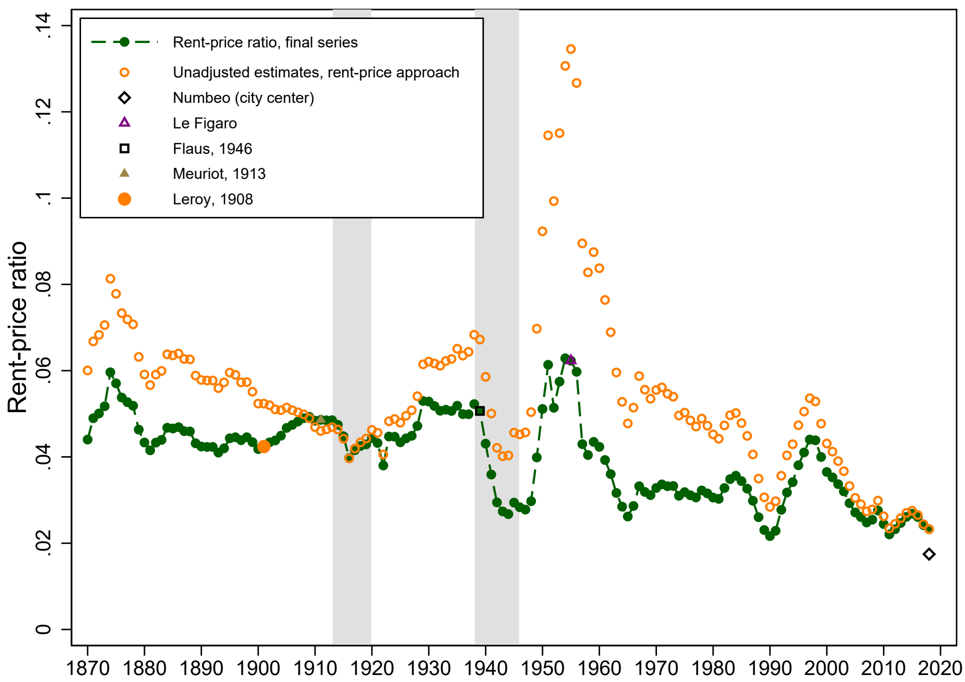

Figure 1.K.8. Paris: plausibility of rental yields

million francs, which produces a gross rental yield of 6.3%. Since the author does not provide many details about how the values were estimated we do not use this benchmark. For the year of 1901 Leroy-Beaulieu (1908), using data on average residential rent and house prices from *Le Livre Foncier de Paris, 1902*, estimates an average gross rental yield of 6.36%, average maintenance costs and taxes of 36.5%, and which gives an average net rental yield of 4.04%. For the year 1911 Meuriot (1913), using data on average residential rent and house price from *Le Livre Foncier de Paris, 1911*, estimates an average gross rental yield of 6.58%, and average maintenance costs and taxes of 26.1%, which gives an average net rental yield of 4.86%. Simonnet, Gallais-Hamonno, and Arbulu (1998) estimate that the gross rent of residential properties purchased by the property investment fund La Fourmi Immobiliere in Paris represented about 6 to 7 percent of property value between 1899 and 1913, which corroborates the estimates by Leroy-Beaulieu (1908) and Meuriot (1913). For the year 1939 Flaus (1946) estimates an average gross rental yield of 7.6% for the city of Paris. For the year 1955 we collected 28 ads from *Le Figaro* for apartment buildings (*maisons de rapport*) in the center of Paris, which reported both the asking price and the gross rental income of the building. By assuming total costs of one third of the gross rental income we arrive at an estimate of 6.2% net rental yield.

Adjusting our rental yield series to the historical benchmarks gives us the adjusted final rental yield series—the green-circled line in Figure 1.K.8.

1.K.6 Germany

The list of the largest cities in Germany in 1900 is in descending order: Berlin, Hamburg, Dresden, Leipzig, Munich, Cologne, Wroclaw and Frankfurt.[108] Of these cities, only Berlin and Hamburg hit the 1% target. The area of Germany, however, changed drastically several times after 1900. This means that we do not include Wroclaw, which does not belong to Germany nowadays or Leipzig and Dresden, which were part of Eastern Germany between 1945 and 1990 and hence price and rent data are missing for a considerable time period. From the remaining cities, there does not exist sufficient data coverage for Munich. To still get close to the 10% target and as Germany covered a considerably larger area in 1900 compared to today, we chose to include all other cities up to and including Frankfurt in our sample. In 1950, the population in both Frankfurt and Cologne was above 1% of Germany's total population.

To the best of our knowledge, there do not exist compiled house price or rent indices on city-level in Germany from public sources. The only readily-available city-level indices we know of are from private companies like *Bulwiengesa*. Recently, researchers started to rely on asking price data from online marketplaces like *ImmoScout24* to analyze German housing markets on a local level. An impressive example is Ahlfeldt, Heblich, and Seidel (2021), who use these data to compile house price and rent indices on arbitrary local levels in Germany between 2007 and 2018. By nature, these data, however, only cover the last one or two decades. As described in the main paper, we construct a novel city-level house price and rent data set for Germany using market reports of the German Real Estate Association (IVD) and one of its predecessors.[109] These market reports surveyed local real estate agents and collected city-level observations for various market and quality segments. We will partly rely on these data to construct our German city long-run series.

There is, however, an alternative source for house price data in Germany. Since 1960, notaries in Germany are obliged to report purchase details for every real estate transaction to the so-called *Gutachterausschüsse* (GA). The GA are comprised of real estate professionals and are organized on city-level. The GA store the transaction price information, along with house characteristics, and compile annual statistics on transaction volumes and price trends that are used to calculate benchmark land prices (*Bodenrichtwerte*) and form the basis for the assessment of real estate values

108. City-level population data are taken from Reba, Reitsma, and Seto (2016) and country-level population data from Jordà, Schularick, and Taylor (2017).

109. The *Immobilienverband Deutschland (IVD)* and the predecessor *Ring deutscher Makler (RDM)*.

for bank loans and insurance purposes. The underlying micro-data in the archives of the GA cover the universe of real estate transactions in (West-)Germany over the past 60 years. So far, this micro-data has not been digitized for academic research. Recently, we started a project that aims to do this. In the course of this project, we are cooperating with the GA in Berlin, Frankfurt and Cologne. These kindly enabled us to use parts of their data to construct hedonic house price indices, which we used to construct our city-level long-run series. We will describe these indices in more detail below.

1.K.6.1 Berlin

House Price Series. For Berlin, we are able to use micro-level house price data from the GA Berlin from 1965 onward. For the period 1958 to 1965 the GA Berlin published data on average price per transaction. Before this period, we had to rely on various statistical publications. Using these we build stratification indices by housing type or district whenever possible. We provide more details below. Due to the special history of Berlin, we only use data on West Berlin after World War II. In the period when Germany was divided only data for West Berlin are available. Afterwards, house prices in East Berlin followed a considerably different trend compared to house prices in West Berlin, mainly because they started at a considerably lower level in 1990. It took a long period for prices in East and West Berlin to converge and conversion might not even be completed today. To not confound our long-run indices with these conversion effects we only use data on West Berlin also after 1990. Prior to the separation of Berlin, the city developed similarly to other cities and there was no fundamental difference between East and West Berlin. As the separation was unforeseen prior to and during World War II, we rely on the data for all of Berlin prior to 1945.

The GA Berlin kindly provided us with transaction-level data for single- and multi-family houses. These data cover the universe of all normal housing transactions in (West) Berlin that were considered to be market prices.[110] We construct separate hedonic indices by housing type (single- or multi-family housing) and eight districts ("*Bezirk*") in West Berlin.[111] We were able to rely on 28,710 transactions for multi-family houses and 63,416 transactions for single-family houses.

To calculate the indices we closely follow the methodology in Eurostat (2013). The final index is constructed in three steps: First, double imputation hedonic indices are calculated separately for single-family and multi-family houses and separated by

110. The GA cleaned the data for non-market price transactions, for example transactions between family members. We also cleaned the data for all cases when not the whole building was sold.

111. We reconstructed the parts from *Bezirk* 1 and 2 that belonged to West Berlin as discrete districts. For multi-family houses the index for district 5 only starts in 1975, as we did not have enough observations before. For single-family houses we were not able to construct indices for districts 1 and 2, as there have not been enough transactions and the index for district 7 only starts in 1970.

district. Chaining is used to connect different years. All hedonic regressions use lot area, state of repair and type of building as exogenous variables. Regressions from 1980 onward also control for floor area[112] interacted with state of repair as a proxy for the quality-adjusted structure. Regressions for multi-family houses from 1980 onward do not use type of building as control anymore,[113] but instead a dummy variable indicating if a part of the building is in commercial use. All regressions are performed using maximum likelihood and assuming normally distributed standard errors. Second, for both housing types separately a Fisher-type stratification index is built from the district-level indices using chaining and transaction value shares as weights. Lastly, the final index is obtained by stratification of the respective indices for single-family and multi-family houses. We again use transaction value shares as weights.

For the period from 1870 to 1918 and 1936 to 1938 we use data from various volumes of the statistical yearbooks of Statistics Berlin[114] as done in Knoll, Schularick, and Steger (2017). The yearbooks contain aggregated data on number of sales and sales volumes for all sales of developed land. In contrast to Knoll, Schularick, and Steger (2017), we use the data on 16 to 21 districts ("*Stadtteile*") for the period from 1870 to 1906 and 1936 to 1938. With this more fine-grained data we build Fisher-type chained stratification indices following Eurostat (2013) using mean price per sales within each stratum. In this way we are able to control for locational shifts in real estate sales. To the extent that building types and locations are correlated, this approach indirectly also controls for shifts in the mix of building types reducing sample selection bias further. For the period from 1907 to 1918 we had to rely on the average price per sale in the former city of Berlin,[115] as data by district were no longer available. We match the index from 1918 to 1936 using the average sales price for the former city of Berlin using the borders of 1918.

We imputed the years in between from two sources: We rely on estimates for the price of rental apartment buildings in Prauser (1941) for the period between 1923 and 1935. For the years 1935 to 1936 we collected the average price per square meter of developed land in the city of Berlin from the statistical yearbooks of German cities.[116] For the period directly after World War I and the period of German hyperinflation no data were available. For the years 1938 to 1940 we collect the average price per square meter of developed land from another publication of Statistics Berlin.[117]

112. Before there have been too many missing observations.
113. The reasons are multicollinearity with floor area and too many degrees of freedom relative to the number of observations.
114. "Statistisches Jahrbuch der Stadt Berlin".
115. Which is nowadays the city-center of Berlin.
116. See Knoll, Schularick, and Steger (2017) data appendix.
117. "*Berlin in Zahlen*" (Volume 1942, p. 86).

During World War II no house price data are available for Berlin. The earliest data after World War II we found start in 1953. From 1953 to 1955 we again rely on a publication by Statistics Berlin.[118] The data cover number of transactions and transaction volume in 20 districts in West Berlin.[119] We again build a Fisher-type stratification index as describes above. For the years 1955 to 1958 we use aggregate data from Pistor (1957) and Pistor (1960). We use the data on sales per housing category (single/multi-family houses, which are not rental apartment buildings, and rental apartment buildings) to again build a Fisher-type stratification index. We also use the data on sales per district for 1958 from the same source to link 1958 to 1938 using stratification per district for all districts belonging to West Berlin.

To fill the remaining gap between 1958 and 1965 we use data from Berger (2010). Again we have to rely on aggregates for all of West Berlin. From 1958 to 1960 we use the average price per transaction for all developed lots without destroyed buildings ("*Trümmergrundstücke*"). For the period from 1960 to 1965 we instead rely on data on price per sold apartment, because apartments are a much more homogeneous good compared to developed lots and the resulting index is, therefore, less volatile. Table 1.K.15 summarizes the components of our final house price index.

Rent Series. The rent index for Berlin is constructed using multiple data sources. Starting in 1975, we use the data we constructed from market reports of the German Real Estate Association and its predecessor. Before, we rely on a rent index calculated during the process of constructing a city-level CPI index. Prior to World War II, we additionally use legal rents in years with strict rent freezes. Prior to and during World War I, we instead rely on rent data collected for tax reasons and published by Statistics Berlin. We provide more details below.

The construction of rent indices from the real estate market reports is described in the Appendix of the main paper. We use the index for Berlin from 1975 onward. This index covers only West Berlin until 2013. From 2014 onward, only data for all of Berlin are available. Rents in West and East Berlin had, however, already come very close in 2013, such that the trends can be assumed to be the same in West and East Berlin after 2013.[120]

For the period from 1950 to 1975 and 1934 to 1938 we use rent indices constructed for the city-level CPI calculation for Berlin from statistical yearbooks pub-

118. "*Berliner Statistik*" (Volume 1958, p. 118).

119. In contrast to the data used for the earlier periods, these data do not only cover sales, but all kinds of real estate transactions. This probably increases measurement error. For this reason we only use these data source for two years, when no other data was available, and make the link from before to after World War II using the data for 1958.

120. We construct the index by chaining and compare the aggregate data in 2014 with the aggregates from West and East Berlin in 2013, such that the sample is always the same comparing consecutive years.

Table 1.K.15. Final house price index for Berlin

PERIOD	SOURCE	DESCRIPTION
1870 - 1906	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated by district from yearbooks; *Method*: Stratification.
1907-1918	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated for former city from yearbooks; *Method*: Average price per transaction.
1923-1935	Prauser (1941)	*Type(s) of dwellings*: Rental apartment buildings; *Type of data*: Price estimates; *Method*: Price per transaction.
1935-1936	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated for former city from yearbooks; *Method*: Average price per square meter of developed land.
1936-1938	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated by district from yearbooks; *Method*: Stratification.
1938-1940	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated for former city from yearbooks; *Method*: Average price per square meter of developed land.
1953-1955	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All transactions aggregated by district from yearbooks; *Method*: Stratification.
1955-1958	Own compilation	*Type(s) of dwellings*: Single/multi-family houses and rental apartment buildings; *Type of data*: All sales aggregated by two housing categories; *Method*: Stratification.
1958-1960	Berger (2010)	*Type(s) of dwellings*: All developed lots except destroyed ones; *Type of data*: All sales aggregated for West Berlin; *Method*: Average price per transaction.
1960-1965	Berger (2010)	*Type(s) of dwellings*: All sold apartments; *Type of data*: All sales aggregated for West Berlin; *Method*: Average price per transaction.
1965-2018	Own compilation	*Type(s) of dwellings*: Universe of single-family and multi-family houses; *Type of data*: Transaction-level data kindly provided by the Gutachterausschuss Berlin; *Method*: Stratified hedonic index.

lished by Statistics Berlin.[121] For the period from 1950 onward the index is intended to track flats for a four-person blue-collar worker household (excluding heating and

121. "*Statistisches Jahrbuch Berlin*" (Volume 1952-1976) and "*Statistisches Jahrbuch der Stadt Berlin*" (Volume 1935-1939).

other costs) and covers West Berlin.[122] For the period prior to World War II, the index covers only old flats (also excluding heating and other costs) for which rent controls applied. The rent index in 1951 is given in 1938 values, such that the linking of the two periods is straightforward.

For the period between 1924 and 1933 we directly rely on legal rents from statistical yearbooks.[123] During this time period, legal authorities dictated a strict rent ceiling in terms of 1914 rents ("*Friedensmiete*"). As this rent ceiling was typically set very low and housing was scarce during this time period, legal rents can be assumed to have been binding in large cities like Berlin. Moreover, for the period from 1934 to 1938, the CPI rent index and legal rents show exactly the same patterns. One drawback is that new construction was excluded from the rent ceiling. There are, however, no other rent data available for this period of time. Both legal rents as well as the CPI rent index for 1934 to 1938 are given in 1914 values, such that linking is straightforward. For the period from 1918 to 1923 no reliable data exist.

For the period from 1870 to 1917 we rely on average rents of the universe of all rented units collected for tax reasons and published in statistical yearbooks.[124] For the period from 1870 to 1907, averages are given by 21 districts ("*Stadtteile*"). As done for house prices, we use these disaggregated data to build Fisher-type chained stratification indices following Eurostat (2013). This way, we are able to control for locational shifts in the sample of rented units. For the period between 1908 to 1917, only averages for the entire city are given, such that we cannot control for locational shifts. For the period before, however, the index controlling and not controlling for locational shifts are similar. One disadvantage of this data source is that it also contains commercial rooms. Therefore, we additionally rely on data from housing censuses between 1880 and 1905, which was collected in five year steps, published in Reich (1912). These data covers the universe of all rented residential apartments in Berlin and contain average rent by number of heated rooms. We construct rent increases in five year steps using average rent weighted by the number of apartments with the respective number of rooms from the 1880 census.[125] We use these data to adjust rent increases between 1880 and 1905 and use the rent index from tax data only for interpolation and extrapolation. For the period for which both sources exist, however, five-year increases from both sources are very similar.

Table 1.K.16 summarizes the components of our final rent index.

Rental Yield Series. Our main benchmark for Berlin is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to all alternative

122. The method or sample to construct the index did, however, change in between, such that different publications give slightly different results for the years 1962 to 1964. We rely on the index given in 1958=100 until 1964.

123. "*Statistisches Jahrbuch der Stadt Berlin*" (Volume 1925-1935).

124. "*Statistisches Jahrbuch der Stadt Berlin*" (Volumes 1877-1917).

125. From "*Statistisches Jahrbuch der Stadt Berlin*" (Volume 1883).

Table 1.K.16. Final rent index for Berlin

PERIOD	SOURCE	DESCRIPTION
1870-1907	Own compilation	*Type(s) of dwellings*: All residential apartments and commercial rooms for rent; *Type of data*: Averages by district collected for tax purposes from yearbooks; *Method*: Stratification (only used for interpolation and extrapolation).
1880-1905	Own compilation	*Type(s) of dwellings*: All residential apartments for rent; *Type of data*: Averages by number of rooms from census from yearbooks; *Method*: Average weighted by number of flats in 1880.
1908-1917	Own compilation	*Type(s) of dwellings*: All residential apartments and commercial rooms for rent; *Type of data*: Averages for entire city collected for tax purposes from yearbooks; *Method*: Average rent.
1924-1933	Statistics Berlin	*Type(s) of dwellings*: Apartments in old buildings subject to rent control; *Type of data*: Legal rent; *Method*: Legal rent in 1914 values.
1934-1938	Statistics Berlin	*Type(s) of dwellings*: Apartments in old buildings subject to rent control; *Type of data*: From CPI-construction; *Method*: CPI rent index.
1950-1975	Statistics Berlin	*Type(s) of dwellings*: Apartments for four-person blue-collar worker household; *Type of data*: From CPI-construction; *Method*: CPI rent index.
1975-2018	Own compilation	*Type(s) of dwellings*: Apartments; *Type of data*: New contract rents by construction period and housing quality reported by local real estate agents; *Method*: Matched model approach using mode prices by quality and construction period bins.

benchmarks we collected for 2018, especially the ones by *Numbeo.com* and the GA (see below). Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.9.

To verify the historical rental yield series, we collected alternative benchmarks from three different sources. First, we use benchmarks from the market reports of the *German Real Estate Association* (IVD). These market reports directly report mode price-rent ratios for residential investment buildings from 1989 onward. The Appendix in the main paper describes how we calculate net rental yields given these. The resulting rental yield benchmarks are plotted as purple triangles in Figure 1.K.9. This series is somewhat above our long-run series, but shows similar (cyclical) patterns. Moreover, the earliest values from the IVD series are very close to our long-run series.

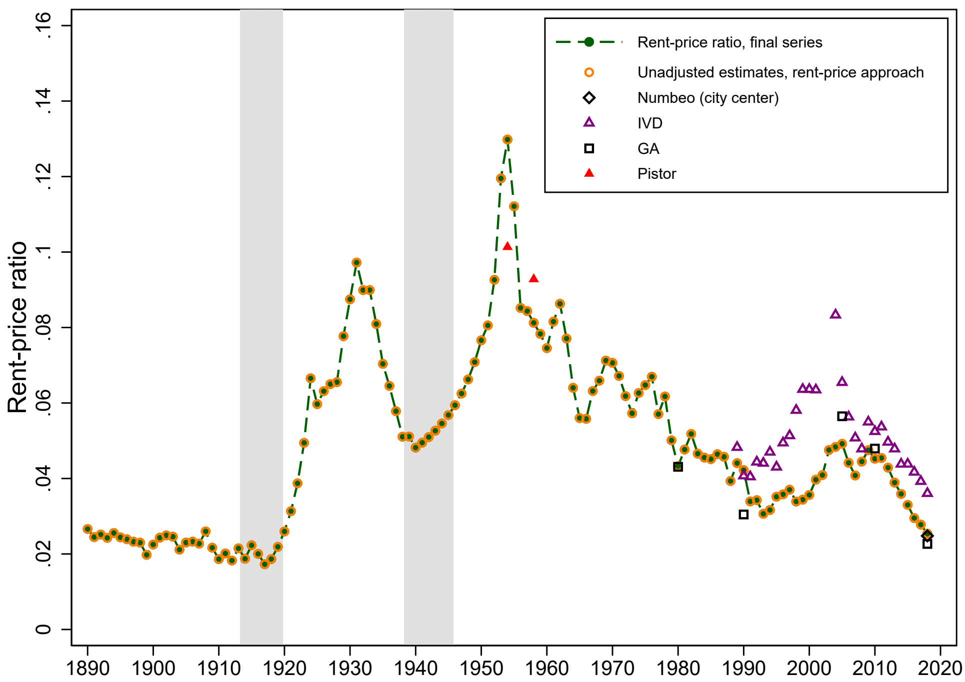

Figure 1.K.9. Berlin: plausibility of rental yields

Second, in their yearly market reports, the GA publishes price-rent ratios for investment buildings. [126] For the years 1980 and 1990, rent values still contain running costs, which needed to be paid by the renter.[127] We assume 20% of rents are running costs to calculate gross rental yields.[128] Following Jordà, Knoll, et al. (2019) we subtract one third of these to get net rental yields. The resulting benchmarks are depicted as black squares in Figure 1.K.9. These benchmarks are very close to our long-run rental yield series, especially the earliest benchmark in 1980. We use the 2018 benchmark from the GA as our alternative benchmark in the robustness section of the main paper, because it has the broadest coverage and relies on micro-data of actual transaction prices. This benchmark is very close to the benchmark from Numbeo.com used as an alternative benchmark for some of the other cities.

126. For 2010 and 2018, we take the average of all investment buildings. For earlier years, we take the average of the price-rent ratio ranges for buildings constructed after 1924. For earlier construction periods, large investments into maintenance might have been necessary. Therefore, we exclude these.

127. For example costs for water or waste disposal ("*Bruttokaltmiete*").

128. We use the micro-level data from the GA to estimate the relation between rent-price ratios with and without running costs for the years between 1995 and 2000. For these years, running costs are between 20 and 25% of rents. As a conservative estimate, we assume that 20% of rents were running costs before 1995.

Third, Pistor (1955) and Pistor (1960) state rent-price ratios for investment buildings in 1954 and 1958. Again, running costs are still included in rent payments and we assume that these have been 20% of rents. The resulting net rental yields are plotted as red triangles in Figure 1.K.9. Both values are reasonably close to our long-run rental yield series.

To summarize, all benchmark values we collected are close to our original rental yield series. Moreover, some benchmarks are above this series and others below. As rental yield benchmarks are subject to measurement error by themselves, we do not adjust our rental yield series for Berlin to any benchmarks.

1.K.6.2 Cologne

House Price Series. The GA in Cologne kindly provided us with micro-level house price data starting in 1989 for all housing types and starting in 1981 for apartments. We use these data to build hedonic house price indices. For the period from 1966 to 1981, we were able to use a subsample of micro-level house transactions from archived data from the GA, which we use to construct a repeat sales index. For the period before 1966, we had to rely on average price per transaction from statistical publications as no other data were available to us. Before World War II, statistical publications contain the average price by district, such that we are able to build stratification indices. We provide more details below.

To construct the hedonic index starting in 1989 we use transaction-level data provided by the GA Cologne. These data cover the universe of transactions for all single-family and multi-family houses as well as apartments that can be assumed to have sold for market prices.[129] We use cleaned transaction prices provided by the GA. These values adjust for sales conditions unrelated to the property itself, for example if the inventory or a kitchen was sold with the house or if specific rights were still granted to the seller. We construct separate hedonic indices by housing type (single-family housing, multi-family housing or apartments) and nine districts ("*Stadtbezirke*") for apartments, eight districts for single-family-houses,[130] and three parts of the city for multi-family-houses.[131] We clean the data for duplicates,[132] out-

129. Clear non-market-price transactions, for example sales between family members, have been excluded by the GA from the sample.

130. We excluded the city-center district ("*Innenstadt*"), because nearly no sales of single-family houses occur in this district.

131. As the number of multi-family houses transacted is considerably lower than for the other types, we aggregates the district to three parts: city center ("*Stadtbezirk Innenstadt*"), districts left of the Rhine ("*Stadtbezirke*": "Chorweiler", "Nippes", "Ehrenfeld", "Lindenthal", and "Rodenkirchen") and districts right of the Rhine ("Stadtbezirke": "Mülheim", "Kalk", and "Porz").

132. For apartments we also delete all duplicates in address, sales price and transaction year, because we want to exclude package deals.

liers in sales prices as well as outliers in the (non dummy) dependent variables.[133] After this procedure, we are able to rely on 12,538 transactions for multi-family houses, 37,972 transactions for single-family houses and 131,744 transactions for apartments.

To calculate the indices we closely follow the methodology in Eurostat (2013). The final index is constructed in three steps: First, double imputation hedonic indices are calculated separately for single-family and multi-family houses and separated by district or part of city. Chaining is used to connect different years. Hedonic regressions for single-family houses use lot area as dependent variable.[134] After 2010, a dummy is additionally included, which controls for houses being detached or attached. Hedonic regressions for multi-family houses use lot area and a dummy for commercial use as exogenous variables.[135] To construct the index for apartments, hedonic regressions use area of the apartment,[136] and a dummy for whether the flat is newly constructed.[137] From 2010 onward, a dummy is included for whether the flat is in a high-rise building (more than ten floors).[138] Starting 2015, we additionally control for quality of the location ("*Lagequalität*"), which was estimated by the GA in four categories.[139] All regressions are performed using maximum likelihood and assuming normally distributed standard errors. Second, for all three housing types separately a Fisher-type stratification index is built from the district-level/part-level indices using chaining and transaction value shares as weights. Lastly, the final index is obtained by stratification of the respective indices for single-family houses, multi-family houses and apartments. We again use transaction value shares as weights. Trends for the indices covering different housing types are very similar.[140]

133. For multi-family houses, we also clean the data for buildings with a large number of floors (more than ten) for all non-missing observations to exclude high-rise buildings. These buildings are rarely transacted and will follow a considerably different pricing compared to other multi-family houses. We assume that for all high-rise buildings the number of floors is non-missing as this variable is of considerably more importance for these buildings.

134. Before 1992, we do not use any dependent variables, as we have too many missing observations for lot area. In consequence, for the first three years, the single-family-house index is in fact a stratification index.

135. All other variables have many missing observations; before 1992, hedonic regressions do not use any dependent variables, such that the multi-family house index is in fact a stratification index for these three years.

136. Observations with extreme or missing values are excluded.

137. The dummy is equal to one if the apartment is less than two years old at the point of sale or marked as "*Neubau*" (new construction).

138. Before, the number of floors of the building is missing for nearly all apartments.

139. The GA defines four quality bins ("*einfach*", "*mittel*", "*gut*" and "*sehr gut*"), which we include as factor variables. Quality of the location does not differ much within districts, such that excluding it from the regression has has no visible effect on the overall index.

140. See Appendix of the main paper.

For the period between 1981 and 1989 only the data for apartments for the first four districts of Cologne are available,[141] which covers 7,600 observations. We use this data to construct hedonic indices separated by district as described above and use stratification with transaction value shares as weights to aggregate the indices for the four districts.

In their archives, the GA Cologne has very detailed maps which show transaction prices and year of sale by address. We use these maps for central parts of the city to extract sales prices for houses before 1989. Unfortunately, the number of transactions marked on these maps decreases for earlier years. Moreover, extracting transaction prices and addresses from maps requires a considerable amount of manual labor. Consequently, the number of observations per year is considerably below the number for the micro-level data we use after 1989.[142] We match the resulting transaction-level data with the data on housing sales after 1990 to build a repeat sales index. Our data features on average 25 matched observations per year between 1966 and 1989. To build the repeat sales index we follow Eurostat (2013) and adjust for heteroscedasticity in standard errors using the weighted least squares (WLS) approach suggested by Case and Shiller (1987). As the number of observations per year used to build the repeat sales index is low, we use the hedonic index for apartments described above after 1981. Both indices, however, follow a similar trend.

Between 1948 and 1975, Statistics Cologne published aggregated data on lot sales for the city of Cologne in the statistical yearbooks.[143] Between 1948 and 1958, we use the average price for a developed lot in the city of Cologne. Between 1959 and 1972, the yearbooks only contain averages for developed and single undeveloped lots prepared for construction ("*Einzelbaustellen*") pooled together. Only after 1972, are statistics again published separately for developed lots. As prices of developed and undeveloped lots might have featured different time trends, we calculate the total increase for developed lots only between 1958 and 1973 and use the average transaction price of the pooled series only to interpolate the index between these years. We use the resulting index only until 1966, when the repeat sales index becomes available.

Prior to World War II, we rely on various publications by Statistics Cologne.[144] These contain aggregate data for transaction volume, area and numbers of sales of

141. For this period, apartment data only exist for the districts "Innenstadt", "Rodenkirchen", "Lindenthal", and "Ehrenfeld". In the apartment data between 1989 and 2018 the first four districts cover 58% of all transactions.

142. We were able to extract 1,577 observations between 1966 and 1988.

143. "*Statistisches Jahrbuch der Stadt Köln*" (Volumes 1948-1957).

144. "*Statistisches Jahrbuch der Stadt Köln*" (Volumes 1948-1957) and "*Cölnische Statistische Vierteljahreshefte*" (Volumes 1906-1913).

developed lots by district.[145] We use the average price per square meter of developed lots to build Fisher-type chained stratification indices following Eurostat (2013) for the period from 1904 to 1941.[146] This way, we control for locational shifts in the sample of sold lots as well as for changes in the average lot size. For the years from 1902 to 1904, we had to rely on the average price per transaction of developed lots within the entire city of Cologne from Neuhaus (1916).[147] We match the data between 1941 and 1948 using the average price per transaction of developed lots within the entire city.

Table 1.K.17 summarizes the components of our final house price index.

Rent Series. As done for Berlin, we use a rent index constructed from market reports of the German Real Estate Association and its predecessor for the most recent period. Before, we use either data from CPI index construction or from other statistical publications. In contrast to Berlin, however, historical rent data for Cologne is less reliable, because it either covers only a subset of the rental market within Cologne or a broader region than just the city of Cologne. To minimize resulting biases in our long-run rent index, we use data from housing censuses in 1890, 1910, 1918, 1950, 1956, 1968 and 1987 to adjust our rental index. The yearly data sources are only used to interpolate the resulting index between census years. This procedure ensures that the trend in our rent index reproduces the rent development of the entire rental market within Cologne, because it minimizes sample selection biases, while the detailed census data also enable us to use standard quality controls.

The data from housing censuses cover the universe of all rented residential units within Cologne. They are taken from various statistical publications.[148] These sources typically report average rent disaggregated by number of rooms, construction period and sometimes even district. We calculate increases between consecutive census years using Fisher-type stratification following Eurostat (2013). Calculating these increases does, however, pose two challenges. First, the way of aggregating and reporting the data changed between census years, such that straightforward matches are not always possible. Second, the city of Cologne grew over time, both by extending its borders spatially and by a high level of new construction. To compare

145. For 1904 and 1905 the data is aggregated by four parts of former Cologne ("Altstadt", "Neustadt", "innere Vororte", and "äußere Vororte"; between 1906 and 1912 13 districts are covered, between 1913 and 1929 sixteen districts and afterwards again thirteen, because three were merged with their neighboring districts.

146. For consecutive years, during which the number of districts/parts change, we aggregate the data for the more detailed year, such that we can build a chained stratification index for the same areas between both years.

147. We collected the number of transactions from statistical yearbooks.

148. 1890, 1910: Neuhaus (1915); 1918: "*Statistisches Jahrbuch Köln*" (Volume 7, 1919); 1950: "*Statistische Mitteilungen der Stadt Köln*" (1955); 1956: "*Statistische Mitteilungen der Stadt Köln*" (1958); 1968: "*Statistisches Jahrbuch Köln*" (Volume 56, 1969); 1987: "*Sonderreihe zur Volkszählung 1987 in Nordrhein-Westfahlen Band Nr. 6.1.*".

Table 1.K.17. Final house price index for Cologne

PERIOD	SOURCE	DESCRIPTION
1902-1904	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated for former city from statistical publications; *Method*: Average price per transaction.
1904-1941	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated by district from statistical publications; *Method*: Stratification.
1948-1966	Own compilation	*Type(s) of dwellings*: All developed (and undeveloped) lots; *Type of data*: All sales aggregated for former city from yearbooks; *Method*: Average price per transaction (the series containing undeveloped lots is used for interpolation only).
1966-1981	Own compilation	*Type(s) of dwellings*: Houses in central locations; *Type of data*: Transaction-level data from archived maps from the Gutachterausschuss Cologne; *Method*: Repeat-sales index.
1981-1989	Own compilation	*Type(s) of dwellings*: Universe of sold apartments in the first four districts of Cologne; *Type of data*: Transaction-level data kindly provided by the Gutachterausschuss Cologne; *Method*: Stratified hedonic index.
1989-2018	Own compilation	*Type(s) of dwellings*: Universe of single-family houses, multi-family houses and apartments; *Type of data*: Transaction-level data kindly provided by the Gutachterausschuss Cologne; *Method*: Stratified hedonic index.

like with like, we always only compare consecutive census years and additionally try to exclude all buildings that were not part of the earlier census year. We provide more details below.

To calculate the stratification increase between 1890 and 1910 we use data disaggregated by number of rooms (1-4) and three parts of the city.[149] Using this stratification, we are able to control for the size of the flat and the rough location. We are not able to exclude buildings constructed after 1890, but as we control for the most relevant quality changes and the two census years are only 20 years apart, we think that this does not bias the result considerably. The data from the census in 1918 are only disaggregated by number of rooms, such that we can only stratify by the size of the dwelling between 1910 and 1918.[150] The census data from 1950 are

149. "Altstadt", "Neustadt" and "Vororte".

150. This can only be taken as an approximation, because the city of Cologne grew in space between 1910 and 1918. To the best of our knowledge there are, however, no better data available for 1918.

given by number of rooms and building period. To calculate the increase between 1918 and 1950, we only rely on flats built prior to 1918, such that we compare the same flats over time. Moreover, we stratify by the number of rooms to control for sample shifts in the size of flats. The data for 1956 are given by building period and district and contain rent per dwelling as well as rent per room. We calculate the increase between 1950 and 1956 using rent per room and stratifying by building period[151] using the total number of rooms as weights. This way, we control for size of the dwelling and time of construction to compare almost the same buildings over time. For 1968, the data contain the average rent per square meter by building period and district. To compare the data to 1956, we calculate the average rent per room using data for the average size of rooms from the same census year.[152] To calculate the increase between 1956 and 1968 we then only consider dwellings built prior to 1948 to exclude new construction after 1968,[153] and stratify the data by district ("*Stadtbezirk*"). This way, we compare only dwellings that already existed in both census years, control for size by taking rent per room, and control for location by stratification. The data from the 1987 census contain rent per square meter disaggregated by district. This enables us to exclude all new districts added to the city of Cologne in between.[154] To calculate the index between 1968 and 1987, we stratify the data by district and use average rent per square meter.[155] This allows us to control for size and location. We are, however, not able to control for building period, because the necessary data are missing in 1987. As we control for the most relevant quality changes over time, we assume that this does not bias the results.

Given these overall increases from the census data, we construct a yearly rent index from various sources to interpolate and extrapolate the index between census years. Starting in 1973, we use the rent data from the German Real Estate Association and its predecessor. The construction for rent indices from the real estate market reports is described in the Appendix of the main paper.

From 1948 to 1975, we have to rely on a rent index constructed for the CPI of North Rhine-Westphalia from statistical yearbooks.[156] This index tracks rents for four-person employee ("*Arbeitnehmer*") households. It covers several cities within the federal state and not only the city of Cologne. Cologne is, however, the largest city within North Rhine-Westphalia and, therefore, presumably was given a large

151. We only use building periods until 1948 in the 1956 data to exclude new construction after 1950.

152. We calculate the average size of rooms as weighted average of the average size of all rented dwellings in Cologne with x rooms divided by the number of rooms x, weighted by total size of flats with x number of rooms; source "*Statistisches Jahrbuch Köln*" (1970).

153. The data for 1968 only breaks building periods in 1948.

154. The city of Cologne grew considerably in 1975; all new districts of Cologne and districts that could not clearly be associated to an old district are dropped from the 1987 data.

155. As total number of square meters is not given by districts, we instead weight by the total number of flats by district.

156. "*Statistisches Jahrbuch der Stadt Köln*" (Volumes 1948-1973).

weight during index construction. Moreover, rent developments of neighboring cities can be assumed to be correlated with the developments in Cologne. Indeed, the overall increases are similar comparing the census data and the CPI rent index.[157]

Prior to World War II, we are able to rely on data collected for a city-level CPI index for the city of Cologne from statistical yearbooks.[158] Between 1921 and 1925 as well as 1928 and 1942, we use a rent index for working-class apartments with two rooms and a kitchen.[159] For the period from 1919 to 1921 as well as for 1926 we use the mode monthly rent (including black market) for working-class apartments with two rooms and a kitchen collected to calculate the CPI index from statistical yearbooks.[160] We interpolate the growth rate for 1927. The CPI index is given in 1914 values, such that we can link the CPI data to the data ending in 1918, correcting for the index increases between 1914 and 1918. We link the data from 1942 to 1948 using the rent increases calculated from the census data.

During World War I, *Statistics Cologne* published average rents of vacant dwellings by number of rooms and city district.[161] We use these data to build a stratification index using housing stock in 1910 as (constant) weights. This way, we can control for locational shifts as well as shifts in dwelling size for the sample of vacant apartments.

For the period between 1904 and 1913, we use market rents for vacant dwellings designed for workers ("*Arbeiterwohnungen*") published by *Statistics Cologne*.[162] The data is given by number of rooms and part of the city.[163]. We aggregate the quarterly data for all non-missing quarters to get yearly averages. We then calculate increases using Fisher-type stratification by number of rooms and part of city.

Table 1.K.18 summarizes the components of our final rent index. It first depicts the years for which census data exist. Then it describes the components of the yearly rent index. We only use the yearly index to interpolate and extrapolate between and after census years, because the yearly data do not cover the ideal sample of the entire rental market in Cologne before 1973, whereas the census data cover the universe of all rented dwellings.

Rental Yield Series. Our main benchmark for Cologne is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to all alternative

157. Between 1950 and 1956, we calculate an increase of rents in Cologne from census data of a factor of approximately 1.28 and the CPI rent index for North Rhine-Westphalia increased by a factor of 1.20. For the period between 1956 and 1968 the indices increased by a factor of 2.18 and 2.03, respectively.

158. "*Statistisches Jahrbuch der Stadt Köln*" (Volumes 1921-1942).

159. 1926 and 1927 are missing from the CPI index data.

160. "*Statistisches Jahrbuch der Stadt Köln*" (Volume 1920-1926).

161. We use 14 out of 16 districts, which feature decent data coverage; source: "*Kölner Statistik 2. Jahrgang Heft 1*" (1919).

162. "*Cölnische Statistische Vierteljahreshefte*" (Volumes 1904-1913).

163. "Altstadt", "Neustadt", and "Vororte"

Table 1.K.18. Final rent index for Cologne

PERIOD	SOURCE	DESCRIPTION
1890, 1910, 1918, 1950, 1956, 1968, 1987	Own compilation	*Type(s) of dwellings*: All rented residential dwellings; *Type of data*: Census data from statistical publications; *Method*: Stratification - we use these data to adjust the trend of the overall index as described in the main text.
1904-1913	Own compilation	*Type(s) of dwellings*: Vacant rental dwellings designed for workers; *Type of data*: Average rent by part of city and number of rooms; *Method*: Stratification.
1914-1918	Own compilation	*Type(s) of dwellings*: All kinds of vacant rental dwellings; *Type of data*: Average rent by district and number of rooms; *Method*: Stratification.
1919-1921	Statistics Cologne	*Type(s) of dwellings*: Working-class apartments with two rooms and a kitchen; *Type of data*: From CPI-construction; *Method*: Mode prices.
1921-1942	Statistics Cologne	*Type(s) of dwellings*: Working class apartments with two rooms and a kitchen; *Type of data*: From CPI-construction; *Method*: CPI rent index.
1948-1975	Statistics North Rhine-Westphalia	*Type(s) of dwellings*: Apartments for four-person employee households; *Type of data*: From CPI-construction for North Rhine-Westphalia; *Method*: CPI rent index.
1973-2018	Own compilation	*Type(s) of dwellings*: Apartments; *Type of data*: New contract rents by construction period and housing quality reported by local real estate agents; *Method*: Matched-model approach using mode prices by quality and construction period bins.

benchmarks we collected for 2018. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.10.

To verify the historical rental yield series, we collected alternative benchmarks from three different sources. First, there only existed a benchmark for Cologne from *Numbeo.com* for 2015. This benchmark is below, but close to our long-run series.

Second, we use benchmarks from the market reports of the *German Real Estate Association* (IVD). These market reports directly report mode price-rent ratios for residential investment buildings from 1989 onward. The Appendix in the main paper describes how we calculate net rental yields. The resulting rental yield benchmarks

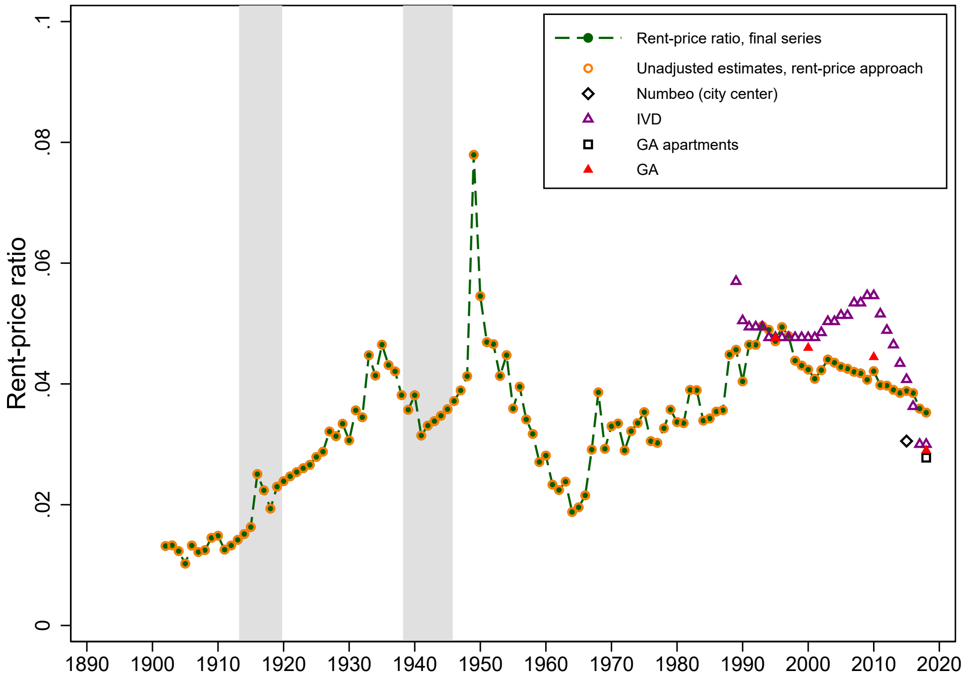

Figure 1.K.10. Cologne: plausibility of rental yields

are plotted as purple triangles in Figure 1.K.10. This series shows somewhat more cyclicality compared to our long-run series, but is overall very close in values.

Third, in their yearly market reports, the GA publishes average price-rent ratios for investment buildings.[164] Following Jordà, Knoll, et al. (2019) we subtract one third of these to get net rental yields. The resulting benchmarks are depicted as red triangles in Figure 1.K.10. These benchmarks are very close to our long-run rental yield series. For 2018, we also calculate rental yields for apartments given the average prices and rents per square meter provided by the GA in the yearly report. The resulting value is plotted as a black square in Figure 1.K.10 and is very close to the rental yields for investment buildings. We use this benchmark as our alternative benchmark in the robustness section of the main paper, because it has a broader coverage compared to just covering investment buildings and relies on micro-data of actual transaction prices.

To summarize, all benchmark values we collected are close to our original rental yield series. As rental yield benchmarks are subject to measurement error by themselves, we do not adjust our rental yield series for Cologne to any benchmarks. We

164. ("*Rohertragsfaktor*"); we take the average value for normal rented investment buildings ("*Mietwohnhäuser*"); in cases when only ranges are given, we take the midpoint from these ranges; source: "*Grundstücksmarktbericht für die Stadt Köln*" (2019, 2011, 2000, 1995).

were not able to find earlier historical rental yield benchmarks for the city of Cologne. However, as the rental yield series seems to be plausible compared to other German cities and we adjust the rent series for sample selection biases in the trend using census data, we assume the resulting rental yield series to be accurate.

1.K.6.3 Frankfurt

House Price Series. The house price series for Frankfurt is constructed using data from the GA Frankfurt starting in 1982. For the period from 1960 to 1982, the GA kindly provided us access to their analogue archives, so we were able to digitize a subsample of their transaction records. Prior to 1960, we had to rely on various publications of Statistics Frankfurt. From these sources, we calculated the average transaction price per square meter of developed lots. Unfortunately, the earlier data do not allow us to control for locational shifts in the sample of transacted buildings within the city of Frankfurt.

The GA Frankfurt kindly provided us with transaction-level data for single- and multi-family houses for the period 1982 to 2018. These data cover the universe of all normal housing transactions in Frankfurt that were considered to be market prices.[165] We use cleaned transaction prices provided by the GA. These values adjust for sales conditions unrelated to the house itself, for example if the inventory or a kitchen was sold with the house or if specific rights were still granted to the seller. We construct separate hedonic indices for single-family and multi-family houses. Moreover, we calculate indices separately for five parts of the city for single-family houses and four parts of the city for multi-family houses.[166] We drop all observations that have missing prices or missing values in one of the dependent variables.[167] After this step, we use 6,093 transactions for multi-family houses and 16,237 transactions for single-family houses between 1982 and 2018.

To calculate the indices we closely follow the methodology in Eurostat (2013). The final index is constructed in three steps: First, double imputation hedonic indices are calculated separately for single-family and multi-family houses and separated by part of the city. Chaining is used to connect different years. Hedonic regressions for single-family houses use lot area as dependent variable as well as a dummy being 1 if the house has a garage and a dummy for houses being classified as a townhouse

165. The GA cleaned the data of non-market-price transactions, for example transactions between family members.

166. We aggregate the districts ("*Ortsbezirke*") 1-4 to form the city center, district 5 forms the part "South", district 6 forms the part "West", we aggregate districts 7 to 11 to form a part in north-east, which is closer to the center, and the districts 12 to 16 as the outer north-east part. For multi-family houses, we aggregate the inner and outer parts in the north-east, because the number of observations in the outer part is very low.

167. This implies that we have to ignore a considerable part of the data set. We use a low number of only highly relevant dependent variables, in order to minimize the number of observations we have to exclude.

("*Stadthaus*").[168] After 2005, we do not use these dummies anymore, but additionally control for floor area.[169] Starting 2006, we also use dummy variables to control for type of house.[170] Regressions for multi-family houses use volume of structure ("*Bruttorauminhalt*") as exogenous variable.[171] Starting in 2005, we instead use floor area in the regressions. All regressions are performed using maximum likelihood and assuming normally distributed standard errors. Second, for both housing types separately a Fisher-type stratification index is built from the part-level indices using chaining and transaction value shares as weights. Lastly, the final index is obtained by stratification of the respective indices for single-family houses and multi-family houses. We again use transaction value shares as weights.

For the period from 1960 to 1981, we hand-collected transaction-level data from the archives of the GA. The reporting of the data changed from 1970 to 1971, so we have to use different methodology before and after 1971. For the period between 1971 and 1981, we were able to collect data on nearly all normal sales of residential houses that can be assumed to have been market prices. We clean the data of all buildings, which seem to be commercial or public, which were demolished after the transaction or which were bought to construct a road. Moreover, we also exclude transactions in which only a part of the right to the land was sold ("*ideeller Anteil*"). Again, we drop all observations that have missing prices or missing values in one of the dependent variables. This leaves us with 3,854 observations. During this period, we are not able to separate between single-family and multi-family houses. Consequently, it is of special importance to control for the size of the building and location. The final index is constructed in two steps: First, double imputation hedonic indices are calculated separately for six parts of the city.[172] Chaining is used to connect different years. Hedonic regressions use an interaction between lot size and allowed floor space ratio ("*Geschossflächenzahl*") as dependent variable. This interaction term describes a legal ceiling on the total floor space allowed on a developed lot

168. Luxurious buildings that sell for a premium.

169. The data structure changes in 2006, such that we have to adapt hedonic regressions.

170. The data classifies eight different types of single-family-houses: semidetached house ("*Doppelhaushaelfte (Einfamilienhaus)*"), detached house ("*Einfamilienhaus (freistehend)*"), row house ("*Reihenhaus (Einfamilienhaus)*"), row-end house ("*Reihenendhaus (Einfamilienhaus)*"), villa ("*Villa*"), individual building style ("*individuelle Bauweise*"), two-family house ("*Zweifamilienhaus*"), and three-family-house ("*Dreifamilienhaus*").

171. The number of missing observations is lower for volume of structure compared to floor area before 2005. Both variables are reasonable proxies for the size of the building.

172. We use the separation described above, but separate the last part again, such that districts ("*Ortsbezirk*") 12 to 15 form the outer north part and district 16 the outer east one. As we are not able to separate between housing types and because different districts typically feature different housing types, the local separation becomes even more important here. Moreover, districts 12 to 15 were only incorporated by Frankfurt in August 1972 and district 16 in January 1977. This implies that the indices for these districts start later (1974 and 1977, respectively). During chaining, we only compare parts of the city for which an index exists for both consecutive years.

and therefore approximates the size of structure.[173] We do not control for lot area separately.[174] All regressions are performed using maximum likelihood and assuming normally distributed standard errors. Second, a Fisher-type stratification index is build from the part-level indices using chaining and transaction value shares as weights.

To connect the data of 1981 and 1982, we build dummies to categorize the floor space ratio (FSR).[175] To do so, we approximate the FSR for the data in 1982 using actual floor area and lot area.[176] Then, we again calculate indices separately for the five parts of the city also used after 1982. Hedonic regressions use the FSR dummies interacted with lot size. Afterwards, a Fisher-type stratification index is build from the part-level indices using chaining and transaction value shares as weights.

Before 1971, we collected data on all normal sales of residential houses at market prices for most districts of the former city of Frankfurt. We were, however, not able to collect a sufficient number of observations in the west of Frankfurt,[177] such that we had to exclude the western part from our sample. Moreover, we exclude all parts that were not officially part of Frankfurt between 1960 and 1971.[178] We cleaned the data for all buildings, which were (mainly) commercial or public, which were demolished after the transaction, which were used to construct a road or did not have an address. We also exclude all transactions in which only a part of the right to the land or a part of the building was sold ("*ideeller Anteil*"). We drop all observations that have missing prices or that could not be classified as either single-family or multi-family houses.[179] After this procedure, we have 1,525 observations for multi-family houses and 1,235 observations for single-family houses. We again build a two-step stratification index stratifying by housing type and within types by three

173. As Frankfurt is a densely populated city, this variable will be highly correlated to actual size of structure. Even in cases when the ceiling was not actually reached, the value of the land will still be highly dependent on the allowed floor space in a dense city.

174. Coefficients for lot area have been very unstable in these regressions. The reason might be that controlling for the size of the structure, lot area and micro-location will be correlated with larger lots typically being in less expensive locations.

175. We build four categories: FSR below 1, FSR between 1 and 2, FSR between 2 and 3 and FSR above 3.

176. This approximation will be downward biased, because to calculate the FSR, a different measure of floor area is used ("*Geschossfläche*" instead of "*Wohnfläche*"). To minimize the effect of the resulting bias, we use the categories instead of controlling for FSR directly. This way, the categories still separate different housing types. Indeed, the resulting index increases considerably less than if using the FSR as control variable directly revealing a considerable downward bias in the approximation of the FSR for 1982. At the same time, it increases noticeably more compared to an index only controlling for lot size, implying that there might have been a large sample shift selling relatively more multi-family houses in 1982. The resulting increase is closer but below the increase found in the real estate market reports data between the two years.

177. District ("*Ortsbezirk*") 6.

178. Districts 12-15.

179. We use the utilization ("*Nutzung*") variable for this classification. This variable is missing for approximately 5% of the sample, which we had to exclude to construct the index.

parts of the city. As control variables, we only use lot size and for multi-family houses additionally a dummy for whether they have some parts in commercial use.[180] All regressions are performed using maximum likelihood and assuming normally distributed standard errors. To merge 1970 to 1971, we use the same procedure as for the data between 1960 and 1970.[181]

Prior to 1960, we rely on aggregate transaction data for developed lots from various statistical publications.[182] For 1897 to 1934 and 1952 to 1960 we use the average price per square meter of developed lots calculated from the universe of normal sales within the former city of Frankfurt.[183] We link the data from 1934 directly to 1952 using this average price. Unfortunately, as the data are not given disaggregated by district, we have not been able to control for locational shifts within the sample of developed lots. To interpolate the index for the years 1935 to 1938, we had to rely on aggregate transaction data for developed lots, which included normal sales alongside exchanges of land ("*Tausch*") as well as voluntary auctions ("*freiwillige Versteigerung*"). We again use average price per square meter of developed lots, so that we adjust for sample shifts within the size of transacted lots.

Table 1.K.19 summarizes the components of our final house price index.

Rent Series. Considering all data sources we knew of for city-level rent developments, it proved impossible to build a continuous yearly rent index for Frankfurt. To link different rent indices from various sources over time and to minimize the bias resulting from using different sources covering different market segments, we again use data from housing censuses. These data cover the universe of all rented residential dwellings in Frankfurt and enable us to control for size and location within the mix of rented dwellings. It therefore provides a precise picture of city-level rent development. We start by calculating rent increases between the census years 1895, 1905, 1910, 1956, 1968 and 1987. To obtain a yearly rent index, we then use yearly data from the market reports of the *German Real Estate Association* and its predecessor and from *Statistics Frankfurt* to interpolate and extrapolate between and after census years.[184]

180. Information on floor area or GFZ is completely missing from the data until 1971.

181. As utilization is missing for a larger part of the sample between 1971 and 1984, we had to exclude a large part of transactions in 1971 to construct the index between 1970 and 1971. As it is not possible to control for the size of the structure in 1970, however, being able to differentiate between single-family and multi-family houses is crucial to control for sample shifts between both years.

182. 1952-1960 and 1938-1939: "*Statistisches Jahrbuch für Frankfurt am Main*" (Volumes 1952 - 1962); 1935-1936: "*Statistisches Jahrbuch deutscher Gemeinden*" (Volumes 1937 - 1938); 1927 - 1934: "*Statistische Jahresübersichten der Stadt Frankfurt a. Main*" (Volumes 1927/28 - 1934/35); 1897 - 1926: "*Statistisches Handbuch der Stadt Frankfurt am Main*" (Volumes 1905/06 and 1929).

183. Data for 1923 to 1925 is missing because of the hyperinflation in Germany.

184. The housing census in 1987 was the last census in Germany that surveyed rents for the universe of all residential rental dwellings.

Table 1.K.19. Final house price index for Frankfurt

PERIOD	SOURCE	DESCRIPTION
1897-1934	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated for former city from yearbooks; *Method*: Average price per square meter of developed land.
1935-1938	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales, exchanges and voluntary auctions aggregated for former city from yearbooks; *Method*: Average price per square meter of developed land.
1952-1960	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated for former city from yearbooks; *Method*: Average price per square meter of developed land.
1960-1982	Own compilation	*Type(s) of dwellings*: Near universe of single-family and multi-family houses; *Type of data*: Transaction-level data from archived records of the Gutachterausschuss Frankfurt; *Method*: Stratified hedonic index.
1982-2018	Own compilation	*Type(s) of dwellings*: Universe of single-family and multi-family houses; *Type of data*: Transaction-level data kindly provided by the Gutachterausschuss Frankfurt; *Method*: Stratified hedonic index.

The data from housing censuses are taken from various publications by *Statistics Frankfurt*.[185] They cover the universe of all rented residential dwellings within Frankfurt and provide average rents disaggregated by number of rooms, construction period and district. We calculate increases between consecutive census years using stratification indices following Eurostat (2013). Calculating these increases does, however, pose two challenges. First, the way of aggregating and reporting the data changed between census years, such that it is often not straightforward to match the data. Second, the city of Frankfurt grew over time, both by extending its borders spatially and by a high level of new construction. To compare like with like, we exclude buildings constructed between consecutive census years as far as possible. We provide more details in the following paragraph.

To calculate the stratification index between 1895, 1905 and 1910, we use data disaggregated by number of rooms (1-6) and subdistrict ("*Stadtbezirk*", number of

185. 1895, 1905, 1910: "*Beiträge zur Statistik der Stadt Frankfurt am Main 11. NF*" (1919); 1956: "*Statistisches Jahrbuch für Frankfurt am Main*" (Volume 1958); 1968: "*Statistisches Jahrbuch für Frankfurt am Main*" (Volume 1971); 1987: "*Frankfurter Statistische Berichte*" ("Sonderheft, Bd. 54").

subdistricts ranging between 48 and 54 depending on the census year).[186] We calculate Laspeyres-type stratification indices using 1910 as baseyear and the total number of rented dwellings by number of rooms and district as weights.[187] Using this stratification, we are able to control for sample shifts in the size of dwellings as well as for fine-grained locations. We are not able to exclude buildings constructed prior to 1910, but as the gap between censuses is not longer than ten years and building periods are highly correlated to locations within cities, we are confident that this does not bias our results. To connect the census data from 1910 and 1956,[188] we use average rent per room disaggregated by subdistrict. To do so, for the 1910 data, we calculate average rent per room for each of the subdistricts ("*Stadtbezirk*", n=50) separately using average rent per dwelling and number of dwellings by subdistrict and number of rooms (1-9). For the 1956 data, we exclude all subdistricts that did not belong to Frankfurt already in 1910 and only use data for dwellings built prior to July 1, 1918.[189] By doing so, we exclude nearly all dwellings built after 1910.[190] With these data at hand, we calculate a Fisher-type stratification index using total number of rooms in rented dwellings per district as weight. The census data in 1968 provide rent per square meter by another classification of subdistricts ("*Ortsteil*", n=38) and building period. We calculate average rent per room from rent per square meter using the average size of dwellings and the average number of rooms per dwelling by subdistrict from the same census year.[191] We aggregate the data from 1956 to the same subdistricts ("*Ortsteil*", n=38) and aggregate all dwellings built prior to 1948. Next, we built a Fisher-type stratification index using rent per rooms of dwellings built prior to 1948. This way, we only compare flats from the same construction period and exclude any new construction. Additionally, we control for any sample shifts in the size or location of dwellings. We calculate the index between 1968 and 1987 using rent per square meter disaggregated by subdistrict ("*Ortsteil*", n=38) and building period. We only use the two building periods prior to 1948 and between 1949 and 1968 for the 1987 census, so we exclude any new construction between 1968 and 1987.[192] We calculate the index between both years using Fisher-type stratification by subdistrict and the two building periods. As

186. For each comparison, we match all subdistricts that were part of Frankfurt in both respective years.

187. Data on the number of dwellings by number of rooms and districts are missing for the census years prior to 1910.

188. Data for 1918 and 1950 are missing for Frankfurt. Using 1910 and 1956 we are, however, able to connect rent indices prior to and after World War II accurately.

189. Average rent is only given pooled for all dwellings built prior to 1918, so we are not able to only use dwellings built prior to 1910.

190. Because of World War I, new construction between 1910 and 1918 will have been very low.

191. Source: "*Frankfurter Statistische Berichte*" ("Sonderheft, Bd. 54").

192. We aggregate rent per dwelling for dwellings built with and without public financing for the period between 1949 and 1968 from the 1987 census using a weighted average of rent per square meter with number of dwellings by category and district as weights.

the total number of square meters by building period and subdistrict is missing, we instead use the total number of dwellings in each stratum as weight. Overall, this procedure arguably gives a reliable picture of rent increases for the rental market in Frankfurt, as we use the universe of all flats excluding new construction and are able to control for size, fine-grained inner-city location and in many cases even building period.

Given these overall increases from the census data, we construct yearly rent indices from various sources to interpolate and extrapolate the long-run index between census years. Starting in 1972, we use the rent data from the market reports of the German Real Estate Association and its predecessor. The construction for rent indices from these market reports is described in the Appendix of the main paper.

For the period from 1949 to 1965 and for 1938, we use monthly rent of a three-room apartment (two rooms plus kitchen) in average distance to the city center built prior to 1924, which was published by Statistics Frankfurt.[193] We link the resulting index with the index for 1968 and the data from the real estate agents' market reports using the census index in 1956, 1968 and 1987. To calculate the rental yield series, we had to linearly interpolate the years 1966, 1967 and 1969 to 1971.

Prior to World War II, we use two different kinds of data from historical publications by Statistics Frankfurt. For the period between 1924 and 1935, we rely on the rent component of the city-level CPI index.[194] The CPI is calculated for less well-off ("*minderbemittelt*") five person households (two adults and three children with age 12, 7 and 1½). The rent component excludes heating and lighting costs. We calculate yearly averages from the monthly data. The index is given in 1914 values, such that the linking to the older data is straightforward. Between 1897 and 1920, we instead use the average rent of dwellings that have been newly rented ("*bezogene Wohnung*") and dwellings for which the rent contract has just ended ("*verlassene Wohnung*") by number of rooms (1-6 or more).[195] We calculate a weighted average rent for each category separately using the number of dwellings by number of rooms from the census in 1910 as weights. Next, we build a simple average over both categories. The resulting index is very close to the index calculated from the housing census data.[196] Data for the period of the hyperinflation in Germany is missing.

Table 1.K.20 summarizes the components of our final rent index. It first depicts the years for which we use the housing census data. Then it describes the components of the yearly rent index. We use the census data to connect the different yearly rent indices. For periods during which both census data and a yearly index exist, we

193. Source: "*Statistisches Jahrbuch für Frankfurt am Main*" (Volumes 1951 - 1966)

194. Sources: "*Statistisches Handbuch der Stadt Frankfurt am Main*" (1928) and "*Statistische Jahresübersichten der Stadt Frankfurt a. Main*" (Volumes 1927/28 - 1934/35).

195. Source: "*Statistisches Handbuch der Stadt Frankfurt am Main*" (1928).

196. Between 1905 and 1910, rents increased in total by 5,96% according to the yearly data and by 5,26% according to the data from the housing censuses. During this period, we still correct for the difference by taking the overall increase from the housing census and impute using the yearly series.

use the yearly indices only to interpolate and extrapolate between and after census years, because we deem the census data to be more reliable.

Table 1.K.20. Final rent index for Frankfurt

PERIOD	SOURCE	DESCRIPTION
1895, 1905, 1910, 1956, 1968, 1987	Own compilation	*Type(s) of dwellings*: All rented residential dwellings; *Type of data*: Census data from statistical publications; *Method*: Stratification - we use these data to link the yearly indices as described in the main text.
1897-1920	Own compilation	*Type(s) of dwellings*: Newly rented and just canceled rented residential dwellings; *Type of data*: Average rent by number of rooms; *Method*: Weighted average.
1924-1935	Statistics Frankfurt	*Type(s) of dwellings*: Apartments for a less-well-off five person household; *Type of data*: From CPI-construction; *Method*: CPI rent index.
1938, 1949-1965	Statistics Frankfurt	*Type(s) of dwellings*: Three-room apartments in average distance to the city center built prior to 1924; *Type of data*: Monthly rent; *Method*: Estimated rent of standardized dwelling.
1972-2018	Own compilation	*Type(s) of dwellings*: Apartments; *Type of data*: New contract rents by construction period and housing quality reported by local real estate agents; *Method*: Matched model approach using mode prices by quality and construction period bins.

Rental Yield Series. Our main benchmark for Frankfurt is taken from *MSCI*, as described in the main paper. This benchmark is very close to all alternative benchmarks we collected for 2018. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.11.

To verify the historical rental yield series, we collected alternative benchmarks from three different sources. First, we use benchmarks from the market reports of the *German Real Estate Association* (IVD). These market reports directly report mode price-rent ratios for residential investment buildings from 1989 onward. The Appendix in the main paper describes how we calculate net rental yields. The resulting rental yield benchmarks are plotted as purple triangles in Figure 1.K.11. This series is very close to our long-run rental yield series.

Second, the transaction-level data provided by the GA Frankfurt contains the yearly gross income for investment buildings between 1982 and 2018. We use these data to calculate mean gross rental yields for these buildings. Following Jordà, Knoll,

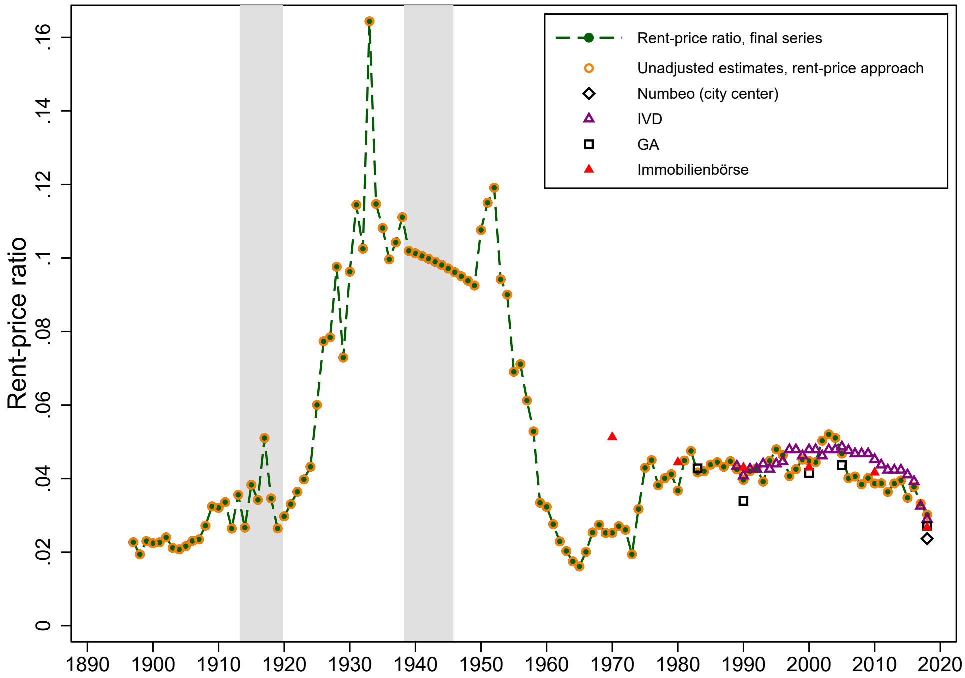

Figure 1.K.11. Frankfurt: plausibility of rental yields

et al. (2019) we subtract one-third to get net rental yields. We plot benchmarks for the years 2018 (calculated from 76 observations), 2005 (91 observations), 2000 (159 observations), 1990 (157 observations) and 1983 (152 observations).[197] The resulting benchmarks are depicted as black squares in Figure 1.K.11. These benchmarks are also very close to our long-run rental yield series.

Third, we calculate rental yield benchmarks using market reports of the real estate agents association of Frankfurt ("*Frankfurter Immobilienbörse*"). These market reports contain price-rent ratios for investment buildings. The level of detail and accuracy of the reports does, however, increase considerably after 2000. We use these values to calculate net rental yields subtracting one-third for maintenance, depreciation and other costs as done in Jordà, Knoll, et al. (2019). The resulting values for the years 1970, 1980, 1990, 2000, 2010 and 2018 are depicted as red triangles in Figure 1.K.11. The resulting series is very close to our long-run series, except for the value for 1970. We use the 2018 value calculated from these reports as our alternative benchmark in the robustness section of the main paper. This value

197. The number of observations is considerably lower for 1982, hence we instead plot the value for 1983.

is nearly indistinguishable from the value calculated using the GA data and close to the value of *Numbeo.com*.

To summarize, all benchmark values we collected are very close to our original rental yield series. The only relevant difference is in 1970, when the benchmark of the reports by the Frankfurt real estate agents association deviates somewhat from our long-run series. The price-rent ratio in this report is, however, only stated as a rough estimate and subject to considerable measurement error. As we use a hedonic house price series built from a high number of observations between 1970 and 2018 and adjust our rent series using housing censuses, we think that the rental yield series calculated using the rent-price approach is more reliable compared to the benchmark in 1970. Therefore, we do not adjust our rental yield series for Frankfurt to any benchmarks.

1.K.6.4 Hamburg

House Price Series. At the time of writing, house price data from the GA Hamburg are not available to us.[198] Instead, we rely on house price data from the market reports of the German Real Estate Association and its predecessor starting in 1972. Prior to 1972, we use data from contemporary publications of Statistics Hamburg throughout. Whenever the data are available by subdivisions of the city of Hamburg, we build stratification indices to control for sample shifts in the location of sold houses. For some years in between, no house price data are available at all. We provide more details below.

In addition to the price data for apartments used in the main paper, the market reports of the German Real Estate Association and its predecessor also provide data on detached and attached single-family houses. To get a broader coverage for the city of Hamburg and to approximate price developments of the entire housing market, and in contrast to the German data set used in the main paper, we use price information for both apartments and single-family houses. We first construct a house price index for apartments and for single-family houses separately. The construction of the apartment index is described in the Appendix of the main paper. Single-family houses are separated into two sub-categories: detached single family houses (with a surrounding plot of land and a garage) and attached single-family houses (without a garage).[199] Again, we start with constructing separate indices for both sub-categories. Within these sub-categories, mode prices are given separately for three to four different quality bins. To get a constant quality index, as done for apartments, we use a chained matched-model approach and simple aver-

198. In an upcoming project, we cooperate with the GA Hamburg to digitize their archived house price data and construct hedonic house price series for Hamburg.

199. "*Freistehende Eigenheime (inkl. Garage und ortsübl. großem Grundstück)*" and "*Reihenhäuser (Mittelhaus ohne Garage)*".

ages over the non-missing quality bins.[200] To construct an overall index for single-family houses, we take a simple average over both sub-categories. Our final house price index for Hamburg between 1972 and 2018 is a simple average of the resulting single-family and apartment indices. This index accurately controls for quality changes over time, as it separates different housing categories and within these categories, is constructed using model dwellings from different quality bins. Quality bins in the original data not only incorporate size and quality of the dwelling itself, but also take the location of the dwelling into account. The main weakness of the index is that it does not adequately reproduce the quality mix within the housing stock in Hamburg, because weights for the different quality bins are missing. To assess the effect of this weakness, we compare house price indices constructed either from the market reports or from transaction-level GA data for Cologne, as for Cologne the GA data are available for both single-family houses and apartments. Figure 1.K.12 plots the results. It shows that the resulting indices for both categories using either the GA data or the market reports are similar. We therefore assume that the bias induced by the missing weights for the quality bins used in the market reports is small.

For the period from 1956 to 1970, we use average prices of developed lots from statistical yearbooks published by Statistics Hamburg.[201] The data covers all sales and voluntary auctions ("*Verkäufe und freiwillige Versteigerungen*") of developed lots within the city of Hamburg. It is given by district ("*Bezirk*") and within districts by two to five subdivisions. We use these subdivisions (22), which are the smallest non-overlapping regional units available, to stratify the data. We build Fisher-type stratification indices using the average sales price of developed lots following Eurostat (2013). This way we are able to control for sample shifts in the location of transacted dwellings. To the extent that housing types and also the size of houses are correlated with location, the fine-grained locational units additionally control for sample shifts along these dimensions. For the years 1955 and 1956, we instead had to rely on average transaction prices of developed lots for the entire city of Hamburg from Matti (1963). House price data for 1971 and before 1955 are missing. To match the stratification index to the later data, we use a market report from the predecessor of the German Real Estate Association[202] from the year 1969. We match these data to the data in 1973 using only the middle categories for each detached and attached single-family houses and apartments and averaging over these hous-

200. We use a simple average, as data on the distribution of the different bins within the housing stock are not available to us.

201. "*Statistisches Jahrbuch für die freie und Hansestadt Hamburg*" (Volumes 1957-1971).

202. The "*Ring deutscher Makler*".

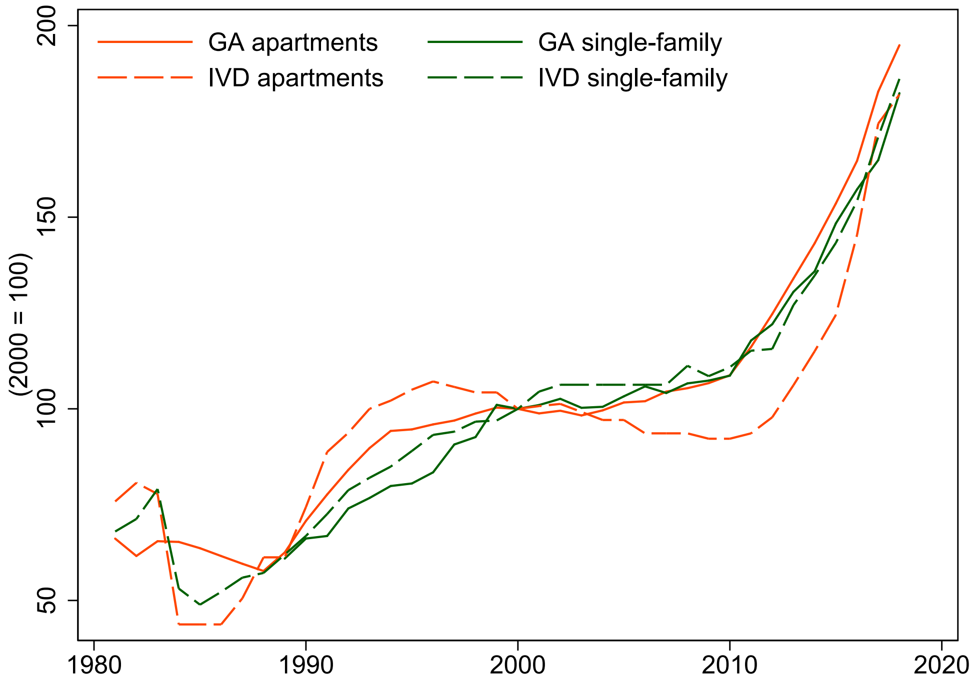

Figure 1.K.12. Nominal house price series from IVD and GA for Cologne, 2000=100

Notes: The figure shows house price indices for apartments and single-family houses in Cologne. The IVD indices are constructed from the market reports of the German Real Estate Association and its predecessor as described for Hamburg in the text. The GA indices are stratified hedonic indices from micro-level data provided by the GA Cologne as described in the text about the house price series for Cologne.

ing types as described above.[203] We linearly interpolate the house price series for 1971 to calculate housing returns.

Prior to World War II, we again use average prices of all sales of developed lots published by Statistics Hamburg.[204] For the period from 1928 to 1937, the data are provided for different subdivisions of the federal state of Hamburg. We exclude all subdivisions that did not belong to the city of Hamburg during this time period and all aggregates comprised of smaller subdivisions. From the remaining 33 subdivisions, for each pair of consecutive years, we drop all subdivisions that featured less than three sales in one of the two years. Using the remaining subdivisions, we again build Fisher-type stratification indices using the average sales price as described

203. The data in 1969 are reported in a different format. In these data, quality bins only incorporate the quality of the location. We assume that by taking only the middle quality bin by housing type, mode prices of the different reporting formats are comparable.

204. 1928 - 1937: "*Statistisches Jahrbuch für die freie und Hansestadt Hamburg*" (Volumes 1928-1938); 1903 - 1928: "*Statistisches Handbuch für den Hamburgischen Staat*" (1920) and "*Statistisches Jahrbuch für die freie und Hansestadt Hamburg*" (Volumes 1925-1928).

above. To link the data between 1937 and 1956, we match the old subdivisions from 1937 to the new ones from 1956 as accurately as possible. As the city of Hamburg grew considerably in between by incorporating surrounding cities and villages, we exclude the parts of the city in 1956 that did not already belong to the city of Hamburg in 1937.[205] After the matching, we again calculate a Fisher-type stratification index between both years. For the years 1903 to 1928, only average sales prices for developed lots within the entire former city of Hamburg are available.

For the period from 1870 to 1889, Statistics Hamburg reports average sales for publicly sold developed lots.[206] Between 1870 and 1885 the data is given by district ("*Stadtteil*"). We include all districts belonging to the former city ("*Stadt und Vorstadt*") or suburbs ("*Vororte*") of Hamburg (22), but exclude the rural districts around the city of Hamburg that belonged to the federal state ("Landgebiet"). Again we build a Fisher-type stratification index using average transaction price by district. For the years 1885 to 1889, the data is only given by the three categories city, suburbs and rural areas. We again exclude the rural areas and build a Fisher-type stratification index stratifying by the two remaining categories. No house price data is available between 1889 and 1903. We match the data from 1889 to 1903 by comparing average price per transaction for public sales in the former city of Hamburg in 1889 to price per transaction for all sales in the former city of Hamburg in 1903. This procedure is potentially biased, because public sales might be a selected sample of all sales. These are, however, no alternative data available that allows for a more precise match.

Table 1.K.21 summarizes the components of our final house price index.

Rent Series. Rent data for Hamburg are taken from various different sources. Starting in 1972, we again rely on the rent data from market reports of the German Real Estate Association and its predecessor. Prior to 1967 and during the interwar period, we use rent indices constructed by Statistics Hamburg during the process of constructing a city-level CPI index. During and prior to World War I, we use average

205. We additionally drop one subdivision of the city of Hamburg in 1956 that only covered a small part already belonging to Hamburg in 1937, but a larger part that did not belong to Hamburg before. As the city is subdivided differently in 1937 and 1956, the matching is not always perfect, so some remaining subdivisions in 1956 additionally cover small parts not belonging to Hamburg in 1937.

206. Developed lots that have been sold at a public exchange market. These encompassed normal sales as well as forced sales (excluding forced sales at the court) and represented approximately 10% of all sales of developed lots. According to Statistics Hamburg, these transactions give a good overview of price movements of all developed lots (see: "*Statistik des Hamburgischen Staats*",Volume 1886, page 176; German definition: "*In der Börse öffentlich verkaufte[..] Grundstücke und zwar sowohl freihändige Verkäufe wie auch die Zwangsverkäufe abseiten des Amtgerichtes Hamburg, desgleich vom Jahre 1883 an die abseiten der Amtsgerichte Bergedorf und Ritzebüttel (umfassend die Landherrenschaften gleichen Names) daselbst öffentlich verkauften Grundtsücke.*"; Source: "*Statistik des Hamburgischen Staats*" (Volumes 1880 & 1886) and "*Statistisches Handbuch für den Hamburgischen Staat*" (1891).

Table 1.K.21. Final house price index for Hamburg

PERIOD	SOURCE	DESCRIPTION
1870-1885	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All public sales aggregated by district from statistical publications; *Method*: Stratification.
1885-1889	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All public sales aggregated by city or suburbs from statistical publications; *Method*: Stratification.
1903-1928	Statistics Hamburg	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales within the former city of Hamburg; *Method*: Average price.
1928-1937	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated by subdivision from statistical publications; *Method*: Stratification.
1955-1956	Matti (1963)	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales and voluntary auctions within the city of Hamburg; *Method*: Average price.
1956-1970	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales and voluntary auctions aggregated by subdivision from statistical publications; *Method*: Stratification.
1972-2018	Own compilation	*Type(s) of dwellings*: Single-family houses and apartments; *Type of data*: Mode prices by housing quality bins reported by local real estate agents; *Method*: Matched model approach using mode prices by housing types and quality bins.

rental data from statistical publications and from Wischermann (1983). No yearly rent data exist for Hamburg between 1967 and 1971. To link the index ending in 1966 to the data starting in 1972 and to get additional data for the year 1968, we use data from housing censuses. We provide more details on the sources and construction of the indices below.

The data collected during the housing censuses are taken from various publications by *Statistics Hamburg*.[207] They cover the universe of all rented residential dwellings within Hamburg and provide average rents disaggregated by construction period and district ("*Bezirk*"). We use housing census data for the years 1956, 1968 and 1987. The data for 1956 contain rents per dwelling additionally disaggregated by number of rooms. For 1968, instead, Statistics Hamburg only published rents per square meter by district and building period, but number of rooms and rents per

207. 1956: "*Statistik des Hamburgischen Staats*" (1958); 1968: "*Hamburg in Zahlen*" (1970, "*Sonderheft*" 2); 1987: "*Statistik des Hamburgischen Staats*" (1992).

dwelling are missing. To match the data from different years as precisely as possible, we match the data in both 1956 and 1968 to the data in 1987. These data comprise both rents per dwelling by number of rooms as well as rents per square meter by district and construction period. We calculate an index relative to 1987 for both 1956 and 1968 as described in the next paragraph and infer the price increase between 1956 and 1968 from these two indices. As the city of Hamburg did not grow spatially between 1956 and 1987 and we are able to control for construction period and location, we think that this procedure measures the total rent development in Hamburg very accurately.

To calculate the index between 1956 and 1987, we only use dwellings constructed prior to 1948, such that we are able to exclude all dwellings built after 1956.[208] Still, we are able to stratify the data by two distinct construction periods.[209] Additionally, we stratify the data by district ("*Bezirk*") and number of rooms (2-4).[210] We use rent per dwelling for each stratum to build a Fisher-type stratification index following Eurostat (2013) for the years 1956 and 1987. The index for 1968 and 1987 is calculated using average rent per square meter stratified by district and construction period. Again, we exclude all dwellings built after 1948 and use the same construction period and district bins to calculate a Fisher-type stratification index.

We use the indices from the housing census data to link the yearly index ending in 1966 and the index starting in 1972 and to get an index value for 1968. Starting in 1972, the yearly index uses the rent data from the market reports of the German Real Estate Association and its predecessor. The construction for rent indices from these market reports is described in the Appendix of the main paper. For the period from 1950 to 1966, we use a rent index calculated by Statistics Hamburg collected from statistical yearbooks.[211] This index was calculated to construct a city-level CPI index and is intended to track rental costs for a four-person employee household ("*4-Personen-Arbeitnehmer-Haushalt*") with a medium consumption basket ("*Mittlere Verbrauchsgruppe*"). It is given in 1938 values, such that the link to the index used prior to World War II is straightforward. As we were not able to find any rent data for Hamburg for the years 1967, 1969, 1970 and 1971, we linearly interpolate the index for these years to calculate housing returns.

Prior to World War II, we again use a rent index calculated by Statistics Hamburg for CPI construction.[212] It covers the rent of apartments with two rooms and

208. The data for dwellings built between 1948 and 1956 and for dwellings built after 1956 are pooled in the 1987 census.

209. All dwellings built prior to 1918 and dwellings built between 1918 and 1948.

210. We exclude all dwellings with only one room, as this data are missing for 1956, and with more than four rooms, because these data are pooled in the 1987 census. We still cover more than 80% of all dwellings within the relevant construction period bins from the 1987 census.

211. "*Statistisches Jahrbuch für die freie und Hansestadt Hamburg*" (Volumes 1952-1966/67).

212. Source: "*Statistisches Jahrbuch für die freie und Hansestadt Hamburg*" (Volumes 1926/27-1937/38)

a kitchen and excludes heating and lighting costs. We use yearly averages of the monthly index.[213] For the earlier years, the index is given in 1914 values, so we can directly link the index to the earlier data.

Between 1870 and 1918, average rent for all occupied rented residential dwellings is given in Wischermann (1983) and a publication by *Statistics Hamburg*.[214] We exclude the suburbs ("*Vororte*") only incorporated into Hamburg in 1913 and calculate an index from the average rent for the former city of Hamburg. Between 1875 and 1890, these data are only given in five-year steps and additionally in 1867. We interpolate the missing years by using the average rent for all rented dwellings (residential and commercial, "*Gelasse*") also given in Wischermann (1983).[215] No data are available between 1918 and 1923.

Table 1.K.22 summarizes the components of our final rent index.

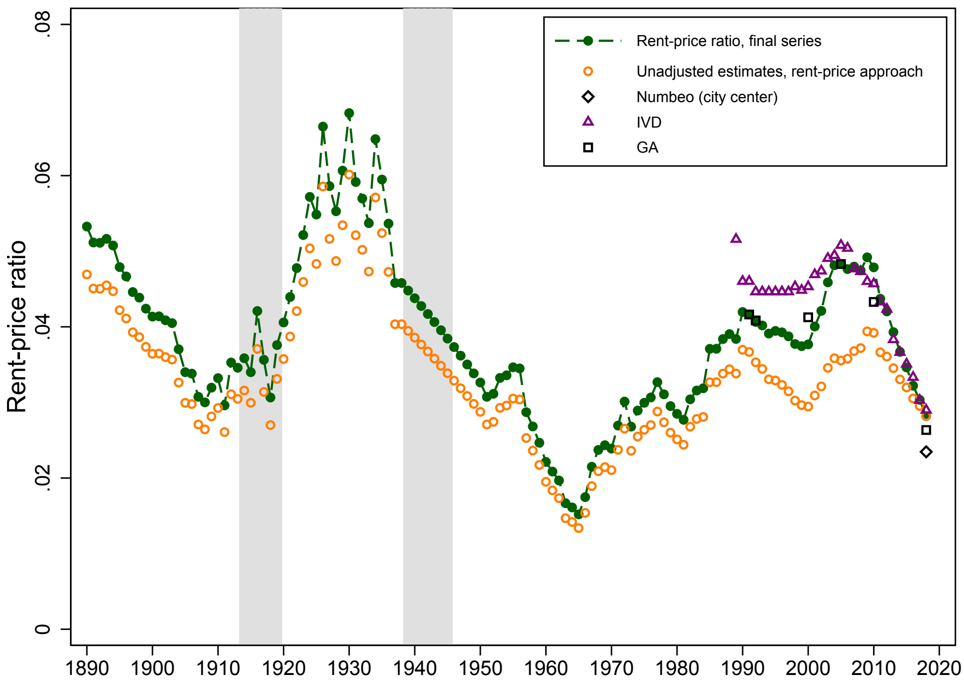

Figure 1.K.13. Hamburg: plausibility of rental yields

213. For 1923, we instead use the end of year value, because only the December value is given in the yearbooks.

214. We use Wischermann (1983) until 1913 and afterwards rent data from "*Statistisches Handbuch für den Hamburgischen Staat*" (1920). The data in Wischermann (1983) is also originally collected from statistical publications.

215. Between 1867 and 1873, we have to rely on linear interpolation instead, because other data are missing.

Table 1.K.22. Final rent index for Hamburg

PERIOD	SOURCE	DESCRIPTION
1956, 1968, 1987	Own compilation	*Type(s) of dwellings*: All rented residential dwellings; *Type of data*: Census data from statistical publications; *Method*: Stratification - we use these data to link the yearly indices as described in the main text.
1870-1913	Wischermann (1983)	*Type(s) of dwellings*: All rented residential dwellings; *Type of data*: Averages from statistical publications interpolated using data for all dwellings between 1870 and 1890; *Method*: Average rent.
1913-1918	Own compilation	*Type(s) of dwellings*: All rented residential dwellings; *Type of data*: Averages from statistical publications; *Method*: Average rent.
1923-1938	Statistics Hamburg	*Type(s) of dwellings*: Apartments with two rooms and a kitchen; *Type of data*: From CPI-construction; *Method*: CPI rent index.
1950-1966	Statistics Hamburg	*Type(s) of dwellings*: Apartments for a four-person employee household; *Type of data*: From CPI-construction; *Method*: CPI rent index.
1972-2018	Own compilation	*Type(s) of dwellings*: Apartments; *Type of data*: New contract rents by construction period and housing quality reported by local real estate agents; *Method*: Matched model approach using mode prices by quality and construction period bins.

Rental Yield Series. The main benchmark for Hamburg is taken from *MSCI*, as described in the main paper. This benchmark is close to the alternative benchmarks we collected for 2018. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.13.

To verify the historical rental yield series, we collected alternative benchmarks from two different sources. First, we use benchmarks from the market reports of the *German Real Estate Association* (IVD). These market reports directly report mode price-rent ratios for residential investment buildings from 1989 onward. The Appendix in the main paper describes how we calculate net rental yields. The resulting rental yield benchmarks are plotted as purple triangles in Figure 1.K.13. This series is above our unadjusted long-run series and decreases faster after 2005, the year with the maximum value after 1990.

Second, in their yearly market reports, the GA Hamburg publishes average price-rent ratios for investment buildings.[216] We use the average values for all buildings for 2005, 2010 and 2018 from the reports. For 1991, 1992 and 2000, we calculate averages for all buildings over construction period and location bins weighted by the number of observations in each bin. Following Jordà, Knoll, et al. (2019) we subtract one-third of the resulting rent-price ratios to get net rental yields. We use the 2018 benchmark as our alternative benchmark in the robustness section of the main paper, because it relies on micro-level data of actual transaction prices from a large number of observations (182). The resulting benchmarks for all mentioned years are depicted as black squares in Figure 1.K.13. These benchmarks are also above our unadjusted series for all years prior to 2018 and depict a larger fall after 2005. As the Hamburg house price series for the last decades, in contrast to the other German cities, is not constructed from micro-level data, but from the real estate agents' market reports, and because the rental yield benchmarks from the two independent sources depict a similar pattern, we believe that these benchmarks are more reliable in comparison to our long-run series. We therefore adjust our long-run series using the GA benchmarks in 2005, the height of the benchmarks after 1990, and in 1991. This gives us the final rental yield series plotted as the green-circled line in Figure 1.K.13. This series is also very close to the GA benchmarks in 2010, 2000 and 1991 and much closer to the IVD series compared to the unadjusted long-run series.

1.K.7 Italy

Italy is a country with a long urban tradition and by 1900 it hosted many large cities. The largest cities were Naples, Milan, Rome and Turin and each had one percent or more of the total population and therefore entered our sample. In absolute terms, Genova, Palermo or Florence also counted large populations of more than 200,000, but failed to meet the 1% threshold. For house prices, we follow the sources of the seminal publication of the Bank of Italy (Cannari, D'Alessio, and Vecchi (2016)) which published a national house price series. We mainly follow its source trail to break numbers down to the regional level. For the early years between 1927 and 1941, *Federazione Fascista dei Costruttori Edili* in principle reports city-level house prices, which however we exclude here because of a data gap for the wartime years. For rents, we largely follow the regional break-down of the national CPI series.

Overall, we think that the quality of the house price series for the Italian cities for the period between 1950 and 1966 is not very high. We were not able to find data to build alternative series, with the exception of Milan. Nevertheless, since the patterns we observe in terms of the price evolution across the Italian cities in our

216. "*Verhältnis des Kaufpreises zur Jahresnettokaltmiete*".

sample and the national series from Knoll, Schularick, and Steger (2017) are very similar, we decided to keep the series in our final data set.

1.K.7.1 Milan

House Price Series. House prices in Milan are taken from a combination of different sources (Table 1.K.23). For the time period 1950 to 1955, we start from the building plot data from Cannari, D'Alessio, and Vecchi (2016). As these exaggerate price movements and ignore building costs, we factor in regional construction costs, taken from Forte and Di Stefano (1970) and Italy's statistical yearbook ("*Anuario Statistico Italiano*"). Following Jordà, Knoll, et al. (2019), we use a construction cost weight of 20 percent. Due to missing regional data, we have to bridge the gap years until 1955 using the national deflator on housing investments from Cannari, D'Alessio, and Vecchi (2016).

We then constructed our own series based on newspaper ads for the period between 1955 and 1966. As the number of ads was insufficient in the early 1950s, we have to start our series in 1955. We use the time dummy hedonic log-linear regression based on newspaper advertisements for flats in Milan taken from *La Notte* and from *Il Corriere della Sera*. We collected a total of 641 ads for the period between 1956 and 1966, which contained information on the price, location and size of the flats. Log prices are regressed on the number of rooms, on a dummy for location for a total of nine neighborhoods in Milan and on dummies for whether the flat has: a bathroom, a kitchen, a heating system, a balcony, a garden, a garage or furniture. Since the two newspapers do not cover the complete period, we first do a regression using the data from both newspapers between 1956 and 1962 and then do a regression using only the ads from *Il Corriere* for the years between 1961 and 1966. Finally, we splice the two indices to build the final index for the years between 1956 and 1966.

From 1966 onward until the current day, we rely on the commercial company *Nomisma's* data from its *Real Estate Market Observatory* which collects data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper *Sole*. The data allow a break down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within sub-segments both minimum and maximum price ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take simple averages of annual square meter prices reported in Euro across all urban market segments. Table 1.K.23 gives an overview of the data used.

Rent Series. Data for rent series are taken from official national statistical sources which report on rent as component of the CPI in regional break downs. We make use of multiple editions of yearbooks to connect the indices reported with varying base years over time. For the early period, we draw on the CPI-series as reported in

Table 1.K.23. Final house price index for Milan

PERIOD	SOURCE	DESCRIPTION
1950-1955	Cannari, D'Alessio, and Vecchi (2016)	*Type(s) of dwellings*: New dwellings; *Type of data*: Official price and cost statistics; *Method*: Weighted indices of all urban centers
1955-1966	Own compilation	*Type(s) of dwellings*: Flats; *Type of data*: Newspaper ads from *La Notte* and *Il Corriere*; *Method*: Time-dummy hedonic index
1966-2018	*Nomisma*	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average price of house by quality and geographic housing market segment

the city statistical yearbooks and connect it to the indices reported in the national statistical yearbook from 1950 to 1965. The index is based on household surveys on 3-4-room apartments of working-class families in the city.

From 1966 onward, we rely on the commercial company *Nomisma's* data from its *Real Estate Market Observatory* which has been collecting data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper *Sole*. The data allow a break-down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within subsegments both minimum and maximum rent ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take simple averages of annual square meter rents reported in euros across all urban market segments. Table 1.K.24 gives an overview of the data and methods used.

Table 1.K.24. Final rent index for Milan

PERIOD	SOURCE	DESCRIPTION
1950-1966	*Instituto Centrale di Statistica: Annuario Statistico Italiano*	*Type(s) of dwellings*: 3-4 room working class apartments; *Type of data*: Household survey; *Method*: Index based on CPI-rent component
1966-2018	*Nomisma*	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys ; *Method*: Average rent of house by quality and geographic housing market segment

Rental Yield Series. Our main benchmark for Milan is taken from *MSCI*, as described in the main part of the paper. This benchmark is reasonably close to, but below the alternative benchmarks we collected for 2018 from *Numbeo.com* (which measures city-center yields), to which we subtract one-third costs following Jordà, Knoll, et al. (2019). Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.14. The series shows a rise in rental yields with the gradual lifting of wartime rent controls and a decline ever since 1970, with a short boom in the 1980s. This is the pattern that we find in all Italian cities in our sample. As a result and, since we did not find historical rental yield benchmarks, we decided to keep the unadjusted series as our final rental yield series.

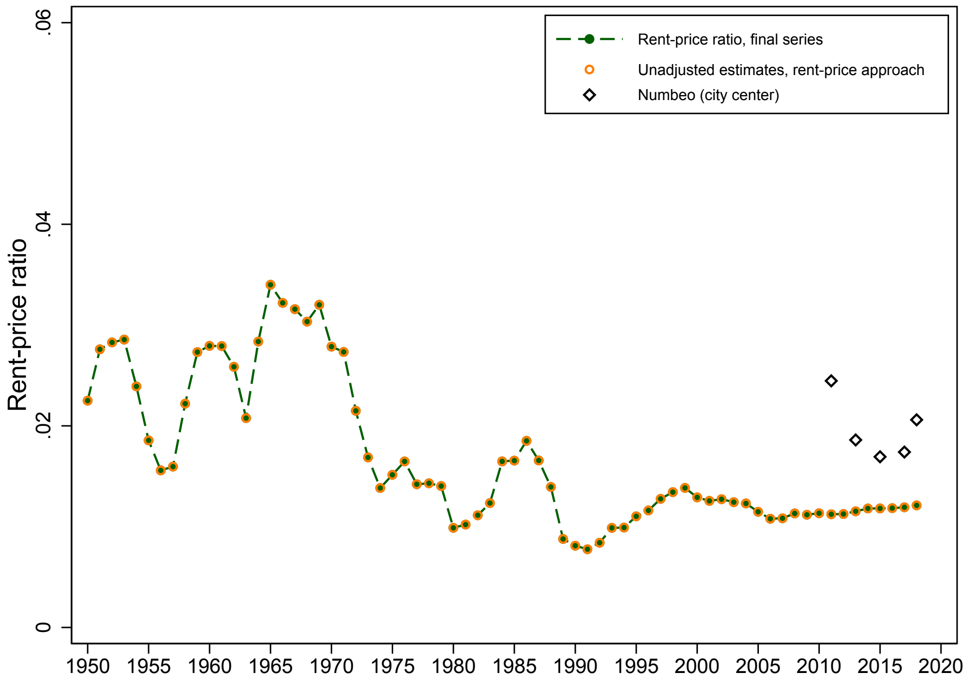

Figure 1.K.14. Milan: plausibility of rental yields

1.K.7.2 Naples

House Price Series. House prices in Naples are taken from a combination of different sources (Table 1.K.25). For the time period 1950 to 1960, we start from the building plot data from Cannari, D'Alessio, and Vecchi (2016). As these exaggerate price movements and ignore building costs, we factor in regional construction costs, taken from Forte and Di Stefano (1970) and Italy's statistical yearbook ("*Anuario Statistico Italiano*"). Following Jordà, Knoll, et al. (2019), we use a construction cost weight of 20 percent. Due to missing regional data, we have to bridge the gap years until 1965 using the national deflator on housing investments from Cannari, D'Alessio, and Vecchi (2016). We have to fall back to this interpolation as we could not identity an accessible newspaper with decent reporting on housing ads as in the case of Milan to build a series of our own.

From 1966 onward until the current day, we rely on the commercial company *Nomisma's* data from its *Real Estate Market Observatory* which has been collecting data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper *Sole*. The data allow a break-down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within sub-segments both minimum and maximum price ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take a simple averages of annual square meter prices reported in euros across all urban market segments.

Table 1.K.25. Final house price index for Naples

PERIOD	SOURCE	DESCRIPTION
1950-1960	Cannari, D'Alessio, and Vecchi (2016)	*Type(s) of dwellings*: New dwellings; *Type of data*: Official price and cost statistics ; *Method*: Weighted indices
1960-1966	*Banca d'Italia*	*Type(s) of dwellings*: New dwellings; *Type of data*: Housing investment deflator; *Method*: Index
1966-2018	*Nomisma*	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average price of house by quality and geographic housing market segment

Rent Series. Data for rent series are taken from official national statistical sources which report on rent as component of the CPI in regional break-downs. We make use of multiple editions of yearbooks to connect the indices reported with varying base years over time. For the early period, we draw on the CPI-series as reported in the city statistical yearbooks and connect it to the indices reported in the national statistical yearbook from 1950 to 1965. The index is based on household surveys on 3-4-room apartments of working-class families in the city.

From 1966 onward, we rely on the commercial company *Nomisma's* data from its *Real Estate Market Observatory* which has been collecting data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper *Sole*. The data allow a break down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within sub-segments both minimum and maximum rent ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take a simple averages of annual square meter rents reported in euros across all urban market segments. Table 1.K.26 gives an overview of the data and methods used.

Table 1.K.26. Final rent index for Naples

PERIOD	SOURCE	DESCRIPTION
1950-1966	*Instituto Centrale di Statistica: Annuario Statistico Italiano*	*Type(s) of dwellings*: 3-4-room working class apartments; *Type of data*: Household survey; *Method*: Index based on CPI-rent component
1966-2018	*Nomisma*	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average rent of house by quality and geographic housing market segment

Rental Yield Series. Our main benchmark for Naples is taken from *Numbeo.com* for 2018, since *MSCI* does not produce estimates for Naples. Following Jordà, Knoll, et al. (2019) we adjust the estimates for one-third costs. Additionally, we collected benchmarks for recent years also from *Numbeo.com*. The absolute levels of recent yields are relatively low. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.15. The series shows a rise in rental yields with the gradual lifting of wartime rent controls and a decline ever since 1970, with a short boom in the 1980s. This pattern is consistent with what we find for the other Italian cities. As a result and, since we did not find historical rental yield benchmarks, we decided to keep the unadjusted series as our final rental yield series.

1.K.7.3 Rome

House Price Series. House prices in Rome are taken from a combination of different sources (Table 1.K.27). For the time period 1950 to 1960, we start from the building plot data from Cannari, D'Alessio, and Vecchi (2016). As these exaggerate price movements and ignore building costs, we factor in regional construction costs, taken from Forte and Di Stefano (1970) and Italy's statistical yearbook (*Annuario statistico Italiano*). Following Jordà, Knoll, et al. (2019), we use a construction cost weight of 20 percent. Due to missing regional data, we have to bridge the gap years until 1965 using the national deflator on housing investments from Cannari, D'Alessio, and Vecchi (2016). We have to fall back to this interpolation, as we could not identity an accessible newspaper with decent reporting on housing ads as in the case of Milan to build a series of our own.

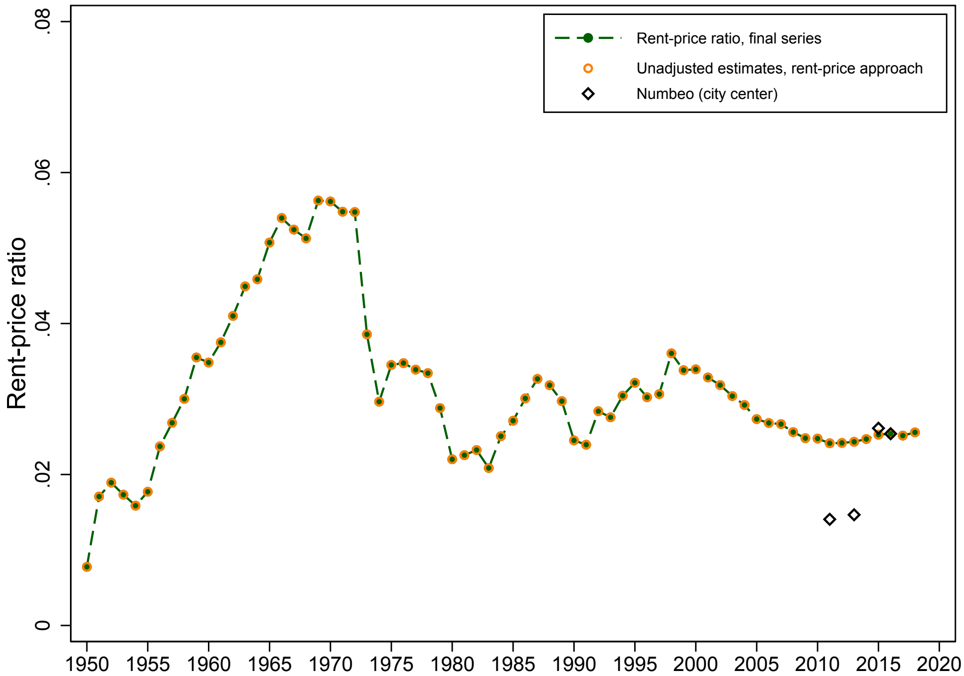

Figure 1.K.15. Naples: plausibility of rental yields

From 1966 onward until the current day, we rely on the commercial company Nomisma's data from its *Real Estate Market Observatory* which has been collecting data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper *Sole*. The data allow a break-down into new and old housing stock and into different geographic sub-segments (Center, semi-center, periphery). Within sub-segments both minimum and maximum price ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take a simple averages of annual square meter prices reported in euros across all urban market segments.

Rent Series. Data for rent series are taken from official local and national statistical sources which report on rent as a component of the CPI in regional break downs. We make use of multiple editions of yearbooks to connect the indices reported with varying base years over time. For the early period, we draw on the CPI-series as reported in the city statistical yearbooks and connect it to the indices reported in the national statistical yearbook from 1950 to 1965. The index is based on household surveys on 3-4-room apartments of working-class families in the city.

From 1966 onward, we rely on data commercial company *Nomisma's* data from its *Real Estate Market Observatory* which collects data on different segments of the

Table 1.K.27. Final house price index for Rome

PERIOD	SOURCE	DESCRIPTION
1950-1960	Cannari, D'Alessio, and Vecchi (2016)	*Type(s) of dwellings*: New dwellings; *Type of data*: Official price and cost statistics; *Method*: Weighted indices
1960-1966	Banca d'Italia	*Type(s) of dwellings*: New dwellings; *Type of data*: Housing investment deflator; *Method*: Index
1966-2018	Nomisma	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average price of house by quality and geographic housing market segment

residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper Sole. The data allow a break-down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within sub-segments both minimum and maximum rent ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take simple averages of annual square meter rents reported in euros across all urban market segments.

Table 1.K.28. Final rent index for Rome

PERIOD	SOURCE	DESCRIPTION
1950-1951	*Annuario Statistico di Roma*	*Type(s) of dwellings*: 3-4-room working class apartments; *Type of data*: Household survey; *Method*: Index based on CPI-rent component
1951-1966	*Instituto Centrale di Statistica: Annuario Statistico Italiano*	*Type(s) of dwellings*: 3-4-room working class apartments; *Type of data*: Household survey; *Method*: Index based on CPI-rent component
1966-2018	*Nomisma*	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average rent of house by quality and geographic housing market segment

Rental Yield Series. Our main benchmark for Rome is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to the alternative benchmark we collected for 2018 from *Numbeo.com* (which measure city center yields), for which we assume one-third costs as in Jordà, Knoll, et al. (2019). Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.16. The series shows a rise in rental yields with the gradual lifting of wartime rent controls and a decline ever since 1970, with a short boom in the 1980s. Again, this is the same pattern as in the other Italian cities, although in the case of Rome it seems to be less pronounced. As a result and, since we did not find historical rental yield benchmarks, we decided to keep the unadjusted series as our final rental yield series.

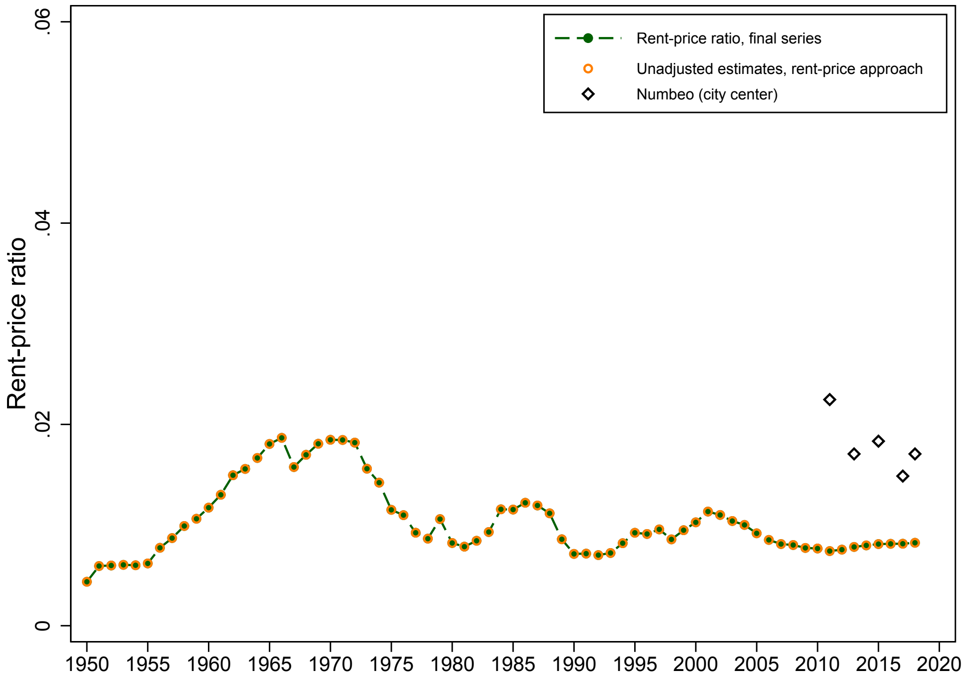

Figure 1.K.16. Rome: plausibility of rental yields

1.K.7.4 Turin

House Price Series. House prices in Turin are taken from a combination of different sources (Table 1.K.29). Lacking city-specific sources as used in the cases of other Italian cities, we have to fall back on the big-city average taken from Cannari, D'Alessio, and Vecchi (2016), as we could not identity an accessible newspaper with decent reporting on housing ads as in the case of Milan to build a series of our own.

From 1966 onward until the current day, we rely on the commercial company *Nomisma's* data from its Real Estate Market Observatory which has been collecting data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente*

Immobiliare and the newspaper *Sole*. The data allow a break down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within sub-segments both minimum and maximum price ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take simple averages of annual square meter prices reported in euros across all urban market segments.

Table 1.K.29. Final house price index for Turin

PERIOD	SOURCE	DESCRIPTION
1950-1966	Cannari, D'Alessio, and Vecchi (2016)	*Type(s) of dwellings*: New dwellings; *Type of data*: Official price and cost statistics ; *Method*: Weighted indices of all urban centers
1966-2018	Nomisma	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average price of house by quality and geographic housing market segment

Rent Series. Data for rent series are taken from official national statistical sources which report on rent as a component of the CPI in regional break downs. We make use of multiple editions of the national yearbooks ("*Annuario Statistico Italiano*") to connect the indices reported with varying base years over time. We use these data to build an index from 1950 to 1965. The index is based on household surveys on 3-4-room apartments of working-class families in the city.

From 1966 onward, we rely on the commercial company *Nomisma's* data from its *Real Estate Market Observatory* which collects data on different segments of the residential real estate market itself since 1988. For the period 1966 to 1987, it relied on pre-existing data mainly from *Consulente Immobiliare* and the newspaper *Sole*. The data allow a break-down into new and old housing stock and into different geographic sub-segments (center, semi-center, periphery). Within sub-segments both minimum and maximum rent ranges are reported. As there are no data on the prevalence of these segments in the overall urban housing stock, we take simple averages of annual square meter rents reported in euros across all urban market segments. Table 1.K.30 gives an overview of the data and methods used.

Rental Yield Series. Our main benchmark for Turin is taken from *Numbeo.com*, as *MSCI* does not provide estimates for Turin. Following Jordà, Knoll, et al. (2019) we assume one-third costs to have an estimate of the net yield. This benchmark is reasonably close to the alternative benchmarks we collected for 2018 from the *Deloitte Property Index*. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in in Figure 1.K.17. The series shows a rise in rental yields with the gradual lifting

Table 1.K.30. Rents in Turin

PERIOD	SOURCE	DESCRIPTION
1950-1966	*Instituto Centrale di Statistica: Annuario Statistico Italiano*	*Type(s) of dwellings*: 3-4-room working class apartments ; *Type of data*: Household survey; *Method*: Index based on CPI-rent component
1966-2018	*Nomisma*	*Type(s) of dwellings*: Old and new dwellings; *Type of data*: Commercial market surveys; *Method*: Average rent of house by quality and geographic housing market segment

of wartime rent controls and a decline ever since 1970, with a short boom in the 1980s. This is the same pattern we find for the other cities. As a result and, since we did not find historical rental yield benchmarks, we decided to keep the unadjusted series as our final rental yield series.

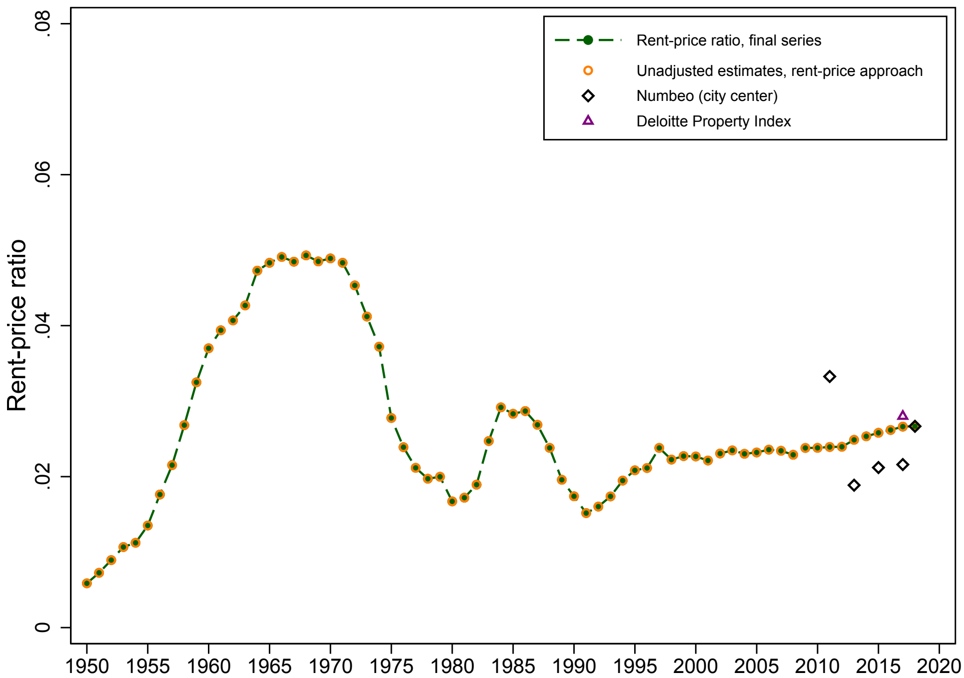

Figure 1.K.17. Turin: plausibility of rental yields

1.K.8 Japan

In 1900 the largest four cities in Japan were in descending order: Tokyo, Osaka, Kyoto and Nagoya. Of these four cities, only Tokyo and Osaka represented more than 1% of the national population in 1900. Tokyo was by far the largest city in Japan with more than two times the population of Osaka and represented around 3.5% of the national population.

To the best of our knowledge there do not exist regional quality-adjusted house price series for Japan for the pre-1975 period.[217] LaPoint (2020) has assembled a very extensive data set containing house price indices for 300 different cities in Japan for the period between 1975 and 2015. Unfortunately, rent data and housing return data are missing for most of these cities; sufficient rent data is only available for Tokyo. Additionally, there has been very exhaustive work on the evolution of prices in Tokyo around the housing bubble at the end of the 1980s.[218] Hence, we only include Tokyo in our long-run data set.

As we explain in more detail below, our main contribution to the history of housing market developments in Japan is the construction of a brand-new hedonic house price index for the period between 1950 and 1975.

1.K.8.1 Tokyo

House Price Series. Shimizu and Nishimura (2007) built a hedonic land price index for the Setagaya Ward in Tokyo for the period between 1975 and 1999. The authors collected a total of 7,991 residential land transaction prices in the Setagaya Ward, which is a famous and expensive residential neighborhood in Tokyo, from the local land registry offices. They used these data to build a time dummy hedonic index, in which they regressed the log price of land on the following set of controls: land lot size, width of road frontage, distance to nearest station, time to Central Business District, floor-to-lot ratio and a railway dummy factor.

Hill, Scholz, Shimizu, and Steurer (2018) built a hedonic house price index for the city of Tokyo for the period between 1986 and 2016 using asking prices of apartments collected from the *Residential Information Weekly* (*Shukan Jyutako Joho*). In total the authors use 237,190 observations, which are spread around the 23 special wards of the prefecture of Tokyo. The index is a rolling time dummy index with a four-quarter rolling window which controls for the following variables: log of floor area, age, time to nearest station, time to Tokyo Central Station and a ward dummy.

LaPoint (2020) built a residential price index for Tokyo from 1975 to 2015 using appraisals data from official real estate appraisers. The author built an extensive

217. Bank of Japan (1986) published house price indices for the six largest metropolitan areas in Japan for the periods 1913-1930 and 1936-1965, based on average residential land appraised values, i.e. without any quality adjustments.

218. See for example the work by Shimizu, Nishimura, and Watanabe (2016)

data set containing almost the universe of land plot appraisals in Japan since 1975. A land plot appraisal takes into account not only the land, but also the building on top of it. Using residential land plots, which are appraised on a yearly basis, the author builds a "repeat appraisal" index in which he regresses the log appraisal value on a time dummy and a land plot fixed effect, which controls for all time-invariant observed and unobserved characteristics of the land or buildings existing on the plot. One concern with this series is the fact that it is based on appraisals and not on actual transaction prices. However, the author shows that the repeat appraisal indices he built correlate very strongly with actual repeat sales indices for overlapping periods.

The Japan Real Estate Institute (JREI) publishes a house price index for Tokyo based on existing apartment sales in the 23 special wards of Tokyo on a monthly basis since 1993. The so-called Fudoken Housing Price Index for the prefecture of Tokyo is based on transaction prices of existing apartments from the official database of real estate registry in Japan, the Real Estate Information Network System (REINS) (Institute, 2021).[219] The index is a weighted least squares repeat sales index.

Since there were no quality-adjusted price series available for the period pre-1975, we built a hedonic price index based on newspaper asking prices. We collected a total of 3,766 observations on single-family houses for the period between 1950 and 1975 from the real estate ad section of the newspaper *Yomiuri*.[220] We built a rolling time dummy index with a five-year rolling window in which we regress the log asking price on the log square meters of the house and dummy, which divides the city into an expensive and a cheap part. To do so, we follow the classification from Diewert and Shimizu (2015) and define the wards 1-4, 7-11 and 13-14 as the expensive part of Tokyo, and the wards 5-6, 12 and 15-21 as the cheap part of the city.

Table 1.K.31 summarizes the components of our final house price index. For the period between 1950 and 1975 we use our own index, since it is the only quality-adjusted house price index for the pre-1975 period. From 1975 to 1986 we rely on the repeat-appraisals index from LaPoint (2020), since it covers the whole city of Tokyo and not only the Setagaya ward. From 1986 to 1993 we opted to use the index by Hill et al. (2018), which is based on a very extensive and representative set of asking prices. This is the period of the famous housing boom in Japan, which peaked in 1989. Shimizu, Nishimura, and Watanabe (2016) make an extensive comparison of repeat sales and hedonic indices around this period and find that the turning points in the repeat-sales indices usually lag behind the ones in the hedonic indices.[221] As such, we chose to use the index by Hill et al. (2018) instead of the index by LaPoint

219. Newly constructed apartments are not considered for the index.

220. We are very grateful to Masashi Tanigaki, without whom we would not have been able to build this index.

221. They attribute these differences to the non-randomness in the repeat sales sample.

(2020) for this period. For post-1993 we can rely on the official index from JREI, which is based on the universe of transaction prices in Tokyo.

Table 1.K.31. Final house price index for Tokyo

PERIOD	SOURCE	DESCRIPTION
1950-1975	Own compilation	*Type(s) of dwellings*: Single-family houses; *Type of data*: Asking prices from the newspaper *Yomiuri*; *Method*: Five-year rolling window hedonic index.
1975-1986	LaPoint (2020)	*Type(s) of dwellings*: Residential land plots; *Type of data*: Appraisals from official real estate appraisers; *Method*: Repeat appraisal index.
1986-1993	Hill et al. (2018)	*Type(s) of dwellings*: Apartments; *Type of data*: Asking prices from the *Residential Information Weekly*; *Method*: Four-quarter rolling-window hedonic index.
1993-2018	Institute (2021)	*Type(s) of dwellings*: Existing apartments; *Type of data*: Transaction prices from the Real Estate Information Network System data set; *Method*: Weighted least squares repeat sales index.

Rent Series. The *Tokyo Statistical Yearbook* has published a yearly rent index for Tokyo since 1947, which is used to build the official CPI series for Tokyo. The index is based on data from the survey on family consumption and expenditures, which is conducted by the *Statistics Bureau of Japan*. The index is based on a matched-model approach, according to which the rent price change over time is calculated based on rents for the same dwellings.[222] Since the classification of rent indices changed over time in the *Tokyo Statistical Yearbook*, we had to splice different series. From 1950 to 1970, we use the index on house and land rent and from 1971 onward we rely on the rent index. As an alternative, we could have used a housing index, which also includes imputed rents. However, in order to be consistent with the national rent index for Japan from Jordà, Knoll, et al. (2019), we chose to use the pure rent index. Also, to make the Tokyo rent series comparable to the national rent index from Jordà, Knoll, et al. (2019), we make the assumption that the rent series understated the growth in rents by a factor of 2 between 1960 and 1969 as well.

Table 1.K.32 summarizes the components of our final rent index. To the best of our knowledge the rent index by the Statistics Bureau of Japan is the only existing long-run quality-adjusted index for Tokyo. As such we use it for the complete period between 1950 and 2018.

222. Shimizu, Nishimura, and Watanabe (2010) give a good overview of the methodology employed by the *Statistics Bureau of Japan*.

Table 1.K.32. Final rent index for Tokyo

PERIOD	SOURCE	DESCRIPTION
1950-2018	*Tokyo Statistical Yearbook* (various years)	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Rent prices from the survey on family consumption and expenditures; *Method*: Matched-model index.

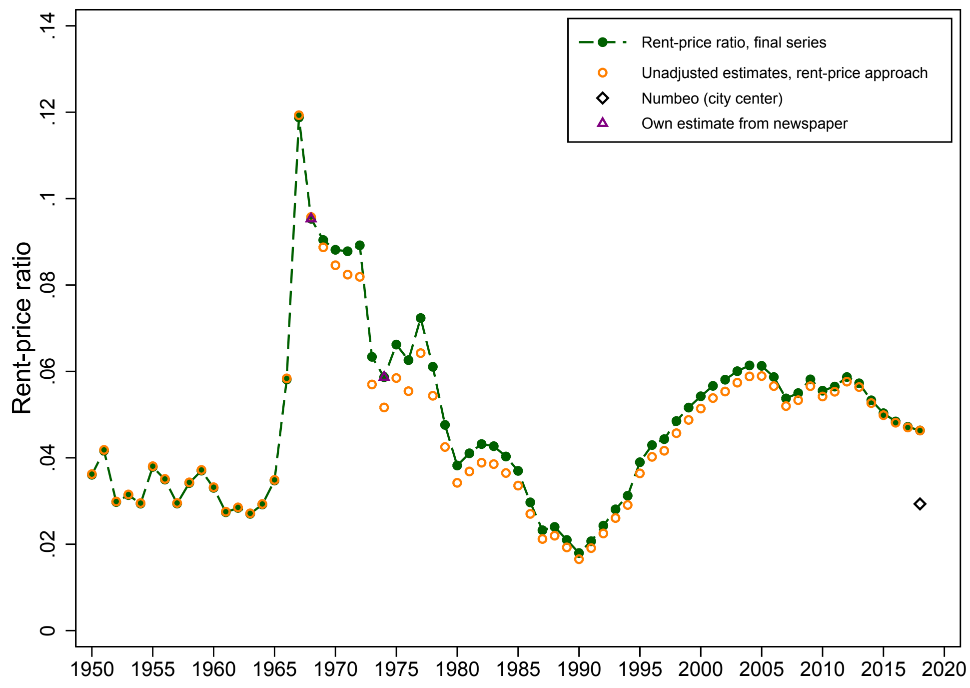

Figure 1.K.18. Tokyo: plausibility of rental yields

Rental Yield Series. Our main benchmark for Tokyo is taken from *MSCI*, as described in the main paper. This benchmark is slightly above the benchmark we collected for 2018 from *Numbeo.com*. According to *Numbeo.com* the gross rental yield in the city center of Tokyo was 4.4% in 2018; adjusting for one-third costs we estimate a net rental yield of 2.9% for 2018. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.18.

Additionally, we collected rental yield benchmarks from the newspaper *Yomiuri* using real estate ads which contained both asking price and gross rental income for the same apartment. In total we collected 110 observations for the years 1968 and 1974 for apartments in the city of Tokyo. For both years we use the average rental yield as benchmark.[223] For 1968 we estimate a gross rental yield of 14.3% and for 1974 8.8%. We then adjust the gross yields for one-third costs, which results in the net yield estimate in Figure 1.K.18. Since due to the high number of observations we think that these estimates are a good approximation of the actual yields in Tokyo at the time, we use them to benchmark our final yield series. This gives us the adjusted final rental yield series—the green-circled line in Figure 1.K.18.

1.K.9 Netherlands

In 1900 the three largest cities in the Netherlands were in descending order: Amsterdam, Rotterdam and The Hague. All of these cities represented more than 1% of the total national population in 1900.[224] There exist very good series on Amsterdam, not only for prices and rents, but also for housing returns, which we explain in more detail below. To the best of our knowledge there do not exist historical house price or rent series for the other Dutch cities. Since Amsterdam alone represented 10% of the total population, we did not build indices for the other cities.

1.K.9.1 Amsterdam

House Price Series. Francke and Korevaar (2019) constructed a long-run house price index for Amsterdam for the period between 1625 and 2017. The authors combined different repeat-sales data sets in Amsterdam to build a long-run repeat-sale house price index. For the period of analysis in this paper (1870-2018), the authors use three different data sets to build the final index. From 1870 to 1975 they combined the repeat-sales data set from Eichholtz (1997), which contains 5,269 repeat sales of houses on the Herengracht canal in Amsterdam, with a data set from Verwey (1943), who collected 2,880 repeat sales of properties in central Amsterdam auctioned between 1840 and 1940. For the period after 1975, the authors use 13,720 repeat sales for the city of Amsterdam from the data set of the Dutch Association of Realtors (NVM). To deal with the low number of observations in certain periods the authors employ the repeat-sales method developed in Francke (2010), which includes a local linear trend model, thereby reducing the noise in the periods with fewer transactions.

We updated the long-run index by Francke and Korevaar (2019) to 2018 using the house price index for Amsterdam constructed by *Statistics Netherlands* (CBS) using the sale price appraisal ratio (SPAR) method and data from the Dutch land registry office. More details about the construction of the series can be found in De Vries, Haan, Van der Wal, and Mariën (2009).

Table 1.K.33 summarizes the components of our final house price index. We decided to use the index by Francke and Korevaar (2019) and not the long-run

223. The median rental yield conveyed very similar results.
224. Reba, Reitsma, and Seto, 2016.

repeat-sales index by Eichholtz (1997), which covers the period between 1628 and 1973, for two reasons. Firstly, the index by Francke and Korevaar (2019) includes significantly more observations, due to the inclusion of the data from Verwey (1943). Secondly, because the index by Eichholtz (1997) only focuses on houses along the Herengracht canal, which is famous for being one of the most expensive in Amsterdam, it might be not representative of the entire city.

Table 1.K.33. Final house price index for Amsterdam

PERIOD	SOURCE	DESCRIPTION
1870-2017	Francke and Korevaar (2019)	*Type(s) of dwellings*: Owner-occupied dwellings; *Type of data*: Transaction and auction data; *Method*: Repeat-sales.
2017-2018	Netherlands (2020)	*Type(s) of dwellings*: Owner-occupied dwellings; *Type of data*: Transaction data; *Method*: SPAR method.

Rent Series. Eichholtz, Korevaar, and Lindenthal (2019) built a long-run rent index for Amsterdam for the period between 1550 and 2018. For the time period considered in this paper (1870-2018) their index is constructed based on data from different sources. For the period between 1870 and 1940 the authors collected rent prices from rental contracts in the Amsterdam City Archives. The largest share of rental contracts comes from the archives of the *Burgerweeshuis* (the Amsterdam orphanage) but also from the archives of the *Brants-Rus Almshouse* (the Roman-Catholic orphanage), and from the archives of different churches in Amsterdam. The authors use a repeat-rent approach based on the method developed in Francke (2010) to build the rent index from 1870 to 1940. For the period after 1940, the authors use the rent price index of the *Amsterdam Statistical Office*, which according to the authors follows the standards of *Statistics Netherlands*. Between 1994 and 2000, the authors rely on the rent component of the national rental index by *Statistics Netherlands*, which is built based on a sample of around 15,000 Dutch households. For the period after 2000 the authors use the index on average rent per square meter in Amsterdam constructed by Dröes, Houben, Van Lamoen, et al. (2017).

The rent series since 1940 mostly do not adjust for quality changes; however, as argued in Eichholtz, Korevaar, and Lindenthal (2019), this period was marked by very strict rent controls in the Netherlands, which were especially strong in Amsterdam. These rent controls were in place from 1927 until the late 1970s. Eichholtz, Korevaar, and Lindenthal (2019) argue that the rent series in that period are, thus, a good approximation of actual market rent growth.

Table 1.K.34 summarizes the components of our final rent index. To the best of our knowledge there do not exist any alternative rent indices for Amsterdam that we could use. As a result, we simply rely on the index by Eichholtz, Korevaar, and Lindenthal (2019).

Table 1.K.34. Final rent index for Amsterdam

PERIOD	SOURCE	DESCRIPTION
1870-2018	Eichholtz, Korevaar, and Lindenthal (2019)	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Rent prices from rental contracts; *Method*: Repeat-rent and average rent per square meter.

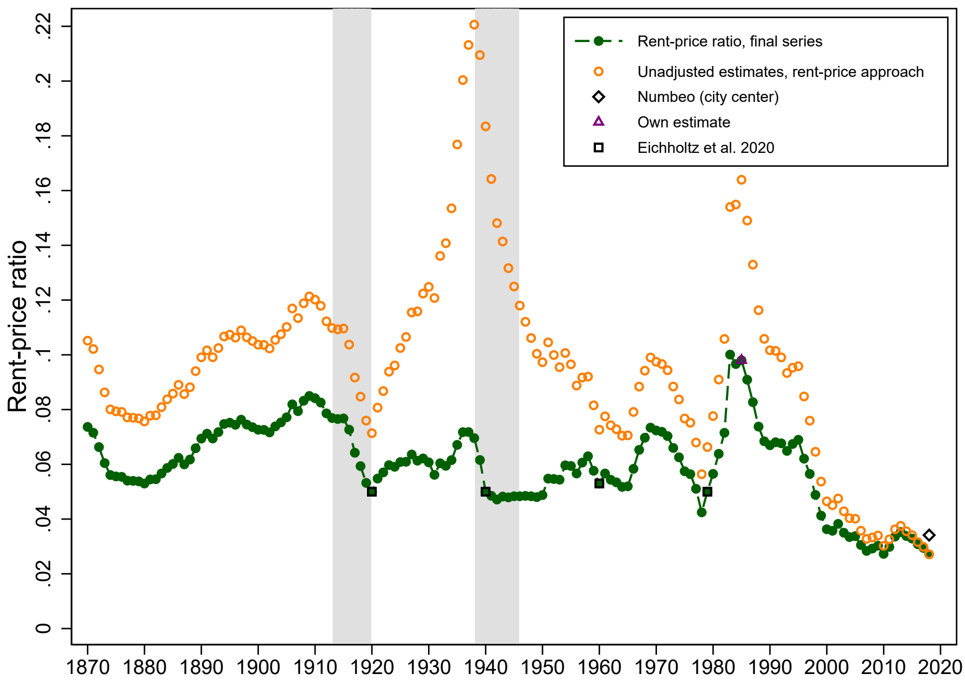

Figure 1.K.19. Amsterdam: plausibility of rental yields

Rental Yield Series. Our main benchmark for Amsterdam is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to the alternative benchmark we collected for 2018 from *Numbeo.com*. According to *Numbeo.com* the gross residential rental yield in the city center of Amsterdam in 2018 was 5.12%, to which we discount one-third costs, which gives us a net yield estimate of 3.41%. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.19.

As is visible from Figure 1.K.19 the unadjusted series rises substantially in the 1980s and in the interwar period. As mentioned above, rent controls were introduced in Amsterdam in 1927 and were slowly lifted after World War II until the end of the 1970s. One concern might be that the house price series over-samples a segment of the housing market, which was not under the rent control, while the rent series is based on a segment of the housing market strongly affected by the rent controls. This could create a strong upward bias in our rental yield series for this period, since our price series grows substantially more than the rent series.

To check whether this is actually the case we collected rental yield benchmarks from two different sources. For the years 1920, 1940, 1960 and 1979 we rely upon the estimates from Eichholtz, Korevaar, Lindenthal, and Tallec (2020), who built a net rental-yield series for Amsterdam using price and rental data for the same properties from the registers of a local real estate agents company. Since the authors have access to the exact rent amount received by each property and the associated costs, they are able to calculate very accurate net rental yields. As is clear from Figure 1.K.19 our unadjusted series lies significantly above these benchmarks. As a result, we use the benchmarks to adjust our series.

Since the series by Eichholtz, Korevaar, Lindenthal, and Tallec (2020) stops in 1979, we collected rental yield estimates for the year 1985 from the newspapers *De Telegraaf* and *Het Parool*. In total we collected 35 observations which feature both asking price and gross rental income for houses situated in Amsterdam.[225] The average gross-yield estimate is 18.8%. We then apply the same cost estimates from Eichholtz, Korevaar, Lindenthal, and Tallec (2020) for the year 1979 to get a net yield estimate of 9.8%. According to Eichholtz, Korevaar, Lindenthal, and Tallec (2020) taxes represented 17.6% and other costs 28% of the purchasing price. As far as we know there were no major changes to property taxes in the 1980s in Amsterdam. Additionally, we also adjust for vacancies, which were around 2% in Amsterdam in that period.[226] Again our benchmark lies substantially below the unadjusted rental yield series, as is clear from Figure 1.K.19. As a result, we also adjust our final series to the benchmark in 1985. This gives us the adjusted final rental yield series—the green-circled line in Figure 1.K.19.

1.K.10 Norway

Oslo was clearly the largest city in Norway, counting 227,000 inhabitants or a bit more than ten percent of the total population in 1900. It has also been covered in a project of the central bank with high-quality long-run house price data (Eitrheim and Erlandsen (2004)). This project also covered the three smaller cities of Bergen, Trondheim and Kristiansand, for which, however, there are no rental data, whereas there is some existing research on Oslo's rental market we can draw on (Oust, 2013). These cities also had only one third or less of Oslo's population in 1900.

225. Originally, we had collected 60 observations. We had to discard 25 of these, since it was not clear whether the rental income reported on the ad was for the whole house or only for a part of it. This was commonly the case when the house had a commercial store on the ground floor and residential apartments in the upper floors.

226. Eichholtz, Korevaar, Lindenthal, and Tallec, 2020.

1.K.10.1 Oslo

House Price Series. The house price index in Oslo is taken from the Norges Bank (Eitrheim and Erlandsen, 2004) and is built using a repeat sales method based on transaction data on different types of dwellings between 1870 and 1985 (cf. 1.K.35). For the time period between 1985 and 2009, the data come from the Norwegian association of real estate agents (*Norges Eiendomsmeglerforbund*) on the basis of which Norges Bank publishes a hedonic price index of detached and semi-detached houses. Since 2009, we draw on the hedonic index of existing detached houses, row houses and multi-family houses by *Statistics Norway* for Oslo and Bærum based on data from *Finn.no*. *Finn.no* is a website which advertises residential properties, but also collects data on actual sales from the largest and most important real estate companies in Norway. We deciced to use this series since 2009, because our rent series for the same period also covers Oslo and Bærum.

Table 1.K.35. Final house price index for Oslo

PERIOD	SOURCE	DESCRIPTION
1870-1985	Eitrheim and Erlandsen (2004) as reported by Norges Bank	*Type(s) of dwellings*: Owner-occupied residential dwellings; *Type of data*: Transaction data from official real property registers; *Method*: Repeat sales
1985-2009	Eitrheim and Erlandsen (2004) as reported by Norges Bank	*Type(s) of dwellings*: Detached and semi-detached houses; *Type of data*: Transaction data from association of real estate agents; *Method*: Hedonic price index
2009-2018	Statistics Norway as reported by Norges Bank	*Type(s) of dwellings*: Different types of existing dwellings; *Type of data*: Transaction data on existing residential dwellings from proivded by *Finn.no*; *Method*: Hedonic price index

Rent Series. The rent index combines different sources (1.K.36). First, for the period 1892-1950, we draw on the city yearbooks, which contain information on absolute rent levels for flats of different sizes and quality for benchmark years, but also a CPI-index for Oslo including a separate rent component which we extracted.

For the time period between 1950 and 1970, we adopted the data source and methodology from Oust (2013) by extracting about 260 annual rent price offers from the local newspaper *Aftenposten* and estimating a time-dummy hedonic regression of the log rent levels on year dummies, controlling for the number of rooms, the geography and the market segment. As market segments, the rental market was composed of two segments: the offers for exchanging rental units and rental market offers (similar as in Gothenburg). As geography, we recoded the addresses and districts from the ads into center, south, west, east, and north, following Oust (2013) on whose index we draw for the period between 1971 and 2008.

Since 2009, we use the rental market survey from Statistics Norway and the average annual rents for three-room apartments in the Oslo and Bærum municipalities.

Table 1.K.36. Final rent index for Oslo

PERIOD	SOURCE	DESCRIPTION
1892-1950	Yearbook of the city of Oslo (various years)	*Type(s) of dwellings*: Flats; *Type of data*: Municipal cost of living survey; *Method*: Rent component of the cpi series for Oslo
1950-1970	Own compilation	*Type(s) of dwellings*: Flats; *Type of data*: Newspaper ads from *Aftenposten*; *Method*: Stratified adjacent period time dummy hedonic rent index
1970-2008	Oust (2013)	*Type(s) of dwellings*: Flats; *Type of data*: Newspaper ads from *Aftenposten*; *Method*: Hedonic rent index
2008-2019	Statistics Norway	*Type(s) of dwellings*: three-room flats; *Type of data*: Rental market survey; *Method*: Simple average

Rental Yield Series. Our main benchmark for Oslo is *Numbeo.com*, as *MSCI* data are not available. The benchmark is for Oslo's city center and reasonably close to the alternative benchmark collected from *Catella* reports.[227] The latter lies slightly above our curve, as it reports on prime real estate only. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.20, which largely follows the trend in the national return series from Jordà, Knoll, et al. (2019). The series shows how the war-related rent control led twice to declines in rental yields with long post-war recoveries, the latter of which built up towards the 1990 Scandinavian house price burst, which was followed by a steep decline in rental yields. Given the high quality of the house price and rent series and the fact that we could not find historical rental yield benchmarks, we decided to keep the unadjusted series as our final series.

227. Source: Catella, "European commercial residential market map 2018".

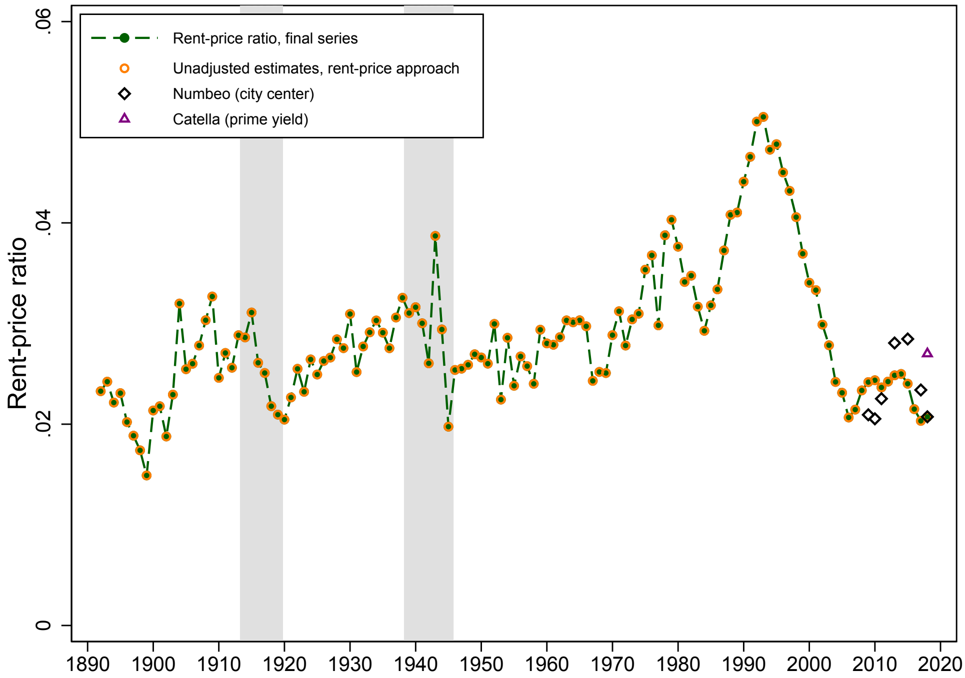

Figure 1.K.20. Oslo: plausibility of rental yields

1.K.11 Spain

In 1900 the three largest cities in Spain were in descending order: Barcelona, Madrid and Valencia. Each of the cities represented more than 1% of the total national population. Barcelona and Madrid were by far the largest and together represented 6% of the population.

Since 2007, the *Spanish Institute of Statistics* (INE) has been publishing high-quality house price series for different Spanish regions, which we explain in more detail below. Before that, and to the best of our knowledge, the only existing regional housing series for Spain cover only short periods of time and are not quality-adjusted.

We capitalize on the fact that the real estate ad section in newspapers grew exponentially after 1960 with the legalization of apartment sales to build our own long-run hedonic house price indices for Barcelona and Madrid.[228] As a result, most of our long-run house price series for both Barcelona and Madrid are composed

228. Before 1960 in Spain the market for apartments was practically non-existent. However, after the introduction of the law *Ley de Propiedad Horizontal* in 1960, the market for apartments started growing exponentially.

of new series. We did not find newspapers with a sufficient ad coverage to build a long-run index for Valencia.

1.K.11.1 Barcelona

House Price Series. For the period between 1954 and 2008 we build a yearly hedonic house price index for the city of Barcelona based on asking prices from the real estate ad section of the newspaper *La Vanguardia*. In total we collected 12,227 observations for apartments and multi-family houses, which contain information on the asking price, the location and the size of the dwelling.[229] To control for the different locations of the apartments, we include district fixed effects, where we use the official borders of ten districts of the city of Barcelona.[230] As mentioned in the introduction to the Data Appendix on Spanish cities, before 1960 there were very few sales of apartments. As a result, we had to build different house price indices for Barcelona, since the number of observations of apartments and multi-family houses changed considerably over time. Our index only starts in 1954, because the newspapers ad section had very few real estate ads for the years between 1950 and 1954.

For the period between 1954 and 1965 we built a time-dummy hedonic index based on multi-family house sales ("*Torres*") by regressing the log asking prices on the time dummies, district fixed effects and on the following set of controls: number of rooms, number of bathrooms and dummy variables for multiple other features (whether the dwelling has a kitchen, heating or balcony).

For the period between 1958 and 1965 we built a time-dummy hedonic index based on multi-family house ads ("*Torres*") and on apartment ads. The reason for this is that for this specific time period the number of observations we were able to collect for each dwelling type is relatively low. As a result, we pool the two different types of dwellings together and run the same regression as above, but now we also include a dummy for the dwelling type.

For the period between 1965 and 2008 we built a three-year rolling-window time-dummy hedonic index by regressing the log asking price on the time dummies, district fixed effects and on the following set of controls: number of rooms, number of bathrooms and dummy variables for multiple other features (whether the dwelling has a kitchen, garden, garage, heating, balcony or whether it was on the top floor of the building).

The *Spanish Institute of Statistics* (INE) publishes house price indices for different Spanish regions since 2007.[231] INE uses data on transaction prices of apartments and single-family houses and dwelling characteristics from the Spanish notaries' data

229. We winsorize the asking prices at the 1% by year.

230. The districts are: Ciutat Vella, Eixample, Gracia, Horta-Guinardó, Les Corts, Nou Barris, Sant Andreu, Sant Martí, Sants-Montjuic, Sarria-Sant Gervasi.

231. The index is built for 19 *Comunidades Autonomas*, which are political and administrative divisions of Spain.

set, which includes information on approximately the universe of transactions in Spain. INE uses these data to build an imputed hedonic index for different dwelling types. The imputed indices are then aggregated using the number of transactions as weights and transformed into a single index using a chained Laspeyres approach.

The Ministry of Development (*Ministerio de Fomento*) published the average dwelling price per square meter for the region of Madrid between 1987 and 2004. The data come from different credit institutions for both new and existing residential dwellings. The price estimates are not quality-adjusted.

In the statistical yearbooks of the city, the *Barcelona City Council* published estimates of average house prices of existing dwellings, i.e. excluding newly built dwellings, for the period between 1975 and 1991. The data were collected from newspapers.

We also calculated yearly average house prices using the total number and total value of residential dwellings transacted in Barcelona for the period between 1950 and 1960. We used the data from the notaries' yearbook (*Anuario de la Dirección General de los Registros y del Notariado*), which contains summary statistics on the universe of real estate transactions recorded by notaries in Spain throughout the year for different autonomous communities (*Comunidades Autonomas*). Due to the lack of quality adjustments, we build a three-year moving average.

Table 1.K.37 summarizes the components of our final house price index. From 1950 to 1954 we use the index based on average prices from the notaries' yearbook. This series is not quality-adjusted, but it is, to the best of our knowledge, the only existing series for this period. From 1954 to 1958 we use the hedonic index based on multi-family ads. We do use this index until 1965, because the number of multi-family houses in the ad section decreases substantially from 1960 onward. From 1958 to 1965 we use the hedonic index based on both apartments and multi-family houses. From 1965 to 2008 we rely on the three-year rolling window index own based solely on apartment ads. Although our hedonic series are based on asking prices, they are the only continuous series for this period which are quality-adjusted. From 2008 onward, we use the official house price index from INE.

Rent Series. The *Spanish Institute of Statistics* (INE) publishes the rent component of the Consumer Price Index regularly for different regions of Spain at least since 1947. For the pre-1985 period INE published the rent indices at the city-level for all province capitals. Since 1985 INE has been publishing the rent indices for all autonomous communities (*Comunidades Autonomas*). And since 2002 INE also publishes rent indices at the province level. The rent index is based on rent prices collected in the Survey on the Active Population (*Encuesta de Población Activa*) from INE, which relies on a rotating sample. This allows INE to use a matched-model approach to calculate average rent changes for the same dwellings over time. As a result, the rent series is mainly composed of existing rental contracts and very little influenced by the values of new rents. Spain introduced rent controls in the 1930s,

Table 1.K.37. Final house price index for Barcelona

PERIOD	SOURCE	DESCRIPTION
1950-1954	Own compilation	*Type(s) of dwellings*: Owner-occupied dwellings; *Type of data*: Transaction prices from the notaries; *Method*: Simple moving-average.
1954-1958	Own compilation	*Type(s) of dwellings*: Multi-family houses; *Type of data*: Asking prices from *La Vanguardia*; *Method*: Time dummy hedonic series.
1958-1965	Own compilation	*Type(s) of dwellings*: Multi-family houses and apartments; *Type of data*: Asking prices from *La Vanguardia*; *Method*: Time dummy hedonic series.
1965-2008	Own compilation	*Type(s) of dwellings*: Apartments; *Type of data*: Asking prices from *La Vanguardia*; *Method*: Rolling-window hedonic series.
2008-2018	INE (2021b)	*Type(s) of dwellings*: Apartments and single-family houses; *Type of data*: Transaction prices from the notaries' database; *Method*: Chained Laspeyres imputed hedonic index.

which lasted until the mid-1960s. Newly constructed buildings were to a large extent exempted from these rent controls and, as a result, the official rent index does not reflect the evolution of rents in the non-controlled rental market. We collected the CPI rent component index using the data from various editions of the Spanish statistical yearbook (*Anuario Estadístico de Espana*).

The department of statistics of the city of Barcelona (*Departament d'Estadística i Difusió de Dades. Ajuntament de Barcelona*) has been publshing the average rent per square meter for new rental contracts in the city of Barcelona since 2000. The averages are based on the universe of new rental contracts, which are collected by the Catalan land institute. In Catalunya all deposits of urban rental contracts must be deposited in the Catalan land institute (*Institut Català del Sòl*).

Table 1.K.38 summarizes the components of our final rent index. To the best of our knowledge there do not exist other historical rent series for Barcelona. As a result we rely for the complete period between 1950 and 2000 on the CPI rent component series from INE. This index measures the evolution of rent prices in Barcelona for the period between 1950 and 1985. For the post-1985 period the index measures the evolution of rent prices in Catalunya. Since Barcelona is by far the largest and most expensive city in the autonomous community Catalonia, we expect the rent series to be a good approximation of the actual rent evolution in the city of

Barcelona.[232] From 2000 onward, we rely on the average rent per square meter from the department of statistics of the city of Barcelona. Although this indicator does not fully control for quality changes in the samples, it covers the city of Barcelona and takes into account new rental contracts. A comparison with the index from INE, shows that it grows more for the period between 2000 and 2018 and it shows cyclical variability around the 2007 financial crisis, which is something that is missing in the index from INE.

Table 1.K.38. Final rent index for Barcelona

PERIOD	SOURCE	DESCRIPTION
1950-2000	INE (2021a)	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Rent prices from survey; *Method*: Matched-model approach.
2000-2018	Difusió de Dades. Ajuntament de Barcelona (various years)	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Rent prices from the Catalan land institute; *Method*: Average rent per square meter.

Rental Yield Series. Our main benchmark for Barcelona is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to the alternative benchmark we collected for 2018 from *Numbeo.com*. According to *Numbeo.com* the gross residential yield in the city center of Barcelona in 2018 was 3.5%, to which we discount one-third costs, which gives us a net yield estimate of 2.7%. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.21.

As discussed above, the rent series from INE is mainly composed of existing rents, thereby ignoring to a great extent the evolution of new rents, which can lead to biases in periods in which newly constructed buildings are excluded from rent freezes. In contrast, our house price index captures both the evolution prices of existing dwellings and the evolution of prices of new dwellings. This could potentially create a bias in our rental yield series. Moreover, the rent series covers the entire autonomous community Catalonia, whereas the house price series focuses on the city of Barcelona for most time periods. This might also bias the rental yield series. To check whether this is the case we collected rental yield benchmarks from the newspaper *La Vanguardia* for the beginning of our sample period. We collected on average 30 observations per year, for which we had both the asking price and gross

232. In 2011, from the 7,519,843 people living in the autonomous community Catalonia, 5,522,565 lived in the Province Barcelona and still 1,611,013 in the municipality of Barcelona (Source: INE, population and housing census 2011.

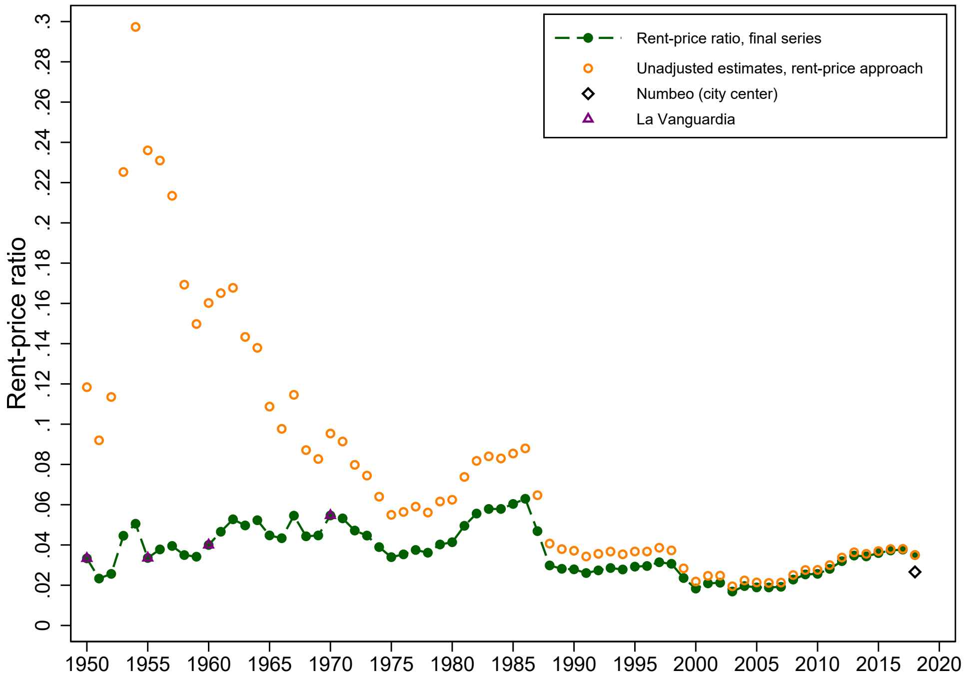

Figure 1.K.21. Barcelona: plausibility of rental yields

rental income for the same multi-family house. Using these data we build average rental yields for the years 1950, 1955, 1960 and 1970.[233] We then adjust the gross estimates for one-third costs and get the net yield estimates, which can be seen in Figure 1.K.21. The net yield estimates from *La Vanguardia* lie substantially below our unadjusted rental yield series. Since this difference is quite large and probably reflects the above-mentioned bias in our unadjusted rental yield series, we use the historical estimates to benchmark our series.

This gives us the adjusted final rental yield series — the green-circled line in Figure 1.K.21.

1.K.11.2 Madrid

House Price Series. For the period between 1950 and 2010 we build a yearly hedonic house price index for Madrid based on asking prices from the real estate ad section of the newspaper *ABC*. In total we collected 12,767 observations, which contain information on the asking price, the location and the size of the dwelling. As we explain in more detail below, the types of dwelling included in our data set as well as the variable we use to control for the size of the dwellings changes over

233. The median yields very similar results.

time.[234] Additionally, our number of observations increases substantially for more recent decades. To account for these differences over time, we built different indices for subperiods between 1950 and 2010. For all these indices we include district fixed effects, where we use the official borders of 20 districts of the city of Madrid to assign the observations.[235] Since the months for which we have data change over time we also control for the quarter in which the ad was posted. While until 1975 most of the ads included information on the number of rooms in the dwelling, this changed in the post-1975 period, where most ads include the total square meters of the dwelling instead.[236]

For the period between 1950 and 1990 we built both a five-year rolling-window hedonic index as well as an adjacent-period hedonic index by regressing the log asking price on year dummies, location and quarter fixed effects and on the following set of controls: number of rooms, type of dwelling (single-family, duplex or apartment), number of bathrooms, and dummies for other features (whether the dwelling has a kitchen, garden, garage, heating or air conditioning).

For the period between 1975 and 2010 we built an adjacent-period hedonic index, which is equal to the index for the period between 1950 and 1990 with one exception: instead of controlling for the number of rooms, we control for the log squared meters of the dwelling. The reason for this is that after 1975 most ads included the square meterage of the dwelling instead of the number of rooms. In order not to lose the observations for which we do not have square meterage but only the number of rooms, we predict the square meterage based on the observations for which we have both the number of rooms and the square meterage.

The *Spanish Institute of Statistics* (INE) publishes house price indices for different Spanish regions since 2007. INE uses data on transaction prices of apartments and single-family houses and dwelling characteristics from the Spanish notaries' data set, which includes information on approximately the universe of transactions in Spain. INE uses these data to build an imputed hedonic index for different dwelling types. The imputed indices are then aggregated using the number of transactions as weights and transformed into as single index using a chained Laspeyres approach.

The Ministry of Development (*Ministerie de Fomento* published the average price per square meter for the region of Madrid between 1987 and 2004. The data come

234. We winsorize the asking prices at the 1% by year.

235. The districts are: Centro, Arganzuela, Retiro, Salamanca, Chamartín, Tetuán, Chamberí, Fuencarral-El Pardo, Moncloa-Aravaca, Latina, Carabanchel, Usera, Puente de Vallecas, Moratalaz, Ciudad Lineal, Hortaleza, Villaverde, Villa de Vallecas, Vicálvaro and San Blas-Canillejas Barajas.

236. In the post-1975 sample we include observations for which we estimated the number of square meters. For the estimation we use the coefficient from a simple linear regression of size in square meters on number of rooms for a sample of observations for which we were able to collect both variables. Excluding the observations for which we estimate the size in square meters only changes our indices slightly.

from different credit institutions for both new and existing residential dwellings. The price estimates are not quality-adjusted.

The city of Madrid also published average prices of residential dwellings for the period between 1960 and 1974 for the metropolitan area of Madrid. The estimates are based on a survey conducted by the planning commission of the city of Madrid (*Comisión de Planeamiento y Coordinación del Área Metropolitana de Madrid*), which collected prices of both privately-owned houses and officially sponsored houses.

The real estate company *Tecnigrama* published average prices per square meter for new dwellings in the city of Madrid between 1976 and 1990.

Table 1.K.39 summarizes the components of our final house price index. From 1950 to 2010 we rely on our own index based on data from *ABC*. More precisely, from 1950 to 1960 we use the five-year rolling-window hedonic index, because the number of observations per year is not as high as in later years, from 1960 to 1980 we use the adjacent period hedonic index, in which we control for the number of rooms, and from 1980 to 2010 we rely on the adjacent period index, in which we control for the total square meters. Although this series is based on asking prices, it is the only continuous series, which is quality-adjusted. From 2010 onward, we use the official house price index from INE.

Table 1.K.39. Final house price index for Madrid

PERIOD	SOURCE	DESCRIPTION
1950-2010	Own compilation	*Type(s) of dwellings*: Apartments and single-family houses; *Type of data*: Asking prices from *ABC* ; *Method*: Rolling-window and adjacent period hedonic indices.
2010-2018	INE (2021b)	*Type(s) of dwellings*: Apartments and single-family houses; *Type of data*: Transaction prices from the notaries' database ; *Method*: Chained Laspeyres imputed hedonic index.

Rent Series. As we explain in Section 1.K.11.1 above for Barcelona, the *Spanish Institute of Statistics* (INE) publishes the rent component of the Consumer Price Index regularly. The index is based on rent prices collected in the Survey on the Active Population (*Encuesta de Población Activa*) from INE, which is based on a rotating sample. This allows INE to use a matched-model approach to calculate average rent changes for the same dwellings over time. As explained above, this official rent index does not reflect the evolution of rents in the non-controlled rental market. We collected the CPI rent component index using the data from various editions of the Spanish statistical yearbook (*Anuario Estadístico de Espana*).

From 2005 onward the real estate website *Idealista.es* has been publishing the average rent per square meter in the city of Madrid.[237] The estimates are based on the average asking rent of all residential rental units located in the city of Madrid, which can be found in the website. We use these estimates of average rent to build a rent index for the city of Madrid between 2005 and 2018.

Table 1.K.40 summarizes the components of our final rent index. To the best of our knowledge there do not exist other historical rent series for Madrid. As a result we rely for the complete period between 1950 and 2005 on the CPI rent component series from INE. This index measures the evolution of rent prices in the city of Madrid for the period between 1950 and 1985. From 1985 to 2002 the index measures the evolution of rent prices in the autonomous community of Madrid. The autonomous community Madrid coincides with the Province of Madrid and approximately covers the metropolitan area of Madrid.[238] Hence, we expect the rent series to be a very good approximation of the actual rent evolution in the city of Madrid. From 2002 to 2005 the index measures the rent developments in the province of Madrid. From 2005 to 2018 we rely on the index from *Idealista.es*, since this index covers new rental contracts and focuses specifically on the city of Madrid.

Table 1.K.40. Final rent index for Madrid

PERIOD	SOURCE	DESCRIPTION
1950-2005	INE (2021a)	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Rent prices from survey; *Method*: Matched-model approach.
2005-2018	*Idealista.es*	*Type(s) of dwellings*: Renter-occupied dwellings; *Type of data*: Asking rent prices from online advertisements; *Method*: Average rent per square meter.

Rental Yield Series. Our main benchmark for Madrid is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to the alternative benchmark we collected for 2018 from *Numbeo.com*. According to *Numbeo.com* the gross residential rental yield in the city center of Madrid in 2018 was 4.37%, to which we discount one-third costs, which gives us a net yield estimate of 2.9%. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.22.

We also plot a net yield benchmark from *MSCI* for 2014, which is exactly on our unadjusted rental yield series.

237. The series can be found in the website.

238. In 2011, from the 6,421,874 people living in the autonomous community Madrid, 3,198,645 lived in the municipality of Madrid (Source: INE, population and housing census 2011.

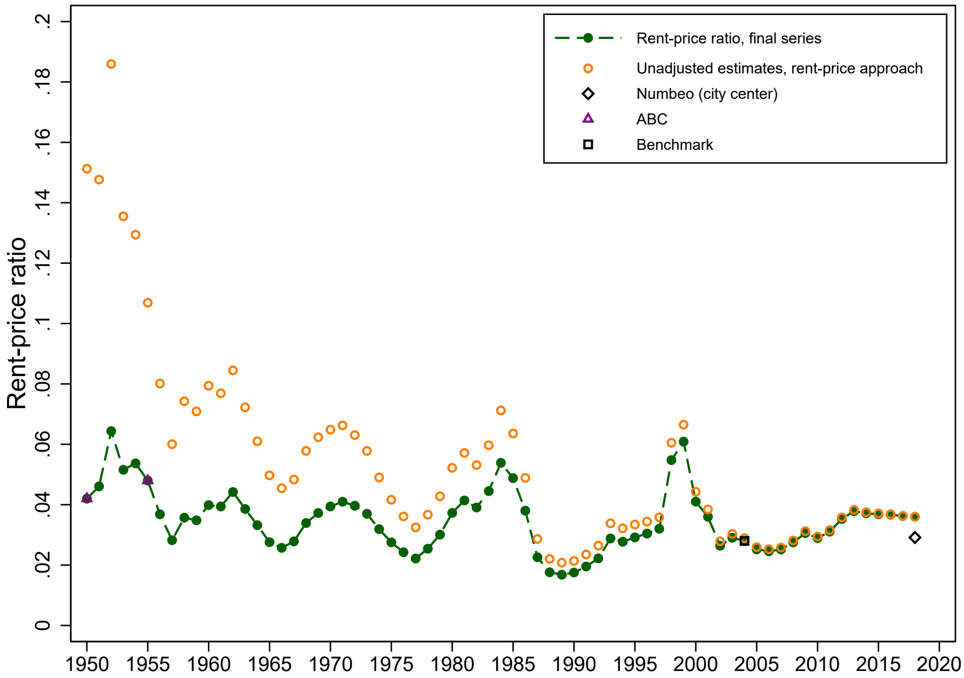

Figure 1.K.22. Madrid: plausibility of rental yields

As discussed above, the rent series from INE is mainly composed of existing rents, thereby ignoring to a great extent the evolution of new rents, which can lead to biases in periods in which newly constructed buildings are excluded from rent freezes. In contrast, our house price index captures both the evolution prices of existing dwellings and the evolution of prices of new dwellings. This could potentially create a bias in our rental yield series. To check whether this is the case we collected rental yield benchmarks from the newspaper *ABC* for the beginning of our sample period. In total we collected 60 observations, for which we had both the asking price and gross rental income for the same multi-family house. Using these data we build average rental yields for the years 1950 and 1955.[239] We then adjust the gross estimates for one-third costs and get the net yield estimates, which can be seen in Figure 1.K.22. The net yield estimates from *ABC* lie substantially below our unadjusted rental yield series. Since this difference is quite large and probably reflects the above-mentioned bias in our unadjusted rental yield series, we use the historical estimates to benchmark our series.

This gives us the adjusted final rental yield series—the green-circled line in Figure 1.K.22.

239. The median yields very similar results.

1.K.12 Sweden

Stockholm and Gothenburg were Sweden's largest cities already by 1900, making up a bit less than six and three % of the country's population respectively. Both are the center of Sweden's now two largest metropolitan regions. Both cities have been subject to detailed and high-quality long-run house price studies, whereas consistent rental data are not readily available. The next largest cities in 1900 were Malmö and Norrköping, whose absolute population, however, were 60,000 or less. We are not aware of existing high-quality long-run price or rent data for these cities.

1.K.12.1 Gothenburg

House Price Series. The long-run house price series for Gothenburg is a high-quality series created by various authors and *Statistics Sweden*. For the time period 1875-1957, we use the index constructed by Bohlin (2014). The data he used are sourced from property tax valuations for houses and combined with purchasing prices from a ledger of properties. The index is built using the SPAR method. From 1957 until 2013, we rely on the index on houses in Gothenburg published in Edvinsson, Blöndal, and Söderberg (2014). The authors used data on small houses from *Statistics Sweden* and built a house price index using the SPAR method. Since 2013, the data we use are from *Statistics Sweden* on transaction and tax valuations of buildings with one or two dwellings, again using the SPAR method. The components of our final house price index are summarized in Table 1.K.41.

Table 1.K.41. Final house price index for Gothenburg

PERIOD	SOURCE	DESCRIPTION
1875-1957	Bohlin (2014)	*Type(s) of dwellings*: Houses; *Type of data*: Property tax valuations and purchasing prices from the *Lagfartsprotokoll*; *Method*: SPAR method.
1957-2012	Edvinsson, Blöndal, and Söderberg (2014)	*Type(s) of dwellings*: Small houses; *Type of data*: Tax valuations and purchasing prices from Statistics Sweden; *Method*: SPAR method.
2013-2018	Statistics Sweden	*Type(s) of dwellings*: One- or two-dwelling buildings; *Type of data*: Individual transaction and tax valuation data; *Method*: SPAR method.

Rent Series. For the rental series, we make use of three different sources: prior to 1950 we draw on the rent surveys for benchmark years reported in Gothenburg's statistical yearbook, using the average rent for a three-room dwelling. We interpolate missing years by using the rent component of the national CPI-index.

Between 1950 and 1978, we collected ads for rental housing in Gothenburg's most important local newspaper, the *Göteborgposten*, which published an ad section covering the local rental market ("*hyresmarknaden*"). This contained predominantly two rental market segments: offers for housing exchanges (due to housing shortages throughout this period) and offers for cooperative rental units ("*insatslägenheter*"). We scanned these relevant pages between 1950 and 1964 from microfilm in the Gothenburg university library and downloaded digitized versions from the website *tidningar.kb.se* for the later period. We extracted all ads containing full information on price, rent and location, yielding a total of 2,181 observations. We then estimated a hedonic price regression of log rent levels using the time-dummy approach with number of rooms, type of offer, quality (new/old) and location as control variables. For the location variable, we recoded the neighborhood and addresses mentioned in the ads into the historical administrative division of ten city districts (plus periphery). The adjusted R^2 is 0.857.

Lastly, we connected the newly constructed series to those reported by *Statistics Sweden* (SCB) on the Greater Gothenburg region, using the three-room average annual rents for existing contracts from 1978 until 1990, the average annual rent per square meter for existing contracts for renting apartments in Greater Gothenburg between 1990 and 1994 and the average rent per square meter for three-room apartments for both existing and new contracts in Gothenburg since 1994. The data are reported in the "*Bostads- och hyresundersökningen serie BO*" for the historical time period and on the SCB website for more recent years. Table 1.K.42 summarizes the components of our final rent index.

Table 1.K.42. Final rent index for Gothenburg

PERIOD	SOURCE	DESCRIPTION
1914-1950	Statistics Sweden	*Type(s) of dwellings*: Three-room dwellings; *Type of data*: Municipal rent surveys; *Method*: Average annual rents (interpolated).
1950-1978	Own series	*Type(s) of dwellings*: Apartments; *Type of data*: Newspaper ads from *Göteborgposten*; *Method*: Hedonic rent index.
1978-1990	Statistics Sweden	*Type(s) of dwellings*: Three-room dwellings; *Type of data*: Rent surveys ("*Hyresräkning*"); *Method*: Average annual rents.
1990-1994	Statistics Sweden	*Type(s) of dwellings*: Apartments in existing stock (Greater Gothenburg); *Type of data*: Rent survey from Statistics Sweden; *Method*: Average annual rents per sqm.
1994-2018	Statistics Sweden	*Type(s) of dwellings*: Existing and new contracts for apartments with three rooms; *Type of data*: Rent survey from Statistics Sweden; *Method*: Average annual rents per sqm.

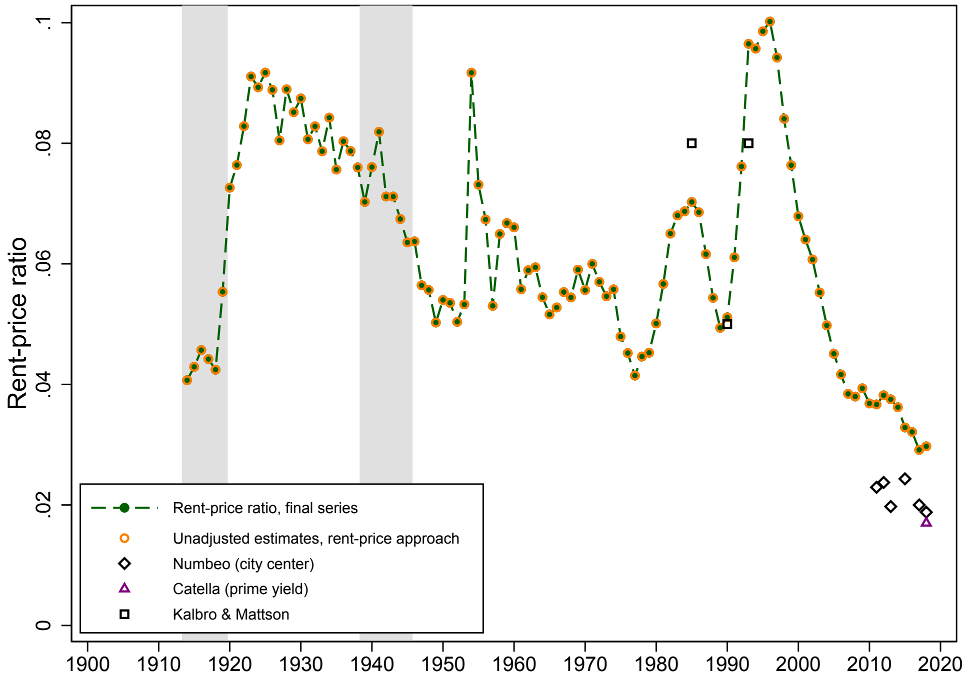

Figure 1.K.23. Gothenburg: plausibility of rental yields

Rental Yield Series. Our main benchmark for Gothenburg is taken from *MSCI*, as described in the main paper. Applying the rent-price approach to this benchmark results in the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.23.

We collected two additional benchmarks for 2018. First, we use the benchmark for the city-center of Gothenburg from *Numbeo.com*, which we adjust to capture net rental yields by subtracting one-third following Jordà, Knoll, et al. (2019). Second, we use the prime yield for Gothenburg from *Catella* reports.[240] Both benchmarks are somewhat below but close to our main benchmark from *MSCI*. We use the benchmark by *Numbeo.com* as our alternative benchmark in the robustness section of the main paper. Additionally, we collected benchmarks from *Numbeo.com* for the years from 2011 to 2017. These benchmarks show a similar pattern compared to our long-run series.

We also collected historical rental yield benchmarks for Gothenburg from Kalbro and Mattsson (1995) for the years 1985, 1990 and 1993. These benchmarks are very close to our unadjusted long-run rental yield series. The unadjusted series shows a long-term decline of rental yields after the 1920s. With the introduction of rent-controls, there were spikes after the two World Wars and a considerable spike around

240. Source: Catella, "European commercial residential market map 2018".

the burst of the 1990 housing bubble. Given that our unadjusted rental yield series is close to the historical benchmarks and shows probable patterns given the historical background, we do not adjust our long-run series.

1.K.12.2 Stockholm

House Price Series. The long-run house price index for Stockholm is composed of three different series and largely following the seminal work by Edvinsson, Blöndal, and Söderberg (2014). Between 1875 and 1956, this index is calculated using the SPAR method and relies on the property tax valuations from the *Stockholms adresskalender* as well as purchase prices from the register of certificates of title to properties, which is in the archives of the Stockholm Magistrates' Court. From 1957 onward until 2012, the authors use the already constructed index by *Statistics Sweden*. The index is built using the SPAR method also using data on tax valuation and purchase prices for small houses and apartment buildings. We rely on the index on houses. Finally, the most recent data are from *Statistics Sweden*, which provides an index built using the SPAR method based on individual transaction and tax valuation data on one- or two-dwelling buildings. Table 1.K.43 summarizes the components of our final rent index.

Table 1.K.43. Final house price index for Stockholm

PERIOD	SOURCE	DESCRIPTION
1875-1956	Edvinsson, Blöndal, and Söderberg (2014)	*Type(s) of dwellings*: Houses; *Type of data*: Property tax valuations and purchase prices from *Stockholms adresskalender*; *Method*: SPAR method.
1957-2012	Edvinsson, Blöndal, and Söderberg (2014)	*Type(s) of dwellings*: Small houses; *Type of data*: Index from Statistics Sweden; *Method*: SPAR method.
2013-2018	Statistics Sweden	*Type(s) of dwellings*: One- or two- dwelling buildings; *Type of data*: Individual transaction and tax valuation data; *Method*: SPAR method.

Rent Series. The long-run rental series are based on two different sources (Table 1.K.44): First, Stockholm's statistical yearbook, which provides benchmark rents for apartments of various room sizes for the city center of Stockholm that we use to build a stratified Fisher index. Strata are defined by the number of rooms of the dwelling. For the period between 1894 and 1914 and the period between 1943 and 1960 there are data every five years. For these two periods we construct a non-chained Fisher Index.[241] For the period in between 1915-1942, there are no missing data, so we construct a chained Fisher index.

The second source starts in 1960. Here we can use the regional break-down of nationally surveyed rental market statistics which are available for the region of Stockholm by different room classes on a yearly basis, published as the "*Bostads- och hyresundersökningen serie BO*" in Stockholm's statistical yearbooks and by *Statistics Sweden*. Strata are defined by the number of rooms in the dwelling. For the period between 1960 and 1978 there are data only in 1960, 1965, 1968 and 1978. For this periods we construct a non-chained Fisher Index. The missing data between 1960 and 1968 are linearly interpolated. Between 1968 and 1978 we use two different rent measures published in Stockholm's statistical yearbooks for interpolation. Between 1968 and 1970, we use the rent per room in the city center and the suburbs of Stockholm. We calculate the increase between 1968 and 1970 for both locational categories and then take a simple average. Between 1970 and 1978 we use yearly data on the rent in multi-family houses built by private enterprise with state housing loans. We calculate a chained Fisher stratification index and use this index for interpolation. For the years after 1978 we construct a chained Fisher stratification index.

Table 1.K.44. Final rent index for Stockholm

PERIOD	SOURCE	DESCRIPTION
1894-1960	Yearbook of the city of Stockholm	*Type(s) of dwellings*: Apartments of different room sizes (city center); *Type of data*: Municipal rent surveys; *Method*: Stratified Fisher Index (interpolated).
1960-2019	Yearbook of the city of Stockholm & Statistics Sweden	*Type(s) of dwellings*: Apartments of different room sizes; *Type of data*: Surveys by Statistics Sweden (Hyresräkning); *Method*: Stratified Fisher index (interpolated for earlier years).

Rental Yield Series. Our main benchmark for Stockholm is taken from *MSCI*, as described in the main paper. Applying the rent-price approach to this benchmark results in the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.24.

We collected two additional benchmarks for 2018. First, we use the benchmark for the city-center of Stockholm from *Numbeo.com*, which we adjust to capture net rental yields by subtracting one-third following Jordà, Knoll, et al. (2019). Second,

241. We use the first year of the respective period as base year. Years in-between are linearly interpolated.

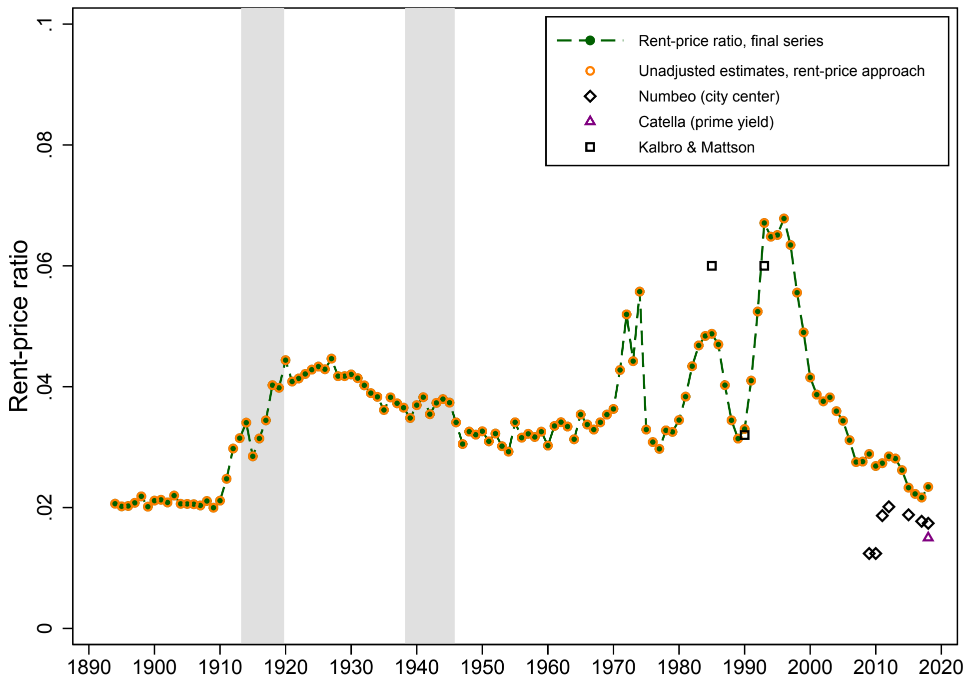

Figure 1.K.24. Stockholm: plausibility of rental yields

we use the prime yield from *Catella* reports.[242] Both benchmarks are somewhat below but close to our main benchmark from *MSCI*. We use the benchmark by *Numbeo.com* as our alternative benchmark in the robustness section of the main paper. Additionally, we collected benchmarks from *Numbeo.com* for the years from 2009 to 2017. These benchmarks are also below our unadjusted series.

We also collected historical rental yield benchmarks for Stockholm from Kalbro and Mattsson (1995) for the years 1985, 1990 and 1993. These benchmarks are very close to our unadjusted long-run rental yield series. Given that our unadjusted rental yield series is close to the historical benchmarks and shows similar patterns as the series for Gothenburg, we do not adjust our long-run series.

1.K.13 Switzerland

The five largest cities in Switzerland in 1900 were in descending order: Zurich, Geneva, Basel, Bern, and Lausanne. All of these cities covered more than 1% of the country's population. We have, however, not been able to find sufficient data to construct long-run house price or rent series for Geneva and Lausanne. By including

242. Source: Catella, "European commercial residential market map 2018".

Zurich, Basel and Bern, we still cover almost 10% of Switzerland's population in 1900.

To the best of our knowledge, there do not exist any compiled house price series on city-level from public sources in Switzerland. Publications from city-level statistical offices within major cities do, however, often contain aggregated housing sales data. In many cases, these data are given for different (inner-city) locations, such that it is possible to control for sample shifts within the location of sold properties. As location is maybe the most important quality characteristic for houses and, within cities, other highly relevant characteristics like size, housing types and construction period are highly correlated with location,[243] this quality control probably minimizes the bias in the house price series.

Apart from these sources, private companies are able to provide regional real estate indices for sale. These indices are typically constructed using private micro-level data sets. The methods are, however, in many cases not made transparent. Due to missing alternatives, we had to rely on such data for Basel and Bern for the most recent decades.

Regarding rent series, statistical offices of large cities also regularly published city-level rent series. These have often been calculated to construct city-level CPI series and are based on reliable and transparent methods.

We discuss the construction of our long-run city-level series in more detail below. The discussion is structured by cities and within cities, by price series, rent series and rental yield series.

1.K.13.1 Basel

House Price Series. To build the long-run house price index for Basel, we use various statistical publications. For much of the period, these publications provide data by district, such that we are able to control for sample shifts in the location of sold properties. For some periods, however, data are only given for the whole Basel-city region ("Kanton Basel-Stadt"). Moreover, in 1982, the statistical office stopped publishing prices of sold properties for a considerable time period. To get a long-run index, we, therefore, had to rely on a private house price index provided by *Wüest Partner*. We describe the different parts of the index in more details below.

For the last four decades, to the best of our knowledge only insufficient house price data is available for Basel from public sources. This means that we have to rely on private house price indices. We use a transaction price index for single-family houses provided by *Wüest Partner*. This index is constructed using hedonic methods

243. This is especially true for Swiss cities, because locational zoning restrictions imply that only buildings of specific types are allowed in specific parts of the city. For Zurich, we did compile an experimental index using house price data disaggregated by these construction zones ("*Bauzonen*"). The resulting index is very similar to an index using location bins instead.

and covers the region Basel-City ("Kanton Basel-Stadt"). In addition to the municipality Basel, this region also covers the two considerably smaller municipalities Riehen and Bettingen, which are part of the MS-region Basel. The index starts in 1985.

Prior to this year, we rely on aggregate house price data from statistical yearbooks.[244] These data cover all voluntary sales of developed lots within the region Basel-City.[245] From 1921 to 1963, the data are given separately for the 19 districts ("*Quartiere*") of the Basel municipality as well as the municipalities of Riehen and Bettingen.[246] We use average price per transaction by district to calculate a chained Fisher-type stratification index following Eurostat (2013).[247] Thereby, we control for locational shifts over time within the sample of sold developed lots. Between 1964 and 1981, the yearbooks only contain average price per transaction for the entire Basel-City region. We therefore have to rely on these averages to construct the house price index for this period. Data for the years 1982 to 1984 as well as for the year 1926 are missing. To link the index ending in 1981 to the index starting in 1985, we use the average price per sale of developed lots by district for 2006 from statistical yearbooks.[248] 2006 is the earliest year for which the data were published again. To control for locational shifts within the sample of transacted properties, which is probably even more important over a long time period, we link the data to the data in 1963 and again calculate a Fisher-type stratification index. We use this index and our yearly indices until 1981 and starting in 1985 to calculate the implied increases in house prices between 1981 and 1985. Afterwards, we linearly interpolate missing years to calculate yearly housing returns.

Prior to 1921, we again have to rely on average price per sold developed lot in the Basel-City region. The data start in 1912 and are collected from a publication of the national statistical office.[249]

Table 1.K.45 summarizes the components of our final house price index.

Rent Series. To construct the long-run rent index for Basel, we mainly rely on data published by the statistical office of the Basel-City region ("Kanton Basel-Stadt"). For the last five decades, we use the rent index calculated by the statistical office. For the

244. "*Statistisches Jahrbuch des Kantons Basel-Stadt*" (Volumes 1921-1982).

245. "*Handänderungen von Freihandverkäufen bebauter Grundstücke*".

246. Approximately 10% of all sales in the region Basel-City have been in the municipalities of Riehen and Bettingen between 1921 and 1963. We still include these municipalities, because they are in commuting distance to the center of Basel and to be consistent with earlier and later periods.

247. To calculate the chained index, for every pair of consecutive years, we drop all districts that feature less than three sales in one of the two years. This ensures that the average is a meaningful characteristic and the influence of outliers is reduced.

248. Source: "*Statistisches Jahrbuch des Kantons Basel-Stadt*" (2006))

249. Source: "*Die Statistik der hypothekarischen Verschuldung und der Handänderungen (Grundbuchstatistik) in einigen Kantonen*", in "*Schweizeriche Zeitschrift für Volkswirtschaft und Statistik*" *(1930) published by the* "*Eidgenössisches Statistisches Amt*".

Table 1.K.45. Final house price index for Basel

PERIOD	SOURCE	DESCRIPTION
1912-1921	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales within Basel-City region from statistical publication; *Method*: Average price per transaction.
1921-1963	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated by district from yearbooks; *Method*: Stratification.
1964-1981	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales within Basel-City region from yearbooks; *Method*: Average price per transaction.
1985-2018	Wüest Partner	*Type(s) of dwellings*: Single-family houses; *Type of data*: Private transaction-level data; *Method*: Hedonic index.

period before, we use average rent by number of rooms published in the statistical yearbooks. Prior to 1920, we use the rent index for Basel published in Curti (1981).

Starting in 1970, our long-run series is composed of the residential rent index for the Basel-City region published by the statistical office.[250] This index was constructed using a random stratified sample covering 5% of all rented dwellings that have one to six rooms.[251]

Prior to 1970, we use residential rent data published in the statistical yearbooks.[252] To ensure constant quality of the sample used to build the rent index, we only rely on average rents of dwellings built prior to 1920 and after 1940. Between 1954 and 1970, we use the averages for dwellings featuring a bathroom that were built between 1900 and 1920 and had between two and four rooms. We build a weighted average for each year over the number of room bins (2-4) using the number of dwellings by number of rooms from the housing census in 1970 as weights. Afterwards, we calculate the increase of the yearly weighted average. As we use the same weights and types of dwellings every year, this procedure minimizes the influence of sample shifts within the sample of rented dwellings. Between 1940 and 1954, we instead use all rented dwellings constructed before 1920 featuring a bathroom but no mansard and having three to four rooms.[253] We use the number of dwellings with three to four rooms from the housing census in 1950 as weights.

250. "*Statistisches Amt des Kantons Basel-Stadt*", Table t09.3.04, last update December 4, 2018.

251. "*Geschichtete Zufallsstichprobe, die 5% der Miet- und Genossenschaftswohnungen mit 1-6 Zimmern umfasst*".

252. "*Statistisches Jahrbuch des Kantons Basel-Stadt*" (Volumes 1925 - 1970).

253. Data for two-room dwellings are missing for some years in between. The data are not given separately anymore for dwellings built between 1900 and 1920 as well as for dwellings built prior to 1920.

Between 1925 and 1940, rented dwellings are sampled to match the mixture of old and new flats in the Basel housing stock that have two to five rooms. We use bins by number of rooms and within rooms by number of mansards and build a weighted average using the number of dwellings in the housing census in 1930 as weights. Between 1920 and 1925, we use a sample of dwellings that was sampled repeatedly in all years since 1920 for the same number of rooms (2-5) and number of mansard (0-2+) bins. We use the number of dwellings by number of rooms and number of mansards from 1929 as weights.

For the period between 1890 and 1920, we use the rent index constructed by Curti (1981). Between 1890 and 1912, the author uses newspaper advertisements for three-room apartments without a mansard in blue-collar worker districts. After 1912, he instead relies on apartments advertised through a public institution.[254] He again uses only three-room apartments, but the apartments are sampled from the entire city of Basel. Next, he chains both indices and calculates three-year moving averages. Finally, Curti adjusts his index for the rent increase between housing censuses in 1910 and 1920. Details can be found in the respective publication.

Table 1.K.46 summarizes the components of our final rent index.

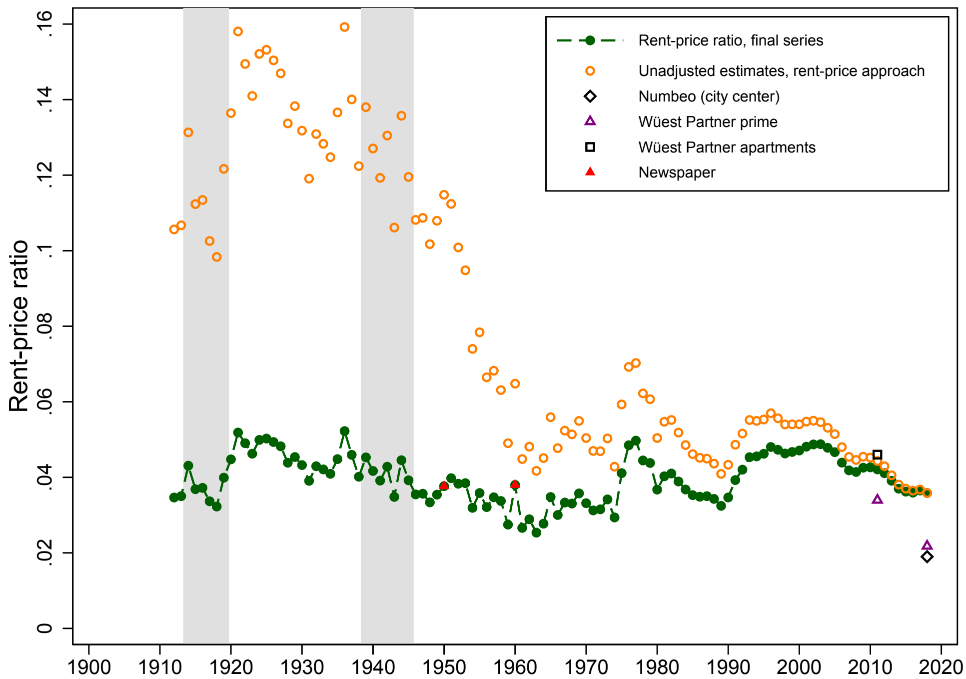

Figure 1.K.25. Basel: plausibility of rental yields

254. Rents published in the *Anzeiger des amtlichen Wohnungsnachweises*.

Table 1.K.46. Final rent index for Basel

PERIOD	SOURCE	DESCRIPTION
1890-1920	Curti (1981)	*Type(s) of dwellings*: Three-room apartments without a mansard; *Type of data*: Newspapers, advertised dwellings in public institution and housing census; *Method*: Three-year moving average adjusted with housing census data in 1910 and 1920.
1920-1925	Own compilation	*Type(s) of dwellings*: Constant sample of rented dwellings with two to five rooms; *Type of data*: Average rent; *Method*: Weighted average using constant weights.
1925-1940	Own compilation	*Type(s) of dwellings*: All rented dwellings with two to five rooms; *Type of data*: Average rent; *Method*: Weighted average using constant weights.
1940-1954	Own compilation	*Type(s) of dwellings*: Dwellings featuring a bathroom built before 1920 with three to four rooms without mansard; *Type of data*: Average rent; *Method*: Weighted average using constant weights.
1954-1970	Own compilation	*Type(s) of dwellings*: Dwellings featuring a bathroom built between 1900 and 1920 with two to four rooms; *Type of data*: Average rent; *Method*: Weighted average using constant weights.
1970-2018	Statistics Basel	*Type(s) of dwellings*: All rented dwellings with one to six rooms; *Type of data*: Random stratified sample covering 5% of all rented dwellings; *Method*: CPI rent index.

Rental Yield Series. In 2018, we used the net rental yield from *MSCI* as our main benchmark as described in the main paper. Applying the rent-price approach to this benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.25. We collected two additional benchmarks for 2018. First, we use the benchmark for the city-center of Basel from *Numbeo.com*, which we adjust to capture net rental yields by subtracting one-third following Jordà, Knoll, et al. (2019). Second, we use the prime yield ("*Spitzenrendite*") for Basel published by *Wüest Partner* in their Swiss market reports.[255] It is calculated as the net initial yield (net earnings/gross purchase price) for fully let prime properties at top locations. We adjust this yield for the vacancy rate in Basel from the same publication. Both benchmarks are somewhat below our main benchmark from *MSCI*. We use the benchmark by *Wüest Partner* as our alternative benchmark in the robustness section of the main paper.

255. "*Immobilienmarkt Schweiz* 2018|4"

We collected two types of historical benchmarks. First, we use the prime yield again adjusted for the vacancy rate and the net-cash-flow yield ("*Netttocashflowrendite*") for rental apartments published by *Wüest Partner* for 2011.[256] Our long-run series is again above the prime-yield, but somewhat below the net-cash-flow yield for apartments, which seems to be plausible.

Second, we collected rental yield benchmarks from newspaper advertisements for the years 1950 and 1960. We take the mean rental yield adjusted for all costs by subtracting one-third as in Jordà, Knoll, et al. (2019). It proved to be hard to find newspaper advertisements that featured a rental yield or a price-rent ratio for properties in the city of Basel, such that the sample size is very small. Considering the strict rent controls in Switzerland during this period,[257] the resulting values are still credible. These rent controls ensured that the variation in rental yields between properties was very small. Indeed, in a public report regarding the rent policy from 1950,[258] the authors describe that the rent in Switzerland was fixed relative to the value of the property. The report states that, after 1946, the outside capital invested into real estate was meant to yield a rate of 3.8%. This value closely matches the newspaper net rental yield benchmark of 3.87% for Basel in 1950.

As the strict rent controls in Switzerland after World War II until the 1960s both makes the newspaper benchmarks more credible and probably induces a bias in our unadjusted long-run rental yield series, we adjust our rental yield series to the newspaper benchmarks in 1950 and 1960. This results in the final rental yield series depicted as the green-circled line in Figure 1.K.25.

1.K.13.2 Bern

House Price Series. The long-run house price index for Bern is constructed using various publications by Statistics Bern. For most of the period, the data contain average property prices by district, such that we can control for sample shifts in the location of sold properties. For the earliest two decades, however, the data are only given for the entire city of Bern. Moreover, data for the most recent decades are missing. Hence, we again had to rely on a house price index provided by *Wüest Partner*. Below, we describe in more details how we construct our long-run index and which sources we use.

Like for Basel, we have to rely on private house price indices for the most recent period, because the publication of property prices by the statistical office stops in 2003. We again use a transaction price index for single-family houses provided by

256. Source: "*Wuest & Partner Immo-Monitoring 2012|2*" p. 173.

257. See e.g. the report "*Die Entwicklung des schweizerischen Mietrechts von 1911 bis zur Gegenwart*" by Helen Rohrbach (2014).

258. "*Die langfristige Neuordnung der Mietpreispolitik*", report of the "*Eidg. Priskontrollkommission (Sub- und Plenarkommission) zuhanden des Vorstehers des Eidg. Volkswirtschaftsdepartements*" (1950).

Wüest Partner. This index is constructed using hedonic methods and covers the MS-region Bern. The index starts in 1985.

Between 1933 and 1985, we use average prices of sold properties from statistical yearbooks.[259] These data cover all voluntary sales of developed lots within the city of Bern.[260] Average prices are given disaggregated by six constant districts ("*Stadtteile*"). We use these to calculate a chained Fisher-type stratification index following Eurostat (2013) using number of transactions by district as weights. In this way we control for locational shifts over time within the sample of sold developed lots. We use the price per transaction instead of the price per square meter, because the price per square meter results in a much noisier series. The reason seems to be that, during the earlier years, some still agrarian developed lots were sold in Bern as can be seen in the yearbooks. These lots had abnormally large areas, such that they bias the average price per square meter considerably. As the number of these lots sold is very low, the average price per transaction is biased considerably less. Data for the year 1939 are missing. We linearly interpolate this year to calculate housing returns.

For the period from 1912 to 1933, we instead had to rely on average price per sold developed lot within the entire city of Bern published by Statistics Bern.[261] We again use price per transaction, because the price per square meter is considerably more volatile probably due to the same reason as described above.

Table 1.K.47 summarizes the components of our final house price index.

Table 1.K.47. Final house price index for Bern

PERIOD	SOURCE	DESCRIPTION
1912-1933	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales within Bern from yearbooks; *Method*: Average price per transaction.
1933-1985	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: All sales aggregated by district from yearbooks; *Method*: Stratification.
1985-2018	*Wüest Partner*	*Type(s) of dwellings*: Single-family houses; *Type of data*: Private transaction-level data; *Method*: Hedonic index.

Rent Series. To construct a long-run rent index for Bern, we use the rent index produced by Statistics Bern for nearly the entire time period. Only for the period

259. "*Statistisches Jahrbuch der Stadt Bern*" (Volumes 1930-1985).
260. "*Handänderungen von Freihandverkäufen bebauter Grundstücke*".
261. *Statistics Bern*: Table T 5.5.3: "*Stadt Bern: Handänderungen von Grundstücken durch Freihandkäufe 1912-2001*". Original sources according to published table: "*Statistisches Handbuch der Stadt Bern, Ausg. 1935; Statistisches Amt der Stadt Bern (1940): Bern und seine Entwicklung - graphisch-statistischer Atlas; Statistisches Jahrbuch der Stadt Bern, Bde. 1926-2002*".

between 1890 and 1914 we instead had to rely on the rent index published by Curti (1981). We describe the two sources in more detail below.

Statistics Bern already started to calculate a rent index in 1914. They kindly provided the yearly rent index starting in 1940 directly to us. For the period between 1914 and 1940, we use the same index from one of their contemporaneous publications.[262] The index is calculated using a chained repeated rent approach. This means that the rent for the same dwellings is compared in two consecutive years. The statistical office adapts the rent changes on dwelling-level for major renovations. In 2018, approximately 2,000 rented dwellings were sampled and the number of missing values is below 5%. The repeated rents are weighted according to the share in the total stock of all rented dwellings to calculate the final index. We use November values from 1950 onward, May values between 1940 and 1950 and yearly values before 1940.

Prior to World War I, we use the rent index constructed by Curti (1981). Between 1890 and 1912, the author uses newspaper advertisements for three-room-apartments without a mansard in blue-collar worker districts. After 1912, he instead relies on aggregate rents published by Bern's housing office ("*städtisches Wohnungsamt*"). He again uses only three-room apartments, but this time pooled for apartments with and without a mansard and for the entire city. Next, he chains both indices and calculates three-year moving averages. The resulting index is used to interpolate between housing census data in 1896, 1913 and 1920 to build a final rent index, which matches the overall rent increase of the entire rented residential housing stock. We use this final index between 1890 and 1914.

Table 1.K.48 summarizes the components of our final rent index.

Table 1.K.48. Final rent index for Bern

PERIOD	SOURCE	DESCRIPTION
1890 - 1914	Curti (1981)	*Type(s) of dwellings*: Three-room apartments; *Type of data*: Newspapers, aggregate data from housing office; *Method*: Three-year moving average adjusted with housing census data in 1896, 1913 and 1920.
1914 - 2018	Statistics Bern	*Type(s) of dwellings*: All rented dwellings; *Type of data*: Micro-level sample of repeated rents; *Method*: Repeated rent index.

Rental Yield Series. As our main benchmark in 2018, we use the net rental yield from *MSCI* as described in the main paper. Applying the rent-price approach to this benchmark results in the unadjusted long-run net rental yield series depicted as

262. "*Statistisches Jahrbuch der Stadt Bern*" (1940) p.106.

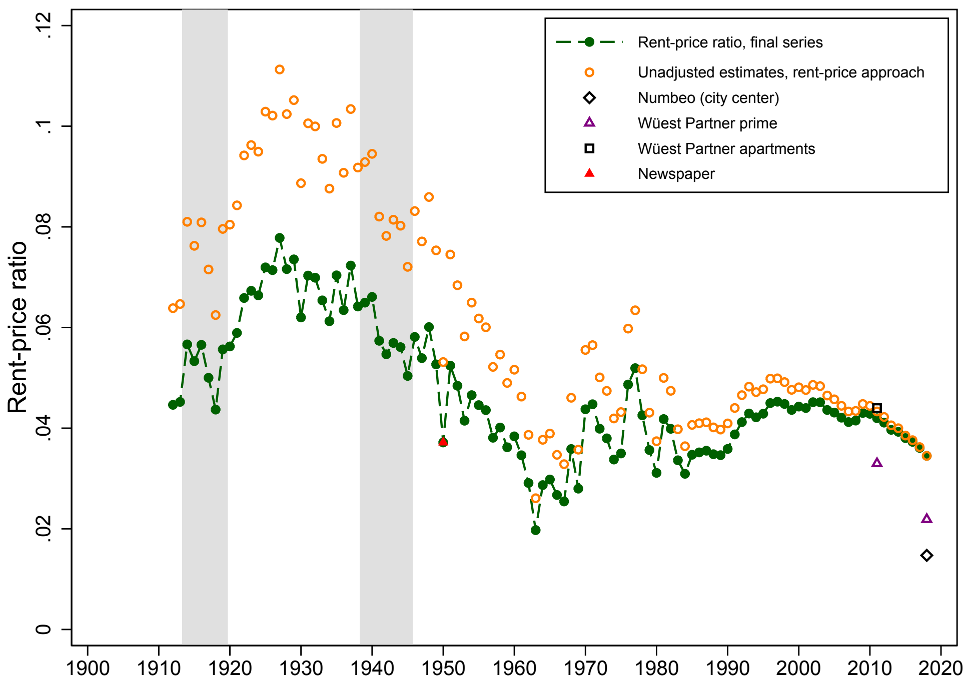

Figure 1.K.26. Bern: plausibility of rental yields

orange circles in Figure 1.K.26. We collected two additional benchmarks for 2018. First, we use the gross rental yield for the city-center of Bern from *Numbeo.com*. We adjust this benchmark to capture the net rental yield by subtracting one-third following Jordà, Knoll, et al. (2019). Second, we use the prime yield ("*Spitzenrendite*") for Bern published by *Wüest Partner* in their Swiss market reports.[263] It is calculated as the net initial yield (net earnings/gross purchase price) for fully let prime properties at top locations. We adjust this yield for the vacancy rate in Bern given in the same publication. Both benchmarks are somewhat below our main benchmark from *MSCI*. We use the benchmark by *Wüest Partner* as our alternative benchmark in the robustness section of the main paper.

We additionally collected two types of historical benchmarks. First, we use the prime yield adjusted for the vacancy rate and the net-cash-flow yield ("*Netttocashflowrendite*") for rental apartments published by *Wüest Partner* for 2011.[264] Our long-run series is again above the prime-yield, but very close to the net-cash-flow yield for apartments.

263. "*Immobilienmarkt Schweiz 2018|4*"
264. Source: "*Wuest & Partner Immo-Monitoring 2012|2*" p. 173.

Second, we collected rental yield benchmarks from newspaper advertisements for the year 1950. We take the median rental yield adjusted for all costs by subtracting one-third as in Jordà, Knoll, et al. (2019). It proved to be hard to find newspaper advertisements that featured rental yields or price-rent ratios for properties in Bern, such that the sample size is very small. Considering the strict rent controls in Switzerland during this period,[265] the resulting values are still credible, as described above for Basel. The resulting net rental yield from the newspaper advertisements is with 3.72% again very close to the rate of 3.8% published in the rent policy report for Switzerland from 1950.

Like we did for Basel, we adjust our rental yield series to the newspaper benchmarks in 1950, because the strict rent controls until the 1960s probably bias our unadjusted rental yield series. The final rental yield series is plotted as the green-circled line in Figure 1.K.26.

1.K.13.3 Zurich

House Price Series. To construct the long-run house price index for Zurich, we use average sales prices of developed properties published by Statistics Zurich throughout. For the entire period, the data are disaggregated by county or even district. This allows us to control for sample shifts in the location of sold properties using stratification methods.

Starting in 1905, we use data by Statistics Zurich published in statistical yearbooks.[266] The yearbooks publish average prices of all voluntary sales of developed lots within Zurich.[267] Average prices are given disaggregated by 11 to 12 counties ("*Kreise*") and later on even by 34 statistical districts ("*statistische Quartiere*").[268] We use the data by counties until 1984 and from there on the data by statistical district. Moreover, we use price per transaction instead of the price per square meter for the earlier period, because the price per square meter results in a noisier series. The reason seems to be that, during the earlier years, some agrarian developed lots ("*Landwirtschaftliche Bauten*") were sold in Zurich, as shown in the statistical yearbooks. These lots had abnormally large areas but low prices, such that the average price per square meter is biased and fluctuates with the number and size of these types of lots sold. As the number of these lots sold per year was very low in general,

265. See e.g. the report "*Die Entwicklung des schweizerischen Mietrechts von 1911 bis zur Gegenwart*" by Helen Rohrbach (2014).

266. "*Statistisches Jahrbuch der Stadt Zürich*" (Volumes 1905-2017); after 2017, no statistical yearbook was published anymore, but Statistics Zurich kindly provided the necessary data.

267. "*Handänderungen von Freihandverkäufen bebauter Grundstücke*".

268. Until 1989, the counties 11 and 12 are aggregated in the data. This implies, that the number of counties is only 11 until 1989 and the number of statistical districts only 32, because county 12 was counted as only one statistical district. Counties 9 to 11 are missing before 1934, because these counties were incorporated into Zurich in this year. When comparing 1933 to 1934, we therefore exclude these counties as well, such that only the same counties are compared for consecutive years.

the average price per transaction is biased considerably less. In 1984, Zurich was already heavily urbanized and no agrarian lots were transacted anymore. This category is even excluded from the yearbooks from 1985 onward. Hence, using price per square meter arguably leads to a less biased series from then on, because it additionally controls for the size of developed lots that are sold. By switching to statistical district-level data, we also ensure that transacted properties within each stratum are even more similar over time. Moreover, the data on number of transactions are missing in the yearbooks from 1989 onward. For both sub-periods we build chained Fisher-type stratification indices following Eurostat (2013). When using price per transaction we use the number of transactions per stratum as weights and for price per square meter the total area transacted in square meters by stratum.

Table 1.K.49 summarizes the components of our final house price index.

Table 1.K.49. Final house price index for Zurich

PERIOD	SOURCE	DESCRIPTION
1905-1984	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: Price per transaction of all sales aggregated by counties from yearbooks; *Method*: Stratification.
1984-2018	Own compilation	*Type(s) of dwellings*: All developed lots; *Type of data*: Price per square meter of all sales aggregated by statistical district from yearbooks; *Method*: Stratification.

Rent Series. Our long-run rent index for Zurich is composed of the rent index produced by Statistics Zurich for nearly the entire time period. Only for the period between 1890 and 1914 we instead rely on the rent index published by Curti (1981). We describe the two sources in more detail below.

Statistics Zurich already started to calculate a rent index in 1914. The data from 1940 onward are available on their website.[269] For the earlier period, we use the same index from one of their contemporaneous publications.[270] The index is calculated to construct a city-level CPI index. It measures rent developments for one to six room dwellings within the city of Zurich. For details about the methodology used please refer to Statistics Zurich. We use yearly averages for the monthly index throughout.

Prior to World War I, we use the rent index constructed by Curti (1981). Between 1890 and 1910, the author uses newspaper advertisements for three-room apart-

269. "*Statistik Stadt Zürich*", "*Zürcher Index der Konsumentenpreise, Mietpreisindex*"; https://www.stadt-zuerich.ch/prd/de/index/statistik/themen/bauen-wohnen/mietpreise/mietpreisindex/mietpreisindex.html.

270. "*Statistisches Jahrbuch der Stadt Zürich*" (1950) p.67.

ments without a mansard in blue-collar worker counties.[271] From 1908 to 1920, he additionally uses statistical data about dwellings advertised through a public institution.[272] He again uses only three-room apartments without a mansard in the same counties. The author chains both resulting indices. For the years the series overlap, Curti builds a weighted average of both indices. Next, he calculates three-year moving averages. The resulting yearly index is used to interpolate between housing census data in 1896, 1910 and 1920 to build a final rent index, which matches the overall rent increase of the entire rented residential housing stock. For details please refer to the aforementioned source. We use the final index between 1890 and 1914.

Table 1.K.50 summarizes the components of our final rent index.

Table 1.K.50. Final rent index for Zurich

PERIOD	SOURCE	DESCRIPTION
1890-1914	Curti (1981)	*Type(s) of dwellings*: three-room apartments; *Type of data*: Newspapers, statistical data about advertised dwellings; *Method*: three-year moving average adjusted with housing census data in 1896, 1910 and 1920.
1914-2018	Statistics Zurich	*Type(s) of dwellings*: Rented residential dwellings with 1-6 rooms; *Type of data*: From CPI index construction; *Method*: CPI rent index.

Rental Yield Series. As described in the main paper, we use the net rental yield from *MSCI* as our main benchmark in 2018. Applying the rent-price approach to this benchmark results in the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.27. We collected two additional benchmarks for 2018. First, we use the gross rental yield for the city-center of Zurich from *Numbeo.com*. We adjust this benchmark to capture the net rental yield by subtracting one-third following Jordà, Knoll, et al. (2019). Second, we use the prime yield ("*Spitzenrendite*") for Zurich published by *Wüest Partner* in their Swiss market reports.[273] It is calculated as the net initial yield (net earnings/gross purchase price) for fully let prime properties in top locations. We adjust this yield for the vacancy rate given in the same publication. Both benchmarks are somewhat below our main benchmark from MSCI. Like for the other Swiss cities, we use the benchmark by *Wüest Partner* as our alternative benchmark in the robustness section of the main paper.

We additionally collected three types of historical benchmarks. First, we use the prime yield adjusted for the vacancy rate and the net-cash-flow yield ("*Nettto-*

271. The author relies on advertisements in the later counties 3, 4, 5 and 6.
272. "*Der städtische Wohnungsnachweis*".
273. "Immobilienmarkt Schweiz 2018|4"

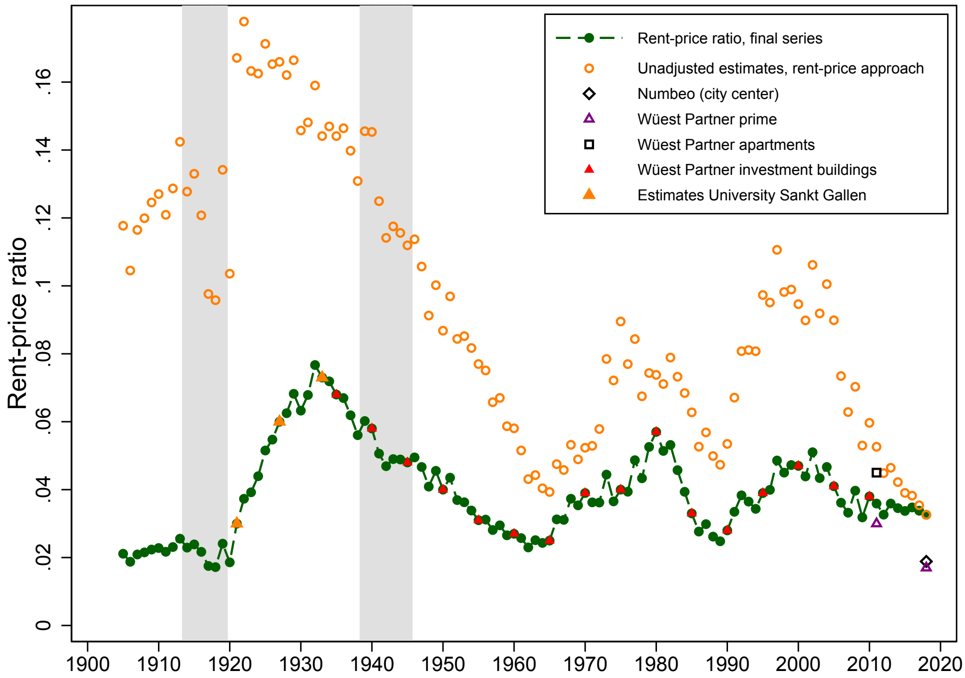

Figure 1.K.27. Zurich: plausibility of rental yields

cashflowrendite") for rental apartments published by *Wüest Partner* for 2011.[274] Our long-run series is again above the prime-yield, but close to the net-cash-flow yield.

Second, another chapter within same publication by *Wüest Partner* provides estimates of net-cash-flow yields ("*Netttocashflowrendite*") for residential investment buildings within the city of Zurich between 1931 and 2010.[275] To construct these estimates, *Wüest Partner* combined their transaction-level sales data with rent estimates calculated using data from Statistics Zurich. We collected rental yield benchmarks for five-year steps using these estimates. The benchmarks are plotted as red triangles in Figure 1.K.27. They show the same cyclical behavior as our unadjusted long-run series, but the values are considerably below our series.

Third, we collected net rental yield benchmarks for Zurich from a publication by the University of Sankt Gallen.[276] They estimate net rental yields of investment buildings for the years 1921, 1927 and 1933. These estimates are depicted as orange triangles in Figure 1.K.27. The resulting benchmarks are also considerably below our unadjusted long-run series.

274. Source: "*Wuest & Partner Immo-Monitoring 2012|2*" p. 173.
275. Source: "*Wuest & Partner Immo-Monitoring 2012|2*" chapter 5 p. 67.
276. "*Die Wohn- und Siedlungspolitik der Kantone und Gemeinden*", in "*Veröffentlichungen der Schweizerischen Verwaltungskurse an der Handels-Hochschule Sankt Gallen*", p.24.

As strict rent controls until the 1960s probably bias our unadjusted rental yield series and all historical benchmarks prior to 2011 are considerably lower compared to our series, we benchmark our series to all benchmarks prior to 2011. Especially the benchmarks for investment buildings by *Wüest Partner* seem to be more reliable compared to our series, because they use micro-level sales data to estimate net rental yields. Moreover, the estimates by *Wüest Partner* are in line with the earlier estimates by the University of Sankt Gallen. Hence, we infer that also the earlier estimates are more plausible compared to our unadjusted series. The final rental yield series is plotted as the green-circled line in Figure 1.K.27.

1.K.14 United Kingdom

The five largest cities in the United Kingdom in 1900 were in descending order: London, Manchester, Birmingham, Glasgow and Liverpool. All of them encompassed more than one percent of the country population in 1900. However, to the best of our knowledge, no complete long-run data on either house prices or rents exists for any of these cities. Consistent house price data in the U.K. on the local planning authority level starts in 1974.[277] Consistent rent data only starts in 1997 for private registered providers and in 2010 for private market rents.[278] Before, we are only aware of city-level data for London. We use these data and compiled additional new series to construct housing return data for London.

1.K.14.1 London

House Price Series. A house price index is available for London from 1968 onward from the *HM Land Registry*. For later periods, there are also other indices available, for example from *Rightmove*, which uses asking prices, *Nationwide* or *Halifax*, which use their own mortgage approvals data. We rely on the *HM Land Registry* index for consistency and as it is based on a large sample using actual transaction prices (see below).[279] Before, there only exist a house price index between 1895 and 1939 from Samy (2015). To the best of our knowledge, there does not exist any house price index for London for other time periods. We therefore built a hedonic house price index for London for the intermediate period using newspaper advertisements.

The *HM Land Registry* publishes together with the *Office for National Statistics* (ONS) a house price index for London under the title *UK House Price Index* (UK HPI). The index is based on all residential real estate transactions collected as part of the official registration process that are sold for full market value. Using this micro-level

277. See **Hilber2015**.
278. See Hilber and Mense (2021).
279. For a comparison of the indices, data and methods used please refer to the documentation of the UK House Price Index, for example here: https://www.gov.uk/government/publications/about-the-uk-house-price-index/comparing-house-price-indices-in-the-uk.

data the ONS calculates a hedonic house price index. We use the index based on all property types for the London region, which is equivalent to the Greater London area. We use yearly averages of the monthly index.

Samy (2015) calculates hedonic house price indices using actual house price data from (1) the yearbooks of the *London Auction Mart* (1895-1922) and (2) the mortgage registers of the *Co-operative Permanent Building Society* (1919-1939). We use the double imputation chained fisher hedonic index presented in the paper. We use the index based on the Auction Mart data until 1922, because it has a larger sample size and focuses more on properties to let, such that it will be more comparable to the rent index.

To fill the gap between the two indices, we use house asking price data from newspaper advertisements to calculate hedonic house price indices. We relied on digitized newspapers, which were available online, and focused on local newspapers as these covered a more standard market segment. Since the availability of the newspapers and the number of advertisements within each newspaper changed over time, we had to use three different newspapers. We collected advertisements for freeholds and leaseholds, which met the following three conditions: First, the advertisement contained information about the price (freehold or leasehold), the type of residence (house, flat...), some indicator of dwelling size (number of rooms or at least number of bedrooms) and the location. Second, we could match the location to a London borough (which we were able to do for nearly all advertisements we collected). Third, the leasehold period (if given) was at least 20 years. Applying these conditions we collected 522 observations from the *West London Observer* (WLO, 1946 - 1954), 3,163 observations from the *Norwood News* (NN, 1946 - 1962) and 521 observations from the *Kensignton Post* (KP, 1962 - 1969). We construct two different indices, one by pooling the advertisements from the first two newspapers together and one only using data from the KP, as the KP covers a significantly different market segment compared to the other two newspapers (unfortunately, data from the NN were not available after 1962). As we collected data for 1962 from both the NN and the KP, we are able to chain the indices. For both periods, we constructed five-year rolling-window time-dummy hedonic indices using a non-parametric approach,[280] because it best fits the trade-off between having enough observations to use a sufficient number of control variables still generating enough statistical power and the time flexibility of coefficients.[281] To construct the index we regressed the log house price on: year dummies; dummies for the total number of all normal rooms (bedrooms + lounges + receptions + sculleries or simply number of rooms if only totals are given); dummies for the number of bathrooms (if given,

280. We follow the methodology used in Keely and Lyons (2020).

281. This way, we only need to assume that coefficients of the hedonic regressions are stable for a period of five years, which might be much more realistic than assuming that coefficients were stable for the entire time period.

otherwise category 0 for missing); dummies for the number of kitchens (if given, otherwise category 0 for missing); a dummy for each London borough; a dummy for the house type (house, flat or maisonette); a separate dummy when the advertisement states the house has at least one garden, garage or is furnished; a dummy for leaseholds, an interaction term of the leasehold dummy and the leasehold period (if given and less than 100 years), a dummy for very long leaseholds (100+ years) and a dummy for leaseholds with a missing period.[282] We do not control for ground rent, as this information seems to be missing in most cases. If we include a control for ground rent, however, the resulting index is very similar.

To link the final house price index for London from 1939 to 1946, we use the Land Registry Index described in Knoll, Schularick, and Steger (2017). According to the authors, the data used to construct the Land Registry Index have to a very large extent been collected from properties in the London area. Using this index, the house price series increases slightly less during World War II compared to the national series. This might, however, be realistic considering that real house prices did not fall in London in the late 1920s and early 1930s, but did decrease considerably during that period for the national house price index in Knoll, Schularick, and Steger (2017). As this index still covers properties outside London, we do not use it to impute the missing years during World War II, but only to link the indices before and afterwards.

Table 1.K.51. Final house price index for London

PERIOD	SOURCE	DESCRIPTION
1895-1939	Samy (2015)	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Micro-data from the London Auction Mart (1895-1922) and the mortgage registers of the Co-operative Permanent Building Society (1923 - 1939) ; *Method*: Hedonic index.
1946-1968	Own compilation	*Type(s) of dwellings*: Houses, flats or maisonettes; *Type of data*: Newspaper advertisements from the *West London Observer*, the *Norwood News* and the *Kensington Post*; *Method*: Hedonic index.
1968-2018	HM Land Registry	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Transaction-level data from the HM Land Registry; *Method*: Hedonic index.

Rent Series. Rent indices for London are available from Eichholtz, Korevaar, and Lindenthal (2019) for the period between 1870 and 1959 and from the *Office for National Statistics* (ONS) from 1997 onward. Apart from these periods, to the best

282. For the leasehold variables see also Samy (2015).

of our knowledge, there only exists a rent series for private registered providers. As these companies face a legal rent ceiling,[283] we do not want to rely on these data to measure rent growth. Instead, we build hedonic rent indices for the period in between relying on advertisements from a range of different newspapers.

Eichholtz, Korevaar, and Lindenthal (2019) build a rent index for London between 1225 and 1959. We use this index from 1870 onward. Between 1870 and 1903, the authors build a repeat sales index using data from Clark (2002). From 1903 to 1909, the authors rely on the hedonic rent index in Samy (2015). Afterwards, they again build a repeat sales index using data on residential rent prices from seal books from the archives of *Trafalgar House Developments Ltd*. For details, please refer to the aforementioned paper.

For the most recent period, we rely on two different rent indices for London compiled by the ONS. Between 1997 and 2005, we use an index compiled using data on regional private rent collected during construction of the consumer price index. The index was published in 2019 in response to an ad-hoc request and covers the London region. From 2005 onward, we use the *Index of Private Housing Rental Prices* for the London region. This index is based on administrative data from the *Valuation Office Agency*. For both indices, we use yearly averages of the monthly index values.

To fill the gap between 1959 and 1997, we collected rent data from newspaper advertisements. We had to rely on different newspapers over time, as most newspapers had only been digitized for specific time periods and not at all times published a sufficient number of housing rent advertisements. We tried to rely on local newspapers as much as possible. As different newspapers arguably covered different housing market segments, we construct separate rent indices for different periods: 1946 to 1962 using the *London Observer* and the *Norwood News*, 1962 to 1969 using the *Kensington Post*, 1966 to 1992 using *The Times* and 1989 to 1997 using the *Hammersmith & Shepherds Bush Gazette*.

From every newspaper we only included advertisements for rents that fulfilled the following three conditions: First, the advertisement contained information about the rent (or income), the type of the residence (house, flat, room, etc.), some indicator of dwelling size (number of rooms or at least number of bedrooms) and the location. Second, we could match the location to a London borough (which we were able to do for nearly all advertisements we collected). Third, the leasehold period (if given) was at least five years. Fourth, we exclude large apartment buildings (when the place had more than three bathrooms, more than two kitchens or more than ten rooms). For all aforementioned time periods, we constructed five-year rolling-window time-dummy hedonic indices using a non-parametric approach,[284] because it best fits the trade-off between having enough observations to use a sufficient number of control variables still generating enough statistical power and the time flexi-

283. See Hilber and Mense (2021).
284. We follow the methodology used in Keely and Lyons (2020).

bility of coefficients.[285] To construct the hedonic index we regressed the log yearly rent on: year dummies; dummies for the total number of all normal rooms (bedrooms + lounges + receptions + sculleries or simply number of rooms if only totals were given); dummies for the number of bathrooms (if given, otherwise category 0 for missing); dummies for the number of kitchens (if given, otherwise category 0 for missing); a dummy for each London borough; a dummy for the house type (house, flat, room or maisonette); a separate dummy when the advertisement states the residence has at least one garden, garage or is furnished; and a separate dummy for whether the advertised residence was also for sale or lease. We do not control for ground rent, as this information seems to be missing in most cases and is less relevant for rents. If we include a control for ground rent, however, the index stays virtually the same.

For the period between 1946 and 1969, we use 219 observations from the *West London Observer* (WLO, 1946 - 1954), 1,742 observations from the *Norwood News* (NN, 1946 - 1962) and 1,176 observations from the *Kensington Post* (KP, 1962 - 1969). We construct two different indices, one pooling the advertisements from the first two newspapers and one with the data from the KP, as the KP covers a significantly different market segment compared to the other two newspapers (unfortunately, data from the NN was not available after 1962).

For the period between 1966 and 1992, we use rent data from advertisements out of the newspaper *The Times* to calculate a hedonic rent index. After 1972, we only collected data for flats, so that observations are more comparable over time.[286] Prior to 1972 we did not find enough observations to focus exclusively on flats.[287] Moreover, we exclude all duplicates and all very cheap[288] and very expensive[289] observations. After the data cleaning we were able to use 4,269 observations (1966-1992, with most between 1969 and 1989).[290]

For the period between 1989 and 1997, we use rent data from advertisements out of the newspaper *Hammersmith & Shepherds Bush Gazette*. We exclude all dupli-

285. This way, we only need to assume that coefficients of the hedonic regressions are stable for a period of five years, which might be much more realistic than assuming that coefficients were stable for the entire time period.

286. As *The Times* was a supra-regional newspaper, housing advertisements sometimes covered luxurious or special real estate, which is less comparable over time. This problem is larger for houses, so that we focused on flats when possible. The other newspapers were local newspapers, which covered more comparable housing segments over time.

287. We, however, only use the index with all buildings constructed using *The Times* for three years between 1969 and 1972.

288. Below 100 pounds per year, mostly typos or not referring to residence.

289. Above 30,000 pounds per year, mostly especially luxurious flats, which are harder to compare over time.

290. There did not exist any data for 1982, so we have to interpolate this year - we use the rent component of the national CPI from Eichholtz, Korevaar, and Lindenthal (2019) for interpolation).

cates and two very expensive[291] observations. Afterwards, we could rely on 3,519 observations.

Table 1.K.52 summarizes the components of our final rent index. For the period between 1946 and 1959 we rely on the index by Eichholtz, Korevaar, and Lindenthal (2019), because it used actual rent data instead of asking prices and is based on a large number of observations. We only use the index we constructed using *The Times* from 1969 to 1989, as it arguably covers a special market segment and is based on fewer observations per year compared to the other newspaper indices for the overlapping periods. As we collected data on overlapping years, we are able to chain the indices.

Table 1.K.52. Final rent index for London

PERIOD	SOURCE	DESCRIPTION
1870-1959	Eichholtz, Korevaar, and Lindenthal (2019)	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Micro-level data from various sources; *Method*: Repeat sales/hedonic index.
1959-1969	Own compilation	*Type(s) of dwellings*: Houses, flats, maisonettes or rooms; *Type of data*: Newspaper advertisements from the West London Observer, the Norwood News and the Kensington Post; *Method*: Hedonic index.
1969-1989	Own compilation	*Type(s) of dwellings*: Houses, flats, maisonettes or rooms; after 1972 only flats *Type of data*: Newspaper advertisements from the Times; *Method*: Hedonic index.
1989-1997	Own compilation	*Type(s) of dwellings*: Houses, flats, maisonettes or rooms; *Type of data*: Newspaper advertisements from the Hammersmith & Shepherds Bush Gazette; *Method*: Hedonic index.
1997-2005	ONS	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Data from CPI construction; *Method*: Stratification.
2005-2018	ONS	*Type(s) of dwellings*: All kinds of residential dwellings; *Type of data*: Administrative data from the Valuation Office Agency; *Method*: Stratification.

Rental Yield Series. Our main benchmark for London is taken from *MSCI*, as described in the main paper. This benchmark is reasonably close to the alternative benchmark collected from *Numbeo.com*, which we use in the robustness section of the main paper. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.28.

291. Above 50,000 pounds per year, large outliers.

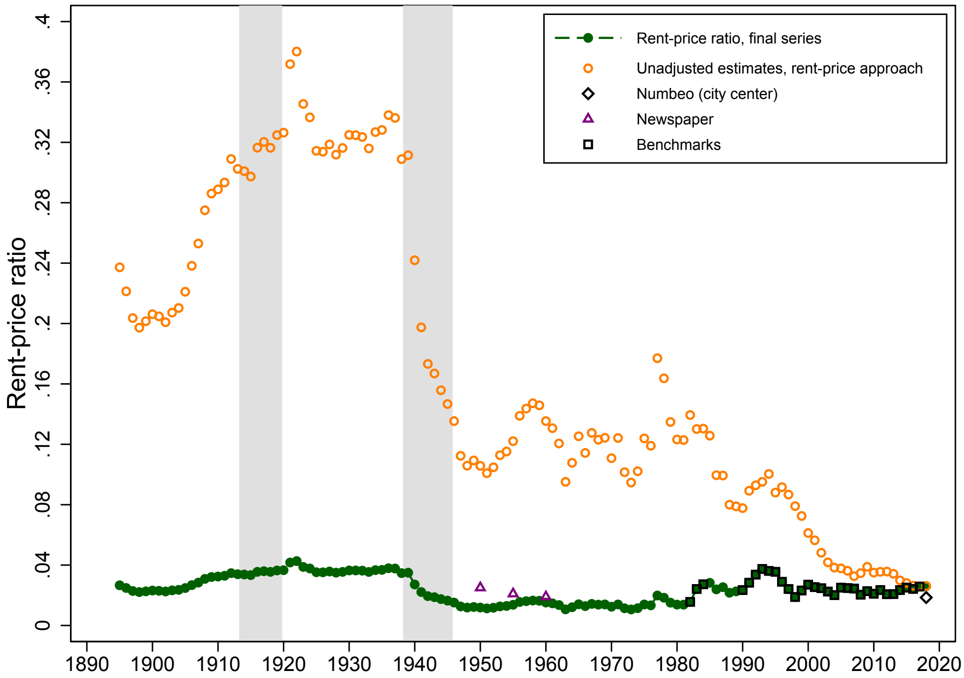

Figure 1.K.28. London: plausibility of rental yields

According to this series rental yields were unrealistically high in the pre-2000 period. The reason is that house prices and rents show fundamentally different trends after 1997. The rent series for London from ONS might, however, be mismeasured, because these are still experimental series and private market rents cover only a small market segment in London. As *MSCI* has extensive data coverage for London, we decided to instead rely on the MSCI-series for the period it is available, meaning for the years 1982-1984 and 1990-2018. Only for the period in between and before do we use the rent-price approach.

This gives us the adjusted final rent-price ratio series—the green-circled line in Figure 1.K.28.

We collected additional benchmarks for gross rental yields from the newspaper data for 1950 (27 observations), 1955 (12 observations) and 1960 (8 observations). Following Jordà, Knoll, et al. (2019) we subtract one-third of these to get net rental yields. Median values are depicted in Figure 1.K.28. These values are reasonably close to our final rental yield series. As sellers have an incentive to overstate rental yields in newspaper advertisements, it does not come as a surprise that the benchmarks from the newspapers are somewhat above our series. For this reason and because benchmarks from newspapers are noisy as well, we do not adjust our rental yield series to these benchmarks.

1.K.15 United States

In 1900 the four largest cities in the United States were in descending order: New York City, Chicago, Philadelphia and Boston. All of these cities represented more than 1% of the national total population in 1900. New York was by far the largest, with almost 6% of the national population.

Although there exist housing series for the other three cities, they cover only specific historical periods in the twentieth century. Fishback and Kollmann (2014) use data from the Home Owners' Loan Corporation (HOLC) surveys, which they combine with census data, to build house price indices for the period between 1920 and 1940 for various different cities in the US. Unfortunately, for all these cities further house prices do not exist until the 1970s, when the Federal Housing Finance Association (FHFA) series start. For Chicago there is extensive literature on the long-run evolution of land values (Ahlfeldt and McMillen, 2018). However, to the best of our knowledge long-run series on house prices or returns do not exist. As such, we were able to construct a continuous long-run housing series only for New York.

1.K.15.1 New York

House Price Series. Nicholas and Scherbina (2012) built a house price index for Manhattan for the period between 1920 and 1939. The authors use 7,500 transaction prices from various issues of the publication *Real Estate Record and Builders' Guide*, alongside information on the characteristics of the properties transacted, to build a hedonic house price index. The data are mostly composed of transactions of multi-family houses, but also include apartment transactions. The hedonic indices control for type of dwelling (tenement, dwelling or loft), size (number of stories and square footage), location (neighborhood), construction material (brick, stone or other) and additional features (e.g. whether the property has a basement). The authors build a constant-relative-value hedonic index, which makes the assumption that the relative prices of the characteristics remain constant over time, and an adjacent-period hedonic index, which allows for time-varying prices of the characteristics. Since both indices produce very similar results we use the former for our long-run New York index.

For the period between 1940 and 1975 we use a house price index based on yearly median prices of multi-family houses (tenements) in Manhattan, which was published in Barr, Smith, and Kulkarni (2018). The transaction data were taken from the annual volumes published by the Real Estate Board of New York. Although the index does not control for quality adjustments over time, Barr, Smith, and Kulkarni (2018) show that it strongly correlates with a hedonic land value index of Manhattan in this period.

For the period after 1975 we rely on the house price county indices from the Federal Housing Finance Agency (FHFA) to create a house price index for the city of New York. The county indices are built using transaction and appraisals data on single-family house purchases and refinances from a mortgage transactions data set, which contains almost the whole universe of mortgages acquired or guaranteed by Fannie Mae or Freddie Mac. The index is based on a repeat-sales approach, which uses a weighted least squares (WLS) approach to handle heteroskedasticity due to constant differences between transactions with different holding periods. For more details about the index please refer to Bogin, Doerner, and Larson (2018). To construct the New York city-level index we use the house price indices for the following counties: Bronx, Kings, New York (Manhattan), Queens and Richmond. We aggregate the county-level indices by using a simple yearly average, since we did not have access to the volume of transactions at the county level.

To the best of our knowledge, there do not yet exist historical real estate indices for the city of New York covering the period between 1920 and 1975. By focusing on Manhattan, we think that we are still approximating the general house price evolution in New York relatively well, since Manhattan was throughout this period the most expensive part of New York and represented a large share of the housing transactions in the city of New York (Barr, Smith, and Kulkarni, 2018).

For the more recent period there are at least two additional indices for New York. The *S&P/Case-Shiller* index covers the Metropolitan Statistical Area (MSA) of New York since 1987 and the Zillow series on house values covers the city of New York since 1996. Zillow uses a hedonic approach to approximate the average value of residential housing in New York. To keep our geographical approach consistent over time, we decided not to use the index from *S&P/Case-Shiller*, which covers the complete MSA of New York. Additionally, we decided to use the FHFA indices, and not the housing values from Zillow, since the FHFA index covers a longer period of time and, as a result, reduces the number of different sources we are using.

Table 1.K.53 summarizes the components of our final house price index.

Table 1.K.53. Final house price index for New York

PERIOD	SOURCE	DESCRIPTION
1920-1939	Nicholas and Scherbina (2012)	*Type(s) of dwellings*: Multi-family houses and apartments ; *Type of data*: Transaction prices from Real Estate Record and Builders' Guide ; *Method*: Constant-relative-value hedonic index.
1940-1975	Barr, Smith, and Kulkarni (2018)	*Type(s) of dwellings*: Multi-family houses ; *Type of data*: Transaction prices from the Real Estate Board of New York; *Method*: Median price .
1975-2018	FHFA	*Type(s) of dwellings*: Single-family houses ; *Type of data*: Transaction prices and appraisals from FHFA mortgage data set; *Method*: Repeat-sales .

Rent Series. Our long-run rent series for New York is based on the residential rent component of the consumer price index (CPI) series constructed by the Bureau of Labour Statistics (BLS) for the MSA of New York for the period between 1914 and 2018. The residential rent component is based on data from the housing surveys conducted by the BLS on a frequent basis on a representative and rotating sample of households within the MSA. The rent series are quality-adjusted in the sense that extra charges or costs are taken into account to estimate the exact price change in rents. For a more detailed overview of the methods used please refer to Labour Statistics (2009).[292] We have two concerns about using the BLS series.

First, the series might under-estimate the impact of new rental contracts on the market, by focusing to a large extent on existing contracts.[293] This could create a wedge between our house price series, which is largely affected by new housing, and our rent series, which mostly relies on existing rental contracts. Second, the BLS series is meant to approximate the rent price growth in the New York MSA, while our house price series focuses on the city of New York.

To attenuate these concerns, we corrected the growth rates of the rental series from the BLS using the average rent values for the city of New York from the census for the years of 1950, 1960, 1970, 1980 and 1990. We do this by adjusting the BLS rent series for the growth in average rents in the city of New York using census data between 1950 and 1990. In Figure 1.K.29 we plot the original nominal rent series from BLS and the rent series we adjusted using the census data on average rents for New York. As expected, rent prices grew more in the city of New York than in the MSA of New York and the differences are quite significant over time. As a result, we use the adjusted rent series in our final rent series for New York. Table 1.K.54 summarizes the components of our final rent index.

As with house prices we are not aware of other historical quality-adjusted rental series for the city of New York. Since 2010 Zillow has published an average rental value for New York based on a hedonic approach using data on asking rents from online real estate ads. To keep a constant method over time, we decided not to use the rental values from Zillow, which nevertheless show a very similar trend for the overlapping years.

Rental Yield Series. Our main benchmark for New York is taken from *MSCI*, as described in the main paper. This benchmark is relatively high compared to the alternative benchmark we collected for 2018 from Demers and Eisfeldt (2021). According to Demers and Eisfeldt (2021) the net residential yield in the New York MSA in 2018 was 1.8%. However, our benchmark from *MSCI* is a bit lower than

292. Although the sampling and estimation methodologies of BLS have changed over time, the points made above have remained a consistent concern of the BLS.

293. A similar concern about the rent component of the CPI in European cities is raised in Eichholtz, Korevaar, Lindenthal, and Tallec (2020).

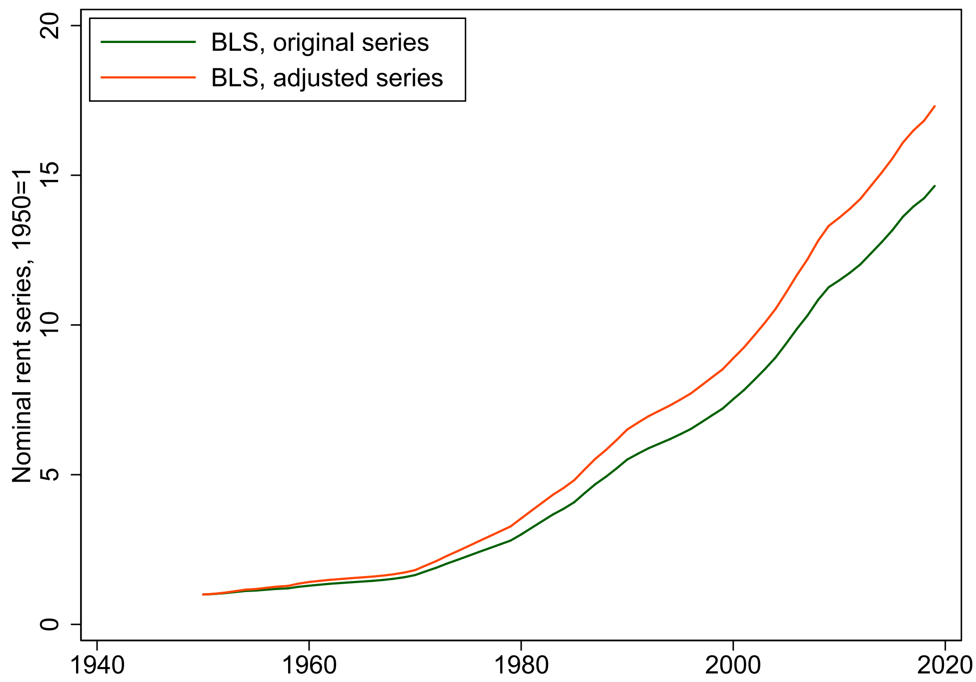

Figure 1.K.29. Nominal rent series for New York, 1950=1

Table 1.K.54. Final rent index for New York

PERIOD	SOURCE	DESCRIPTION
1914 - 2018	Own compilation	*Type(s) of dwellings*: Renter-occupied dwellings ; *Type of data*: Rental values from BLS housing surveys and census ; *Method*: Quality-adjusted rent series.

the benchmark from *Trulia.com* for 2018, which calculated the ratio of the median gross rent to median house price in the city of New York to be 0.0612. We then adjust for one-third costs to get an estimate of net yield of 4%. Applying the rent-price approach to our main benchmark gives us the unadjusted long-run net rental yield series depicted as orange circles in Figure 1.K.30.

Although we try to correct for geographical coverage differences in our rent and price series, there might still exist a wedge between the two series. This could lead to an under-estimation of the rent growth in the city of New York, since we are using the MSA-level rent series. In turn, this could bias our rental yield series upwards for the beginning of our sample. As a result we also collected a net yield benchmark from Demers and Eisfeldt (2021) for 1985. As can be seen from Figure 1.K.30 this benchmark lies substantially below our unadjusted series. As a result, we adjusted

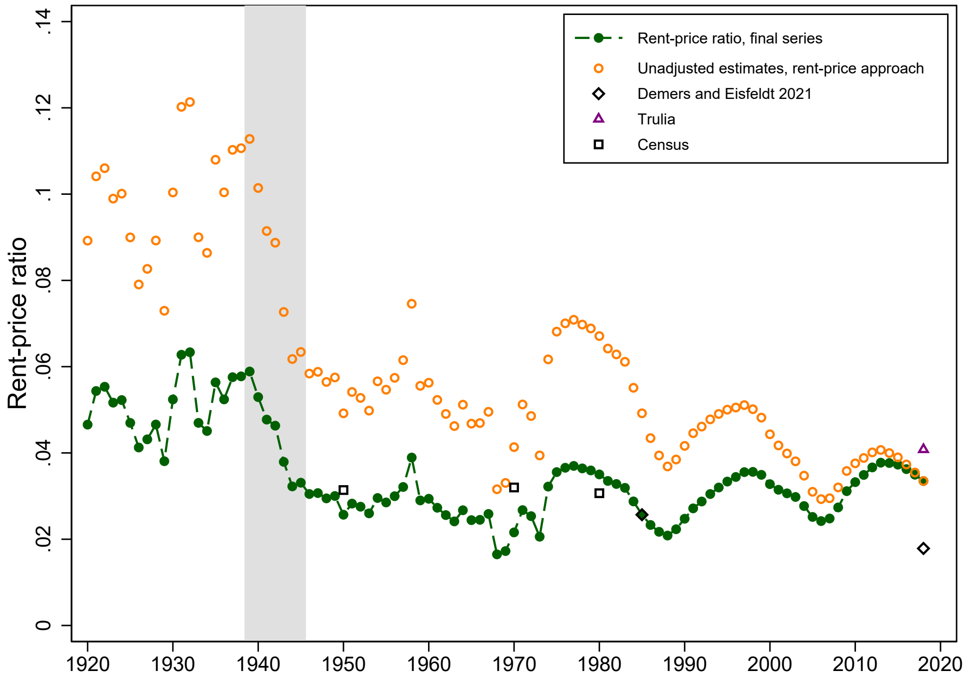

Figure 1.K.30. New York: plausibility of rental yields

our series to the benchmark. This gives us the adjusted final rental yield series—the green-circled line in Figure 1.K.30.

Additionally, we also build historical benchmarks using census data on mean house value and mean rent paid for the years 1950, 1970 and 1980.[294] We then adjust these estimates assuming one-third costs. As can be seen in Figure 1.K.30 these benchmarks lie below our unadjusted series, but very close to our adjusted series.

To the best of our knowledge, there do not exist reliable estimates of rent-price ratios for New York for the period before 1950. As a result, we cannot be sure that our estimated rental yield series is correct for this period. However, a study by Grebler (1955) using data on income properties in Manhattan shows that multi-family houses yielded a net rental income of about 9% in the second half of the 1920s, and a net rental income of 3.5% in the 1930s and 1940s.[295] The strong volatility in the estimates provided by the author question its accuracy. Since relatively few details are given about the methods, we do not use his estimates. Nevertheless, his data

294. Unfortunately, housing data for New York City are missing in the census of 1960.

295. The author collected data on rental income and expenses from the Real Estate Record and Builders' Guide.

seem to show a downward trend between 1930 and 1950, which our final series also reproduces.

References

1961 Census of Canada : Housing Vol. II. 1961. *1961 Census of Canada : Housing Vol. II.* Canada. Dominion Bureau of Statistics. [99, 103]

Abelson, Peter. 1985. "House and Land Prices in Sydney: 1925 to 1970." *Urban Studies* 22 (6): 521–34. DOI: 10.1080/00420988520080881. [90]

Abelson, Peter, and Demi Chung. 2005. "The Real Story of Housing Prices in Australia from 1970 to 2003." *Australian Economic Review* 38 (3): 265–81. DOI: 10.1111/j.1467-8462.2005.00373.x. [85–87, 90, 91]

Abildgren, Kim. 2018. *Danmarks Nationalbank 1818-2018*. Danmarks Nationalbank. [103]

Adam, Klaus, Oliver Pfäuti, and Timo Reinelt. 2020. "Falling Natural Rates, Rising Housing Volatility and the Optimal Inflation Target." Working paper. University of Bonn, and University of Mannheim, Germany. [66]

Ahlfeldt, Gabriel M, and Daniel P McMillen. 2018. "Tall buildings and land values: Height and construction cost elasticities in Chicago, 1870–2010." *Review of Economics and Statistics* 100 (5): 861–75. [207]

Ahlfeldt, Gabriel M., Stephan Heblich, and Tobias Seidel. 2021. "Micro-Geographic Property Price and Rent Indices." [116]

Ahlfeldt, Gabriel M., and Elisabetta Pietrostefani. 2019. "The Economic Effects of Density: A Synthesis." *Journal of Urban Economics* 111 (May): 93–107. DOI: 10.1016/j.jue.2019.04.006. [9]

Amaral, Francisco, Martin Dohmen, Ronan Lyons, and Carla Revilla. 2021. "The Cost of Housing over the Long Run: the Canadian View." *Unpublished Manuscript*, [95–102]

Ansell, Ben W. 2019. "The Politics of Housing." *Annual Review of Political Science* 22 (1): 165–85. DOI: 10.1146/annurev-polisci-050317-071146. [8]

Armona, Luis, Andreas Fuster, and Basit Zafar. 2018. "Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment." *Review of Economic Studies* 86 (4): 1371–410. DOI: 10.1093/restud/rdy038. [66]

Arundel, Rowan, and Cody Hochstenbach. 2019. "Divided Access and the Spatial Polarization of Housing Wealth." *Urban Geography*, 1–27. [8]

Association, Canadian Real Estate. 2021a. "MLS House Price Index." URL: https://www.crea.ca/housing-market-stats/mls-home-price-index/. [101]

Association, Canadian Real Estate. 2021b. "MLS HPI Home Price Index Methodology." [100]

Australian Bureau of Statistics. 1993. "6416.0 - House Price Indexes: Eight Capital Cities, Dec 1993." *https://www.abs.gov.au/AUSSTATS/abs@.nsf*, [86, 90]

Australian Bureau of Statistics. 2018. "Consumer Price Index: Concepts, Sources and Methods, 2018." *https://www.abs.gov.au/ausstats/abs@.nsf/mf/6461.0*, [86, 87, 90, 92]

Australian Bureau of Statistics. 2020. "Residential Property Price Indexes: Eight Capital Cities." *https://www.abs.gov.au/statistics*, [85, 90]

Badarinza, Cristian, and Tarun Ramadorai. 2018. "Home away from Home? Foreign Demand and London House Prices." *Journal of Financial Economics*, [35]

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy* 126 (6): 2224–76. [66]

Bank of Japan. 1986. "Bank of Japan: The First Hundred Years: Materials." [161]

Barr, Jason, Fred H Smith, and Sayali J Kulkarni. 2018. "What's Manhattan worth? A land values index from 1950 to 2014." *Regional Science and Urban Economics* 70: 1–19. [207, 208]

Becker, Richard A., Allan R. Wilks, Ray Brownrigg, Thomas P Minka, and Alex Deckmyn. 2018. “"maps" - Draw Geographical Maps.” *R-package*, [10]

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra. 2018. “House Prices and Consumer Spending.” *Review of Economic Studies* 85 (3): 1502–42. [5]

Berger, Knut. 2010. *1960 - 2010 50 Jahre Gutachterausschuss für Grundstückswerte in Berlin*. Gutachterausschuss für Grundstückswerte in Berlin. [119, 120]

Black, Duncan, and Vernon Henderson. 1999. “Spatial Evolution of Population and Industry in the United States.” *American Economic Review* 89 (2): 321–27. DOI: 10.1257/aer.89.2.321. [7]

Blanchard, Olivier J., and Lawrence F Katz. 1992. “Regional evolutions.” *Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution* 23 (1): [72]

Board, Toronto Regional Real Estate. 2021. “Historic Statistics TRREB MLS System Sales and Average Price.” URL: https://trreb.ca/files/market-stats/market-watch/historic.pdf. [95]

Bogin, Alexander, William Doerner, and William Larson. 2018. “Local House Price Dynamics: New Indices and Stylized Facts.” *Real Estate Economics* 47 (2): 365–98. DOI: 10.1111/1540-6229.12233. [70, 76, 208]

Bohlin, Jan. 2014. “A Price Index for Residential Property in Göteborg, 1875–2010.” *Historical Monetary and Financial Statistics for Sweden: House Prices, Stock Returns, National Accounts, and the Riksbank Balance Sheet* 1620: 25–62. [181]

Bonneval, Loïc, and François Robert. 2010. *Mesurer la rentabilité du placement immobilier. Le cas de l’immeuble de rapport à Lyon (1890-1968)*. Vol. 25, XXV-1. Éditions de l’EHESS. [109]

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. “Overreaction in Macroeconomic Expectations.” *American Economic Review* 110 (9): 2748–82. [65]

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2019. “Diagnostic Expectations and Stock Returns.” *Journal of Finance* 74 (6): 2839–74. [33, 65]

Bracke, Philippe. 2013. “How Long Do Housing Cycles Last? A Duration Analysis for 19 OECD Countries.” *Journal of Housing Economics* 22 (3): 213–30. DOI: 10.1016/j.jhe.2013.06.001. [56]

Brounen, Dirk, Piet Eichholtz, Stefan Staetmans, and Marcel Theebe. 2013. “Inflation Protection from Homeownership: Long-Run Evidence, 1814-2008.” *Real Estate Economics* 42 (3): 662–89. DOI: 10.1111/1540-6229.12023. [16]

Bureau, Census. 2018. “Metropolitan and Micropolitan Statistical Areas and Related Statistical Areas - States.” *Cartographic Boundary Files - Shapefile*, [57]

Busch, Dr. August. 1919. “Beiträge zur Statistik der Stadt Frankfurt am Main.” *Erstes Heft: Die Besiedlung des Frankfurter Stadtgebiets und die Befriedigung des Wohnungsbedürfnisses der Bevölkerung*, [13]

Canada, Statistics. 2021a. “Table 18-10-0004-11 Consumer Price Index, by geography, monthly, percentage change, not seasonally adjusted, provinces, Whitehorse and Yellowknife.” URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810020501. [102]

Canada, Statistics. 2021b. “Table 18-10-0205-01 New housing price index, monthly.” URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810020501. [95, 96, 101]

Canada, Statistics. 2021c. “Table 34-10-0133-01 Canada Mortgage and Housing Corporation, average rents for areas with a population of 10,000 and over.” [97, 98, 102]

Cannari, Luigi, Giovanni D’Alessio, and Giovanni Vecchi. 2016. “I prezzi delle abitazioni in Italia, 1927-2012.” *Questioni di Economia e Finanza (Banca d’Italia)* 333: [150–155, 157–159]

Case, Karl, and Robert Shiller. 1987. “Prices of Single Family Homes Since 1970: New Indexes for Four Cities.” Working paper. DOI: 10.3386/w2393. [126]

Case, Karl E., Robert J. Shiller, and Anne Thompson. 2014. “What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets—A 2014 Update.” [65]

Clarenc, Philippe, Jean-François Cote, Alain David, Jacques Friggit, Philippe Gallot, Stéphane Gregoir, Anne Laferrere, Adélia Nobre, Catherine Rougerie, and Nelly Tauzin. 2014. "The Notaires-INSEE Housing Prices Indexes: Version 3 of Hedonic Models." *INSEE Methodes*, [110]

Clark, Gregory. 2002. "Shelter from the Storm: Housing and the Industrial Revolution, 1550–1909." *Journal of Economic History* 62 (02): DOI: 10.1017/s0022050702000578. [203]

Claveau, Jeannine, Jack Lothian, and Sylvie Gauthier. 2009. "The Rent Component of the Consumer Price Index." *Proceedings of the Survey Methods Section*, [96]

Cloyne, James, Clodomiro Ferreira, and Paolo Surico. 2020. "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism." *Review of Economic Studies* 87 (1): 102–29. [5]

Cochrane, John H. 2009. *Asset pricing: Revised edition*. Princeton university press. [35]

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2019. "The Costs of Agglomeration: House and Land Prices in French Cities." *Review of Economic Studies* 86 (4): 1556–89. [9]

Conseil General de l'Environment et du Developpement Durable. 2021. "House Prices in France: Property Price Index, French Real Estate Market Trends in the Long Run." URL: http://www.cgedd.developpement-durable.gouv.fr/house-prices-in-france-property-price-index-french-a1117.html. [110, 114]

CoreLogic. 2018. "CoreLogic Hedonic Home Value Index, March 2018 Results." *Available at www.corelogic.com.au/news*, [87, 89, 92, 93]

Curti, Marco. 1981. "Ergebnisse 3, Die Entwicklung der Wohnungsmieten." In *Reallöhne schweizerischer Industriearbeiter von 1890 bis 1921*. Universität Zürich. Sozialökonomisches Seminar. Abteilung Wirtschaftsgeschichte. [189–191, 194, 197, 198]

Cvijanović, Dragana, and Christophe Spaenjers. 2021. ""We'll always have Paris": Out-of-country Buyers in the Housing Market." *Management Science* 67 (7): 4120–38. [35]

De Stefani, Alessia. 2020. "House Price History, Biased Expectations, and Credit Cycles: The Role of Housing Investors." *Real Estate Economics*, [65, 66]

De Vries, Paul, Jan de Haan, Erna Van der Wal, and Gust Mariën. 2009. "A house price index based on the SPAR method." *Journal of Housing Economics* 18 (3): 214–23. [165]

Demers, Andrew, and Andrea L. Eisfeldt. 2021. "Total Returns to Single-Family Rentals." *Real Estate Economics*, (May): DOI: 10.1111/1540-6229.12353. [8, 9, 17, 20, 209, 210]

Desmet, Klaus, and Vernon Henderson. 2015. "The Geography of Development Within Countries." In *Handbook of Regional and Urban Economics*. Elsevier, 1457–517. DOI: 10.1016/b978-0-444-59531-7.00022-3. [7]

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban Accounting and Welfare." *American Economic Review* 103 (6): 2296–327. DOI: 10.1257/aer.103.6.2296. [9]

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3): 479–524. DOI: 10.1257/aer.20131706. [9]

Diewert, Erwin, and Chihiro Shimizu. 2015. "Residential property price indices for Tokyo." *Macroeconomic Dynamics* 19 (8): 1659–714. [162]

Difusió de Dades. Ajuntament de Barcelona, Departament d'Estadística i. various years. "Alquiler Medio por m2." URL: https://ajuntament.barcelona.cat/estadistica/castella/Estadistiques_per_temes/Habitatge_i_mercat_immobiliari/Mercat_immobiliari/Habitatges_lloguer/evo/tllogem2.htm. [175]

Dröes, M, A Houben, R Van Lamoen, et al. 2017. "De Prijsstijgingen op de Amsterdamse huizenmarkt onder de loep." *Economisch-Statistische Berichten* 102: [166]

Duon, Gaston. 1946. “Documents sur le Problème du Logement à Paris.” *Études Économiques* 1: [110]
Duranton, Gilles. 2007. “Urban Evolutions: The Fast, the Slow, and the Still.” *American Economic Review* 97 (1): 197–221. DOI: 10.1257/aer.97.1.197. [29]
Duranton, Gilles, and Diego Puga. 2000. “Diversity and Specialisation in Cities: Why, Where and When Does it Matter?” *Urban Studies* 37 (3): 533–55. DOI: 10.1080/0042098002104. [9]
Edvinsson, Rodney, Sölvi Blöndal, and Johan Söderberg. 2014. “A Price Index for Residential Property in Stockholm, 1875-2011.” In *Historical Monetary and Financial Statistics for Sweden: House Prices, Stock Returns, National Accounts, and the Riksbank Balance Sheet, 1620-2012.* Ekerlids Förlag. [49, 181, 184]
Eeckhout, Jan. 2004. “Gibrat’s Law for (All) Cities.” *American Economic Review* 94 (5): 1429–51. [29]
Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. “Spatial Sorting.” *Journal of Political Economy* 122 (3): 554–620. DOI: 10.1086/676141. [7]
Eichholtz, Piet. 1997. “A long run house price index: The Herengracht index, 1628–1973.” *Real Estate Economics* 25 (2): 175–92. [165, 166]
Eichholtz, Piet, Matthijs Korevaar, and Thies Lindenthal. 2019. “500 Years of Housing Rents, Quality and Affordability.” *Working Paper*, [166, 167, 202–205]
Eichholtz, Piet, Matthijs Korevaar, Thies Lindenthal, and Ronan Tallec. 2020. “The Total Return and Risk to Residential Real Estate.” *SSRN Electronic Journal*, DOI: 10.2139/ssrn.3549278. [9, 12, 76, 168, 209]
Eitrheim, Øyvind, and Solveig K Erlandsen. 2004. “House Price Indices for Norway 1819–2003.” In. 35. Norges Bank, 349–78. [168, 169]
Eurostat. 2013. *Handbook on Residential Property Prices (RPPIs).* International Monetary Fund. URL: https://www.amazon.com/Handbook-Residential-Property-Prices-RPPIs-ebook/dp/B00F0LQFWQ?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=B00F0LQFWQ. [15, 111, 117, 118, 121, 125–127, 133, 137, 143, 147, 188, 193, 197]
Ewens, Michael, Charles M. Jones, and Matthew Rhodes-Kropf. 2013. “The Price of Diversifiable Risk in Venture Capital and Private Equity.” *Review of Financial Studies* 26 (8): 1854–89. DOI: 10.1093/rfs/hht035. [40]
Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2020. “Heterogeneity and Persistence in Returns to Wealth.” *Econometrica* 88 (1): 115–70. DOI: 10.3982/ecta14835. [6]
Fama, Eugene F, and Kenneth R French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics*, [30]
Favilukis, Jack, Sydney C Ludvigson, and Stijn Van Nieuwerburgh. 2017. “The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium.” *Journal of Political Economy* 125 (1): 140–223. [6]
Firestone, O. 1951. *Residential Real Estate in Canada.* University of Toronto Press. [94, 95]
Fishback, Price, and Trevor Kollmann. 2014. *6. New Multicity Estimates of the Changes in Home Values, 1920–1940.* University of Chicago Press. [207]
Flaus, Lucien. 1946. “La construction d’immeubles d’habitation en France entre 1928 et 1939.” *Journal de la société française de statistique* 87: 67–80. [115]
Forte, Carlo, and Roberto Di Stefano. 1970. *Analisi storica della rendita urbana.* Napoli: Ed. Scientifiche Italiane. [151, 153, 155]
Fourth census of Canada, 1901. 1901. *Fourth census of Canada, 1901.* Canada. Census, and Statistics Office. [94]

Fox, Ryan, and Peter Tulip. 2014. “Is Housing Overvalued?” *Available at SSRN 2498294*, [88, 92]

Francke, Marc. 2010. “Repeat sales index for thin markets.” *Journal of Real Estate Finance and Economics* 41 (1): 24–52. [165, 166]

Francke, Marc, and Matthijs Korevaar. 2019. “When Birth or Death Hits Home: Demograhic Change and the Housing Market.” *Working Paper*, [165, 166]

Friggit, Jacques. 2002. “Long Term House Prices and Residential Property Investment Performance in Paris in the Time of the French Franc, 1840-2001.” URL: http:%20//%20www.%20cgedd.%20developpement-durable.%20gouv.%20fr/%20IMG/doc/%20house-price-france-1840-2001_%20cle5a8666.%20doc. [110, 112]

Friggit, Jacques. 2008. “Comparing Four Secular Home Price Indices.” *Note de Travail, CGEDD*, [111]

Friggit, Jaques. 2013. “Loyers et Revenus Depuis les Années 1970.” *CGEDD*, [113]

Gennaioli, Nicola, and Andrei Shleifer. 2018. *A Crisis of Beliefs*. Princeton University Press. [65]

Giacoletti, Marco. 2021. “Idiosyncratic Risk in Housing Markets.” *Review of Financial Studies* 34 (8): edited by Stijn Van Nieuwerburgh, 3695–741. DOI: 10.1093/rfs/hhab033. [8, 9, 35, 38, 40, 69–71, 75, 78]

Glaeser, Edward L., editor. 2010. *Agglomeration Economics*. University of Chicago Press. [9]

Goetzmann, William N., and Alok Kumar. 2008. “Equity Portfolio Diversification*.” *Review of Finance* 12 (3): 433–63. DOI: 10.1093/rof/rfn005. [72]

Grebler, Leo. 1955. *Experience in Urban Real Estate Investment: An Interim Report Based on New York City Properties*. Columbia University Press. [211]

Greenwald, Daniel L, Matteo Leombroni, Hanno Lustig, and Stijn Van Nieuwerburgh. 2021. “Financial and Total Wealth Inequality with Declining Interest Rates.” Working paper. National Bureau of Economic Research. [6]

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. “Superstar Cities.” *American Economic Journal: Economic Policy* 5 (4): 167–99. DOI: 10.1257/pol.5.4.167. [6–9, 28, 29, 34, 56–59, 66, 72]

Han, Lu, and William C. Strange. 2015. “The Microstructure of Housing Markets.” In *Handbook of Regional and Urban Economics*. Elsevier, 813–86. DOI: 10.1016/b978-0-444-59531-7.00013-2. [40]

Hilber, Christian, and Andreas Mense. 2021. “Why Have House Prices Risen so Much More than Rents in Superstar Cities?” [9, 32, 200, 203]

Hilber, Christian, and Wouter Vermeulen. 2016. “The Impact of Supply Constraints on House Prices in England.” *Economic Journal* 126 (591): 358–405. DOI: 10.1111/ecoj.12213. [8, 9, 34, 39]

Hill, Robert J, Michael Scholz, Chihiro Shimizu, and Miriam Steurer. 2018. “An evaluation of the methods used by European countries to compute their official house price indices.” *Economie et Statistique* 500 (1): 221–38. [161–163]

Hill, Robert J. 2012. “Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy*.” *Journal of Economic Surveys*, DOI: 10.1111/j.1467-6419.2012.00731.x. [15]

Hub, ArcGIS. 2019. “Bundesländer 2019 mit Einwohnerzahl.” *Esri Deutschland Content*, [60]

Hudson, Robert S., and Andros Gregoriou. 2015. “Calculating and Comparing Security Returns is Harder than You Think: A Comparison between Logarithmic and Simple Returns.” *International Review of Financial Analysis* 38 (March): 151–62. DOI: 10.1016/j.irfa.2014.10.008. [44, 45]

Hull, John C. 2019. “Options, Futures and Other Derivatives.” [44]

INE. 2021a. “Indice de precios de consumo (IPC) - Alquiler de Vivienda.” URL: https://www.ine.es/jaxiT3/Tabla.htm?t=22345. [175, 179]

INE. 2021b. “Índice de precios de vivienda.” URL: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736152838&menu=ultiDatos&idp=1254735976607. [174, 178]

INSEE. various years. *Annuaire Statistique de la France*. INSEE. [114]

Institute, Japan Real Estate. 2021. “Fudoken Housing Price Index.” URL: https://www.reinet.or.jp/?page_id=14347. [162, 163]

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): 1225–98. DOI: 10.1093/qje/qjz012. [6, 9, 10, 16, 17, 20–23, 28, 43, 47–51, 58, 61, 63, 79, 105, 108, 123, 132, 140, 141, 150, 151, 153, 155, 158, 159, 163, 170, 183, 185, 191, 192, 195, 196, 198, 206]

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2017. “Macrofinancial History and the New Business Cycle Facts.” *NBER Macroeconomics Annual* 31 (1): 213–63. DOI: 10.1086/690241. [12, 15, 18, 85, 94, 116]

Jordà, Òscar, Moritz Schularick, and Alan M Taylor. 2019. “The Total Risk Premium Puzzle.” Working paper. National Bureau of Economic Research. DOI: 10.3386/w25653. [9]

Kalbro, Thomas, and Hans Mattsson. 1995. *Urban land and property markets in Sweden*. London: Routledge. [183, 186]

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante. 2020. “Non-Durable Consumption and Housing Net Worth in the Great Recession: Evidence from Easily Accessible Data.” *Journal of Public Economics* 189: 104–76. [6]

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante. 2018. “Monetary Policy According to HANK.” *American Economic Review* 108 (3): 697–743. [5]

Keely, Richard, and Ronan C. Lyons. 2020. “Housing Prices, Yields and Credit Conditions in Dublin since 1945.” (August): DOI: 10.1007/s11146-020-09788-z. [9, 201, 203]

Kemeny, Jim. 1995. *From Public Housing to the Social Market: Rental Policy Strategies In Comparative Perspective*. London et al.: Routledge. [28]

Kholodilin, Konstantin. 2020. “Long-Term, Multicountry Perspective on Rental Market Regulations.” *Housing Policy Debate* 30 (6): 994–1015. DOI: 10.1080/10511482.2020.1789889. [27]

Kholodilin, Konstantin, and Sebastian Kohl. 2021. “Social Policy or Crowding-out? Tenant Protection in Comparative Long-run Perspective.” *Housing Studies*, 1–24. DOI: 10.1080/02673037.2021.1900796. [28]

Kholodilin, Konstantin, Sebastian Kohl, Artem Korzhenevych, and Linus Pfeiffer. 2021. “The Hidden Homeownership Welfare State: An International Long-Term Perspective on the Tax Treatment of Homeowners.” *DIW Discussion Papers*, [64]

Klick, Larissa, and Sandra Schaffner. 2020. “Regional Real Estate Price Indices for Germany, 2008 – 2019: RWI–GEO–REDX.” *Jahrbücher für Nationalökonomie und Statistik* 241 (1): 119–29. DOI: 10.1515/jbnst-2020-0001. [79, 82]

Knoll, Katharina. 2017. “Our Home in Days Gone By: Housing Markets in Advances Economies in Historical Perspective.” Doctoral dissertation. Freie Universität Berlin. [112]

Knoll, Katharina, Moritz Schularick, and Thomas Steger. 2017. “No Price Like Home: Global House Prices, 1870-2012.” *American Economic Review* 107 (2): 331–53. DOI: 10.1257/aer.20150501. [6, 9, 26, 47, 49, 90, 118, 151, 202]

Kohl, Sebastian. 2017. *Homeownership, Renting and Society: Historical and Comparative Perspectives*. London: Routledge. [27]

Kohl, Sebastian, and Jardar Sørvoll. 2021. “Varieties of Social Democracy and Cooperativism: Explaining the Divergence between Housing Regimes in Nordic and German-Speaking Countries.” *Social Science History*, [14]

Kotova, Nadia, and Anthony Lee Zhang. 2019. "Search Frictions and Idiosyncratic Price Dispersion in the US Housing Market." *SSRN Electronic Journal*, DOI: 10.2139/ssrn.3386353. [8, 78]

Kuchler, Theresa, and Basit Zafar. 2019. "Personal Experiences and Expectations about Aggregate Outcomes." *Journal of Finance* 74 (5): 2491–542. DOI: 10.1111/jofi.12819. [65]

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins. 2020. "Income and Wealth Inequality in America, 1949-2016." *Journal of Political Economy* 128 (9): 3469–519. [6]

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2020. "The Expected Return on Risky Assets: International Long-run Evidence." *SSRN 3546005*, [19]

Labour Statistics, Bureau of. 2009. "How the CPI Measures Price Change of Owners' Equivalent Rent of Primary Residence (OER) and Rent of Primary Residence (Rent)." *Bureau of Labour Statistics - Factsheets*, [209]

LaPoint, Cameron. 2020. "You Only Lend Twice: Corporate Borrowing and Land Values in Real Estate Cycles." *Available at SSRN*, [161–163]

Leroy-Beaulieu, Paul. 1908. *L'Art de Placer et Gérer sa Fortune*. Ch. Delagrave. [115]

Lustig, Hanno N, and Stijn Van Nieuwerburgh. 2005. "Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective." *Journal of Finance* 60 (3): 1167–219. DOI: 10.1111/j.1540-6261.2005.00759.x. [9]

Marnata, F. 1961. *Les Loyers des Bourgeois de Paris, 1860-1958*. Paris: A. Colin. [112, 114]

Matti, Werner. 1963. "Hamburger Grundeigentumswechsel und Baulandpreise 1903-1937 und 1955-1962." In *Hamburg in Zahlen: Monatsschrift des Statistischen Landesamtes der Freien und Hansestadt Hamburg*. Statistisches Landesamt der Freien und Hansestadt Hamburg. [143, 146]

Merton. 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance*, [35]

Meucci, Attilio. 2010. "Quant Nugget 2: Linear vs. Compounded Returns–Common Pitfalls in Portfolio Management." *GARP risk professional*, 49–51. [44]

Meuriot, Paul. 1913. "Le livre foncier de Paris 1911. Étude démographique et économique." *Journal de la société française de statistique* 54: 364–407. [115]

Mian, Atif, and Amir Sufi. 2011. "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis." *American Economic Review* 101 (5): 2132–56. [5]

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103. DOI: 10.1257/app.5.1.65. [72]

Morrison, P.S. 1978. "Residential Property Conversion: Subdivision, Merger and Quality Change in the Inner City Housing Stock, Metropolitan Toronto, 1958-1973." Doctoral dissertation. University of Toronto. [94, 95]

Mueller, Elisabeth. 2010. "Returns to Private Equity - Idiosyncratic Risk Does Matter!" *Review of Finance* 15 (3): 545–74. DOI: 10.1093/rof/rfq003. [40]

Netherlands, Statistics. 2020. "Prices of Owner-Occupied Dwellings." URL: https://www.cbs.nl/en-gb/series/prices-of-owner-occupied-dwellings. [166]

Neuhaus, Georg. 1915. *Die Stadt Cöln im ersten Jahrhundert unter Preussischer Herrschaft 1815 bis 1915. Zweiter Band. Die Verwaltung der Stadt Cöln seit der Reichsgründung in Einzeldarstellungen.* Edited by Stadt Cöln. Verlag von Paul Neubner. [127]

Neuhaus, Georg. 1916. *Die Stadt Cöln im ersten Jahrhundert unter Preussischer Herrschaft 1815 bis 1915. Erster Band, II. Teil. Die Entwicklung der Stadt Cöln von der Errichtung des deutschen Reiches bis zum Weltkriege*. Edited by Statistisches Amt der Stadt Cöln. Verlag von Paul Neubner. [127]

Neutze, Max. 1972. “The Cost of Housing.” *Economic Record* 48 (3): 357–73. DOI: 10.1111/j.1475-4932.1972.tb00784.x. [90]

Nicholas, Tom, and Anna Scherbina. 2012. “Real Estate Prices During the Roaring Twenties and the Great Depression.” *Real Estate Economics* 41 (2): 278–309. DOI: 10.1111/j.1540-6229.2012.00346.x. [207, 208]

Observatoire des Loyers de l’Agglomeration Parisienne. various years. “Rapport des Loyers.” URL: https://www.observatoire-des-loyers.fr/rapports-sur-les-loyers/. [112, 114]

Oust, Are. 2013. “Rent Indices, Oslo 1970-2008.” *International Journal of Housing Markets and Analysis* 6 (3): 341–55. DOI: 10.1108/IJHMA-07-2012-0032. [168–170]

Paciorek, Andrew. 2013. “Supply Constraints and Housing Market Dynamics.” *Journal of Urban Economics* 77 (September): 11–26. DOI: 10.1016/j.jue.2013.04.001. [76]

Parker, Jonathan A., and Christian Julliard. 2005. “Consumption Risk and the Cross Section of Expected Returns.” *Journal of Political Economy* 113 (1): 185–222. DOI: 10.1086/426042. [36]

Piazzesi, Monika. 2018. “Asset Pricing Program.” *NBER Reporter* 2018 (2): 1. [9]

Piazzesi, Monika, and Martin Schneider. 2016. “Housing and Macroeconomics.” In *Handbook of Macroeconomics*. Elsevier, 1547–640. DOI: 10.1016/bs.hesmac.2016.06.003. [35]

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel. 2020. “Segmented Housing Search.” *American Economic Review* 110 (3): 720–59. [6]

Piazzesi, Monika, Martin Schneider, and Selale Tuzel. 2007. “Housing, Consumption and Asset Pricing.” *Journal of Financial Economics* 83 (3): 531–69. DOI: 10.1016/j.jfineco.2006.01.006. [9]

Pistor, Ernst-August. 1955. *Die Lage auf dem Westberliner Grundstücksmarkt 1955*. Wilhelm Droste & Co. [124]

Pistor, Ernst-August. 1957. *Die Lage auf dem Westberliner Grundstücksmarkt 1957*. Wilhelm Droste & Co. [119]

Pistor, Ernst-August. 1960. *Die Lage auf dem Westberliner Grundstücksmarkt 1959/1960*. Wilhelm Droste & Co. [119, 124]

Pooley, Colin G. 1992. *Housing Strategies in Europe, 1880-1930*. Leicester University Press. [19]

Prauser, Reinhold. 1941. *Die Grundstücksschätzung*. Vol. 4, Carl Heymanns Verlag. [118, 120]

Reba, Meredith, Femke Reitsma, and Karen C. Seto. 2016. “Spatializing 6,000 Years of Global Urbanization from 3700 BC to AD 2000.” *Scientific Data* 3 (1): DOI: 10.1038/sdata.2016.34. [12, 85, 109, 116, 165]

Reich, Emmy. 1912. *Der Wohnungsmarkt in Berlin von 1840-1910*. Edited by Gustav Schmoller and Max Sering. Duncker & Humblot. [121]

Revilla, Carla. 2021. “Long-run Housing Price Developments (with-)in Toronto.” Mathesis. University of Bonn. [94, 95, 97]

Robinson, David T., and Berk A. Sensoy. 2016. “Cyclicality, Performance Measurement, and Cash Flow Liquidity in Private Equity.” *Journal of Financial Economics* 122 (3): 521–43. DOI: 10.1016/j.jfineco.2016.09.008. [40]

Rozsypal, Filip, and Kathrin Schlafmann. 2020. “Overpersistence Bias in Individual Income Expectations and Its Aggregate Implications.” *Working Paper*, [33]

Sagi, Jacob S. 2021. “Asset-Level Risk and Return in Real Estate Investments.” *Review of Financial Studies* 34 (8): edited by Stijn Van Nieuwerburgh, 3647–94. DOI: 10.1093/rfs/hhaa122. [8, 9, 38, 40, 71, 78]

Saint-Genix, De et al. 1895. “De la valeur des terrains et immeubles à Paris, à différentes époques.” *Journal de la société française de statistique* 36: 179–90. [114]

Saiz, Albert. 2010. “The Geographic Determinants of Housing Supply.” *Quarterly Journal of Economics* 125 (3): 1253–96. [9, 34, 76]

Samy, Luke. 2015. “Indices of House Prices and Rent Prices of Residential Property in London, 1895-1939.” Working paper. University of Oxford, Department of Economics. [200–203]

Schwartz, Herman M., and Leonard Seabrooke. 2008. “Varieties of Residential Capitalism in the International Political Economy: Old Welfare States and the New Politics of Housing.” *Comparative European Politics* 6: 237–61. URL: http://dx.doi.org/10.1057/cep.2008.10. [28]

Shiller. 2015. *Irrational Exuberance*. Princeton university press. [6]

Shimizu, Chihiro, and Kiyohiko G Nishimura. 2007. “Pricing structure in Tokyo metropolitan land markets and its structural changes: pre-bubble, bubble, and post-bubble periods.” *Journal of Real Estate Finance and Economics* 35 (4): 475–96. [161]

Shimizu, Chihiro, Kiyohiko G Nishimura, and Tsutomu Watanabe. 2010. “Residential rents and price rigidity: Micro structure and macro consequences.” *Journal of the Japanese and International Economies* 24 (2): 282–99. [163]

Shimizu, Chihiro, Kiyohiko G Nishimura, and Tsutomu Watanabe. 2016. “Housing prices in Tokyo: A comparison of hedonic and repeat sales measures.” In *Index Number Theory and Price Statistics*. De Gruyter Oldenbourg, 792–813. [161, 162]

Simonnet, François, Georges Gallais-Hamonno, and Pedro Arbulu. 1998. “Un siècle de placement immobilier. L’exemple de la fourmi immobilière.” *Journal de la société française de statistique* 139 (2): 95–135. [115]

Soo, Cindy K. 2018. “Quantifying Sentiment with News Media across Local Housing Markets.” *Review of Financial Studies* 31 (10): 3689–719. [66]

Sorensen, Morten, Neng Wang, and Jinqiang Yang. 2014. “Valuing Private Equity.” *Review of Financial Studies* 27 (7): 1977–2021. DOI: 10.1093/rfs/hhu013. [40]

Stapledon, Nigel. 2007. “Long Term Housing Prices in Australia and Some Economic Perspectives.” *PhD diss., University of New South Wales, Sydney*, [51, 86–89, 91–94]

Stapledon, Nigel. 2012. “Trends and Cycles in Sydney and Melbourne House Prices from 1880 to 2011.” *Australian Economic History Review* 52 (3): 293–317. DOI: 10.1111/j.1467-8446.2012.00359.x. [51, 85–87, 89–91]

Statistical Yearbook of the City of Helsinki. various years. *Statistical Yearbook of the City of Helsinki*. City of Helsinki. URL: https://www.hel.fi/static/tieke/digitoidut_asiakirjat/helsingin_kaupungin_tilastolliset_vuosikirjat/1996.html. [107]

Statistics Denmark. 2021. “EJ55: Price Index for Sales of Property (2006=100) by Region, Category of Real Property and Unit.” URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/sales-of-real-property. [103]

Statistics Finland. 2020a. “Prices of dwellings in housing companies.” [107, 108]

Statistics Finland. 2020b. “Real Estate Prices.” [107]

The Finnish Property Market, 2019. 2019. *The Finnish Property Market, 2019*. KTI. [108]

Tokyo Statistical Yearbook. various years. *Tokyo Statistical Yearbook*. City of Tokyo. URL: https://www.toukei.metro.tokyo.lg.jp/tnenkan/tn-eindex.htm. [164]

Tversky, Amos, and Daniel Kahneman. 1983. “Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment.” *Psychological review* 90 (4): 293. [65]

United Nations. 2018. “World Urbanization Prospects: The 2018 Revision, Online Edition.” *Department of Economic and Social Affairs, Population Division*, [21]

Urban Economics, Centre for, and University of British Columbia Real Estate. 2013. “Canadian Cities Housing and Real Estate Data.” URL: http:%20//%20www.%20sauder.%20ubc.%20ca/

%20Faculty/%20Research_%20Centres/%20Centre_%20for_%20Urban_%20Economics_%20and_%20Real_%20Estate/%20Residential_Real_%20Estate_%202. [94]

Van Dijk, Dorinth. 2019. "Local Constant-Quality Housing Market Liquidity Indices." *SSRN Electronic Journal*, DOI: 10.2139/ssrn.3388211. [41]

Verwey, G. 1943. "Prijsverloop van 426 huizen te Amsterdam, welke voor 1880 reeds van eigenaar wisselden." *Economisch Historisch Jaarboek* 22: 251–303. [165, 166]

White, Eugene N, Kenneth Snowden, and Price V Fishback. 2014. *Housing and Mortgage Markets in Historical Perspective*. The University of Chicago Press. [19]

Wischermann, Clemens. 1983. *Wohnen in Hamburg vor dem Ersten Weltkrieg*. Edited by Hans J. Teuteberg and Peter Borscheid. F. Coppenrath Verlag Münster. [146, 148, 149]

Chapter 2

The Fall in the Risk-Free Rate and Rising House Price Dispersion*

Joint with Francisco Amaral, Sebastian Kohl, and Moritz Schularick

2.1 Introduction

The contemporary housing affordability crisis in large urban centers has exposed growing regional inequality in housing prices. Prices in cities like New York, London, or Paris have reached new record heights, while housing price growth in other parts of the respective countries lags behind. Rising housing prices in the most productive cities are a barrier to migration, forcing productive workers to stay in stagnating regions. The resulting inefficient allocation of labor slows down economic growth considerably (Herkenhoff, Ohanian, and Prescott, 2018; Hsieh and Moretti, 2019). The widening gap in housing prices also entails considerable distributional consequences and contributes to political polarization. Areas excluded from rising housing prices are far more likely to vote for populist causes or parties, such as in the case of Brexit (Adler and Ansell, 2019; Ansell, 2019). While the COVID-19 pandemic has temporarily affected the spatial distribution of housing prices, all scenarios predict an at least partial return to the pre-pandemic trends as working-from-home gradually subsides (Gupta, Mittal, Peeters, and Van Nieuwerburgh, 2021).

To understand why housing prices have risen noticeably more in some cities than in others, it is helpful to start with the formation of housing prices. The price of a housing unit is equivalent to its discounted expected future rental cash flow (Poterba, 1984). As such, rents define the fundamental value of housing prices. If

* This work is part of a larger project supported by the European Research Council Grant (ERC-2017-COG 772332). We thank Gabriel Ahlfeldt, Benjamin Born, Arpit Gupta, Dmitry Kuvshinov, Andrea Modena, Marco Pagano, and the participants of the "The Socioeconomics of Housing and Finance" workshop in Berlin for helpful feedback. The project also received support from the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy – EXC 2126/1 – 390838866.

housing prices rise more in New York than in the rest of the US, the same spatial pattern in rents would be expected. Following this straightforward argument, existing explanations for increasing regional price dispersion focused on local housing demand and supply fundamentals. New data, however, indicate that rent dispersion has increased considerably less than price dispersion over the last decades. This calls into question the extent to which shocks to fundamentals can account for the increase in price dispersion.

In this paper, we present an alternative mechanism for the increase in housing price dispersion that does not depend on increasing rent dispersion but lies at the heart of housing price formation. Assuming initial regional differences in rent–price ratios, we demonstrate that a nationwide fall in discount rates induced by the fall in the risk-free rate increases regional gaps in housing prices. While initial regional differences in rent–price ratios are evident in the data, we argue that they are rooted in differences in local housing risk-premia. New long-run data elicits empirical support for our mechanism.

In the first part of the paper, we argue that previous explanations for rising housing price dispersion based on fundamentals (e.g., Gyourko, Mayer, and Sinai (2013); Nieuwerburgh and Weill (2010)) cannot generate the full increase in price dispersion we observe in the data without invariably overshooting rent dispersion. New research using high-quality granular housing data rejected attempts to explain the absence of rising rent dispersion in the data with measurement error due to market segmentation between owner-occupied and rental housing (Begley, Loewenstein, and Willen, 2021; Demers and Eisfeldt, 2021). We additionally use a new long-run data set of housing prices and rents for 27 large ("superstar") cities in 15 developed countries constructed by Amaral, Dohmen, Kohl, and Schularick (2021), demonstrating that housing price dispersion has indeed grown considerably more than rent dispersion. By analyzing price–rent ratios, we can directly observe the share of price increases that is not produced by rent increases. The data reveal that price–rent ratios in the largest cities grew by almost twice the national average. This phenomenon began already in the 1980s and slowly accumulated over time, coinciding with the secular fall in the risk-free rate. Explanations based on transitory effects, such as Hilber and Mense (2021), can, thus, also not account for the bulk of excess housing price dispersion.

The second part of the paper introduces a new mechanism that produces growing housing price dispersion without increasing rent dispersion. Building on a spatial version of the Gordon growth model (Gordon, 1962), we show that falling discount rates disproportionately affect the valuation of housing in cities, in which rent–price ratios are initially lower. All else equal, a fall in discount rates leads to a linear fall in rent–price ratios but a non-linear increase in price–rent ratios. This is because the price–rent ratios are the inverse function of rent–price ratios. As the linear fall in rent–price ratios in the superstar cities began at a lower level, the same fall in discount rates will lead to larger increases in price–rent ratios.

The initial difference in rent–price ratios is apparent in the data. In our model, this difference can be generated by either a difference in local rent growth expectations or by differences in local housing risk-premia.[1] According to the data, realized rent growth differed only slightly between the superstar cities and the national average. We, thus, argue that a difference in risk-premia is a more plausible explanation. This is supported by the observation that total housing returns in superstar cities have been considerably below the national average over the long-run and more recent decades (Amaral et al., 2021). As the paper also presents evidence that housing risk is lower in these cities, a lower risk-premium is the evident explanation for this observation.

This new mechanism also finds support in the data. Regional dispersion in price–rent ratios began to increase in the 1980s, coinciding with the fall of ex-ante real risk-free rates (Negro, Giannone, Giannoni, and Tambalotti, 2019; Rachel and Summers, 2019). According to Kuvshinov and Zimmermann (2020), this fall in risk-free rates also led to an overall fall in discount rates. Since the 1980s, rent–price ratios have fallen steadily (abstracting from cyclical variation), which has been evident for superstar cities and has been mirrored by national data starting on a higher level. This observation perfectly fits the model prediction.

Finally, we calibrate our model to the data in 1985 to demonstrate that it can generate the same increase in levels as well as dispersion of price–rent ratios observed in 2018. To do so, we only need to assume a fall of 1.3 percentage points in the discount rates between 1985 and 2018, which is moderate compared to existing estimates in the literature. We can thus link both the overall increase in national housing prices and the rising regional dispersion to the secular fall in real risk-free rates. While several papers attempted to explain the increase in national housing prices with declining interest rates (e.g., Garriga, Manuelli, and Peralta-Alva (2019); Himmelberg, Mayer, and Sinai (2005); Miles and Monro (2019)), to the best of our knowledge, this is the first study to establish a connection between growing regional housing price dispersion and the fall in risk-free rates.

Our mechanism has important implications for the literature regarding secular stagnation (Rachel and Summers, 2019). While most of this work focuses on the causes of the decline in real risk-free rates, our findings suggest that the implied fall in discount rates has considerable real effects. The closest work is Kroen, Liu, Mian, and Sufi (2021), who demonstrate that falling interest rates contribute to the rise of superstar firms in the stock market, especially when interest rates are already low.

The remainder of this paper is organized as follows. The next section examines existing explanations for the increase in housing price dispersion and new evidence

1. Note that this holds under more general conditions. Using a simple "discount rate – cash flow" decomposition (Campbell and Shiller, 1988), it can be shown that a difference in rent–price ratios is driven by a difference in local rent growth expectations or by differences in local housing discount rates.

suggesting that these explanations are insufficient. Section 2.3 presents new empirical evidence that housing price dispersion has notably increased more than rent dispersion since the 1980s. The subsequent section presents the new mechanism and confirms that it matches the empirical evidence and can generate the excess price dispersion observed in the data. The final section concludes.

2.2 Current explanations for increasing price dispersion

Different mechanisms have been proposed to explain the increasing regional dispersion in housing prices that focus on the role of local housing market fundamentals. The different mechanisms are typically embedded in spatial housing models at the center of which there is a variant of the present value equation for housing:

$$P_t^i = \sum_{j=1}^{\infty} \mathbb{E}\left(Rent_{t+j}^i * \left(\frac{1}{1+r_t} \right)^j \right), \tag{2.1}$$

where P_t^i is the housing price in city i at time t, $Rent_t^i$ is the rent payment net of costs, and r_t is the discount rate at time t. Such models typically assume a constant discount rate across regions based on a representative marginal investor. The equation directly links current local housing prices and current and future local rents. Changes in local fundamentals, such as wages, affect local demand for housing services and, thus, rents. These, in turn, affect housing prices. In the next section, we discuss two mechanisms that have received considerable attention in the context of spatial housing models.

2.2.1 Two spatial housing models

Nieuwerburgh and Weill (2010) construct a spatial, dynamic equilibrium model in the tradition of Rosen (1979) and Roback (1982), reproducing the entire distribution of metropolitan areas in the US. These metropolitan areas are hit by idiosyncratic and persistent productivity shocks. Households with heterogeneous abilities move freely across metropolitan areas in reaction to these shocks. Housing supply is limited by supply regulations, meaning that rents will adjust to compensate for resulting regional wage differences. This, in turn, affects housing prices. The authors calibrate productivity shocks to match the increase in the observed regional wage dispersion between metropolitan areas from 1975 to 2007. They demonstrate that the model matches the increase in housing price dispersion observed in the data. Notably, the model produces an increase in rent dispersion three times larger than observed empirically. To explain this divergence between model and data, the authors argue that owner- and renter-occupied markets are segmented, resulting in the measured rent dispersion not mirroring the implied imputed rent dispersion in the owner-occupied housing market.

Gyourko, Mayer, and Sinai (2013) developed a two-location model to show that increasing national demand generated by population growth potentially affects regions differently, depending on local housing supply elasticities. Under the assumption that enough people prefer to live in a supply-constrained city, the model predicts that in response to increasing national demand, a supply-constrained city will experience a stronger rental increase than an unconstrained city. This increase in rent increases housing prices following the present-value equation. The authors call the cities that display a combination of low supply elasticities and strong housing price growth superstar cities. To test the model's predictions, the authors use census data on 279 US metropolitan statistical areas (MSAs) from 1950 to 2000, finding that the superstar cities experienced significantly stronger housing price growth in reaction to a growing number of high-income households nationally. However, the authors do not test the model predictions for rents.[2] In Appendix 2.A, we use the same data to demonstrate that housing prices increased considerably more than rents in superstar cities, which cannot be accounted for by the model in the aforementioned paper.

One concern regarding these contradictory results is that housing prices and rents are measured using data from different market segments. This segmentation can lead to selection bias when comparing rental values, which are typically taken from lower quality segments of the housing market, with housing prices primarily from higher-quality segments. Glaeser and Gyourko (2007) argue that this limits the use of no-arbitrage conditions, such that owner-occupied and rental housing markets can develop differently over time. This could explain why rent dispersion increased substantially less than price dispersion.

2.2.2 New evidence on price-rent ratios

Recently, several researchers have made use of high-quality granular data, demonstrating that these selection problems do not resolve the puzzle of large movements in prices relative to rents. Begley, Loewenstein, and Willen (2021) use micro-data from Corelogic on prices and rents for the same property to estimate price–rent ratios, thereby avoiding this type of selection bias. They show that the price variation in owner- and renter-occupied housing markets are almost perfectly correlated. Between 2000 and 2018, the price–rent ratio accounts for most of the variation in housing prices of owner-occupied dwellings. If anything, they observe that renter-occupied prices have risen more than owner-occupied prices. This could even magnify the differences between price and rent dispersion observed in the data.

Other research observed a considerable downward trend in rent–price ratios, even when correcting for selection problems. Demers and Eisfeldt (2021) use micro-

2. Using their data, in Appendix 2.A, we demonstrate that the authors' results do hold for rents, but the effect on rents is less than half the size of the effect on housing prices.

data from the American Housing Survey to build rent–price ratios for 15 different US cities from 1985 to 2020, using hedonic models and non-parametric methods to adjust for rental homes being more prevalent in lower price tiers. The resulting rent–price ratios indicate a downward trend, particularly in the superstar cities.[3]

In summary, the evidence suggests that housing price growth has outpaced rent growth, particularly in superstar cities. Hilber and Mense (2021) present a model to account for this observation. In the model, an increase in housing prices over rents in superstar cities can be generated by a combination of lower short- and long-run supply elasticities and serially correlated housing demand shocks. After a positive housing demand shock, rents and housing prices rise. The adjustment of the housing supply depends on the short-run supply elasticity. As demand shocks are serially correlated, agents expect another positive demand shock. If the long-run housing supply is sufficiently inelastic, agents expect rents to grow more in the future. This increases housing prices relatively more than rents in the supply-constrained city. The predictions of this model are tested using regional data on housing prices and rents for the UK from 1997 to 2018. The authors demonstrate that more supply-constrained regions, like London, have experienced a greater increase in price–rent ratios in response to labor demand shocks. While the model can explain the cyclical behavior of price–rent ratios, whether this mechanism works in the medium and long run is unclear. In fact, according to the model and the data, housing demand shocks mean-revert over a sufficiently long period (Piazzesi and Schneider, 2016). This implies that the increase in the divergence of price–rent ratios generated by the model is transitory.

All in all, previous literature focused on models in which regional housing price dispersion is generated by increasing rent dispersion. However, this is difficult to reconcile with existing data, which indicates that the increase in rent dispersion has been considerably smaller than the increase in housing price dispersion.

2.3 Long-run empirical evidence

In this section we use the international long-run data from Amaral et al. (2021) to show that: (i) dispersion in housing prices between superstar cities and national averages grew substantially more than dispersion in rents over the last 70 years, (ii) dispersion in prices began increasing in the 1980s - much earlier than the dispersion in rents, and (iii) superstar cities' price–rent ratios also began increasing relative to national averages in the 1980s.

The data cover 27 international superstar cities in 15 OECD countries and include housing price series, rent series, and rent–price ratio estimates from 1870 to

3. Table A.5 in the online Appendix of Demers and Eisfeldt (2021) shows that net rental yields have fallen, particularly for the highest price quintile.

2018. Superstar cities are defined as the largest cities within each country in terms of 1900 population statistics, including cities like London, New York, Paris and Tokyo. The city-level data are merged with national housing data from Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019) that is extended to 2018. For details, please refer to the aforementioned paper.

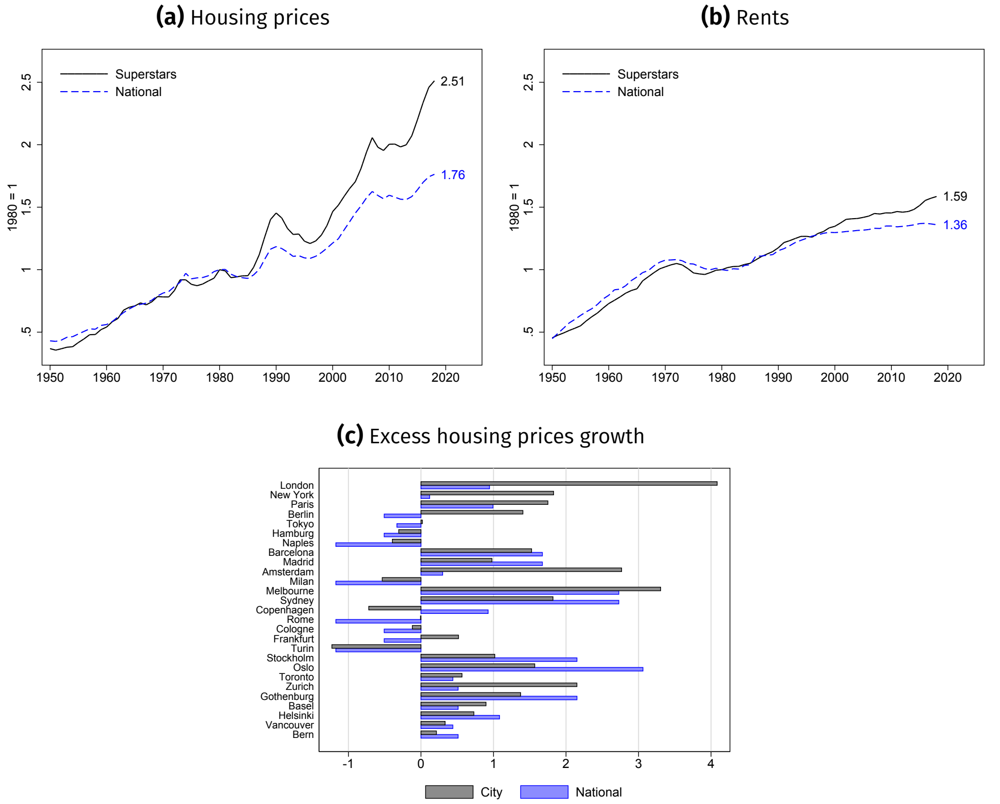

Figure 2.1. Average housing price and rent indices for 27 superstar cities and the national level

Notes: Indices of equally-weighted average real housing price increases (Panel (a)) and real rent increases (Panel (b)) of 27 national superstar cities in black are from Amaral et al. (2021) and average national real housing price and rent increases from Jordà et al. (2019) weighted by the number of sample cities in the respective country in blue. 1980=1. Panel (c) presents the mean difference between superstar cities and the national average in log housing price growth and log rent growth for the period 1980-2018.

Figure 2.1 panel (a) plots the average real housing price increases for the 27 superstar cities and the average real national housing price increases.[4] The national series are weighted by the number of sample cities in the respective country. The right–hand side of Figure 2.1 presents the same indices for rents.

4. We first compute the average increase for every period and then calculate the increases over time. This is equivalent to a portfolio from the 27 superstar cities that is rebalanced every period such that cities that grew considerably in the beginning of the sample period are not subsequently assigned a larger weight.

The data confirm two facts that were uncovered in earlier work. First, housing prices have grown much more than rents in both superstar cities and at national levels, and second, housing price dispersion between superstar cities and national averages increased considerably more than rent dispersion. The data also show that the price divergence began in the 1980s, while the rent divergence only started more recently, around the 2000s. These observations call into question the sole importance of fundamentals for the increase in housing prices over the long-run and for the increase in regional price dispersion.

Panel (c) presents mean excess log housing price growth over log rent growth between 1980 and 2018 at the city and national levels, demonstrating that housing prices have grown more than rents in most cities and countries.[5] In contrast, excess housing price growth has been higher at the city-level than nationally for most cities in the data set. This is particularly true for the largest cities, like London or New York. Exceptions are mainly smaller cities within each country, such as Turin or Bern, or cities in countries that experiences a severe real estate crisis between 1980 and 2018, like the Scandinavian countries in the 1990s and Spain after the financial crisis.

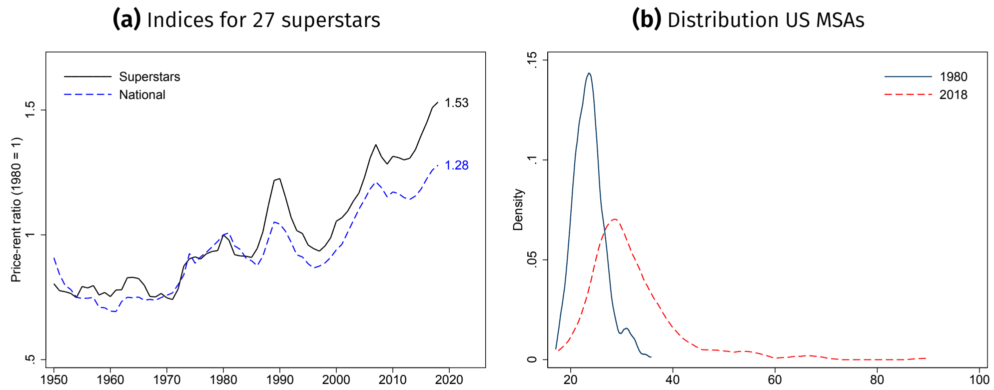

Figure 2.2. Price-rent ratios

Notes: Panel (a): Index of equally-weighted average increases of price-rent ratios of 27 national superstar cities from Amaral et al. (2021) and average national increases of price-rent ratios from Jordà et al. (2019) weighted by the number of sample cities in the respective country. 1980=1. Panel (b): Kernel density of price-rent ratios of 316 US MSAs in 1980 and 2018 calculated from net rental yields from Amaral et al. (2021).

The Figure 2.2 panel (a) plots the average increases in price-rent ratios for the superstars and on the national level to show the proportion of the housing price dispersion that cannot be accounted for by rent dispersion. Changes in price–rent ratios indicate how much housing prices changed after accounting for changes in

5. The exceptions are Germany, which featured a peak in housing prices in 1980; Italy, where housing prices have not yet recovered from the financial crisis in 1980; Japan, which featured a severe real estate crisis in the 1990s; and Denmark.

rents. From previous observations, price–rent ratios are expected to have increased considerably since the 1980s. More importantly, the data show that price–rent ratios have increased considerably more in superstar cities than the national average. While the gap in price–rent ratios varies over the cycle, a phenomenon that could be explained by the mechanism proposed in Hilber and Mense (2021), it shows a strong persistence over the last decades and seems to be increasing over time.

To show that the increase in dispersion in price-rent ratios does not solely apply to the superstar cities but is a feature of the entire city size distribution we use the U.S. MSA–level data from Gyourko, Mayer, and Sinai (2013) that was extended to 2018 in Amaral et al. (2021). Figure 2.2 panel (b) plots the distribution of MSA-level price–rent ratios in 1980 and 2018, demonstrating a considerable increase in the dispersion of price–rent ratios. This phenomenon is particularly strong for the distribution's right tail, where the superstar cities are located. As expected, mean price–rent ratios have also increased over time. Still, the coefficient of variation (CV) increased from 23.7 to 32.1. This increase in the level and dispersion of price–rent ratios since the 1980s is what we will investigate in the rest of the paper.

2.4 New mechanism

This section constructs a parsimonious, spatial Gordon growth model of the housing market. Then we present a new mechanism for the increase in housing price dispersion that does not follow from increasing rent dispersion but results mechanically from differences in rent–price ratios between cities.

To construct the model, we begin with present value equation (2.1), the only difference being that we allow for differences in discount rates between cities:

$$P_t^i = \sum_{j=1}^{\infty} \mathbb{E}\left(Rent_{t+j}^i * \left(\frac{1}{1 + r_t^i} \right)^j \right). \tag{2.2}$$

From a theoretical perspective, a combination of local market segmentation and incomplete markets imply that discount rates do not need to equalize between cities.[6] Piazzesi, Schneider, and Stroebel (2020) show that housing markets are locally segmented, using data on online searches to document large differences in housing search behavior across different municipalities in California.[7] Housing markets are also incomplete because housing assets are indivisible, and homeowners are typically non-diversified. The lack implies limitations to arbitrage precluding discount rates from equalizing (Piazzesi and Schneider, 2016). Empirically, Amaral et al.

6. Sagi (2021) builds a housing search model, showing that heterogeneity in discount rates is an essential condition to explain the dynamics in real estate prices.

7. They also demonstrate that differences in housing search between different quality segments within municipalities are less pronounced.

(2021) show that total housing returns have been persistently lower in superstar cities than in the rest of the country. Differences in housing returns are argued to be due to differences in housing risk, as housing prices co-vary less with income in larger MSAs and idiosyncratic housing price risk is lower. This evidence strongly indicates that discount rates differ between cities.

In the following, we assume that discount rates are composed of a risk-free component, that is equal for the entire country and a risk-premium that can differ by the city invested in; $r_t^i = \text{risk-free}_t + \text{risk-premium}_t^i$.

To simplify the following discussion, we make two additional assumptions: First, we assume that rents in city i are expected at time t to grow at a constant rate g_t^i. Second, we assume that $r_t^i > g_t^i$, such that housing prices are finite. This allows us to rewrite equation (2.2) as the Gordon (1962) growth valuation formula:

$$P_t^i = \sum_{j=1}^{\infty} \left(Rent_t^i * \left(\frac{1+g_t^i}{1+r_t^i} \right)^j \right) \qquad \Longleftrightarrow \qquad P_t^i = Rent_t^i * \frac{1+g_t^i}{r_t^i - g_t^i}. \tag{2.3}$$

Following this equation, the rent-price ratio is equal to:

$$\text{Rent-price ratio}_t^i = \frac{Rent_t^i}{P_t^i} = \frac{r_t^i - g_t^i}{1+g_t^i}. \tag{2.4}$$

To analyze housing price dispersion, we next consider a setting with two cities: reservation city A and superstar city B. The reservation city can be understood as the average of all other locations within a country except the superstar city. To compare both cities, we make three additional assumptions. First, as argued in the urban economics literature (Gyourko, Mayer, and Sinai, 2013; Hilber and Mense, 2021) we assume expected rent growth in the superstar city is higher than or equal to the reservation city; $g_t^B > g_t^A \;\; \forall t$. Second, as argued above, following Amaral et al. (2021), we assume that risk-premia are lower or equal for housing investments in the superstar compared to the reservation city, such that $r_t^B \leq r_t^A \;\; \forall t$. Third, we assume that at least one of the two previous inequalities is strict, such that rent–price ratios are lower in the superstar city and:

$$r_t^A - g_t^A > r_t^B - g_t^B > 0. \tag{2.5}$$

From equation (2.3) we can write the log price difference between cities B and A as:

$$\begin{aligned} &log(P_t^B) - log(P_t^A) = \\ &\qquad log(Rent_t^B) + log\left(\frac{1+g_t^B}{r_t^B - g_t^B}\right) - log(Rent_t^A) - log\left(\frac{1+g_t^A}{r_t^A - g_t^A}\right). \end{aligned} \tag{2.6}$$

2.4.1 A fall in the risk-free rate

In this subsection, we derive the predictions of our model after a fall in the risk-free rate. We assume that the risk-free rate decreases by Δ, such that:

$$\begin{aligned} log(P_t^B) - log(P_t^A) = \\ log(Rent_t^B) + log\left(\frac{1+g_t^B}{r_t^B-\Delta-g_t^B}\right) - log(Rent_t^A) - log\left(\frac{1+g_t^A}{r_t^A-\Delta-g_t^A}\right). \end{aligned} \quad (2.7)$$

If we differentiate with respect to Δ and under the assumptions made above, we get:

$$\frac{\partial\left(log(P_t^B) - log(P_t^A)\right)}{\partial\Delta} = \frac{1}{r_t^B-\Delta-g_t^B} - \frac{1}{r_t^A-\Delta-g_t^A} \quad > \quad 0.$$

This demonstrates that a uniform fall in discount rates across both cities, generated by a fall in the risk-free rate, increases housing price dispersion if rent–price ratios initially differ.

The intuition for this observation is presented in Figure 2.3. Panel (a) plots the rent–price ratio in the model as a function of $r-g$ for a varying r, wherein the rent–price ratio changes linearly in r. Following equation (2.5), we assume that $r-g$ is lower in the superstar city at time t, resulting in a lower rent–price ratio. Next, we assume that between t and $t+1$ r falls in both cities by one percentage point. This leads to an approximately equal fall in the rent–price ratio in the superstar (B) and in the reservation city (A).

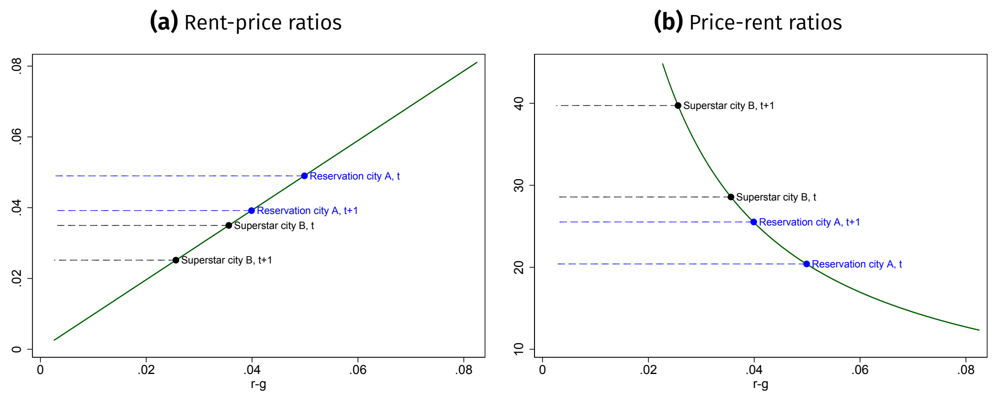

Figure 2.3. A fall in discount rates in the model

Notes: Panel (a) plots the rent–price ratio in our model as a function of $r-g$. To calculate the points, we assumed that $g = 0.0175$. Panel (b) shows the corresponding price–rent ratio.

Figure 2.3 panel (b) plots the corresponding price–rent ratio. As the price–rent ratio is the inverse function of the rent–price ratio, when *r* changes, the price–rent ratio changes in a non-linear fashion. Since the initial price–rent ratio is higher in the superstar city, an equally large fall in *r* leads to a larger increase in the price–rent ratio in the superstar than in the reservation city. Subsequently, the price dispersion between the superstar city and the reservation city increases when *r* falls, even when rents are constant in both cities.

2.4.2 Rent–price ratios in the data

The previous section we determined that price dispersion increases in response to a fall in the risk-free rate if rent–price ratios initially differ. Our model also predicts a parallel fall in rent–price ratios across cities due to a fall in the risk-free rate.

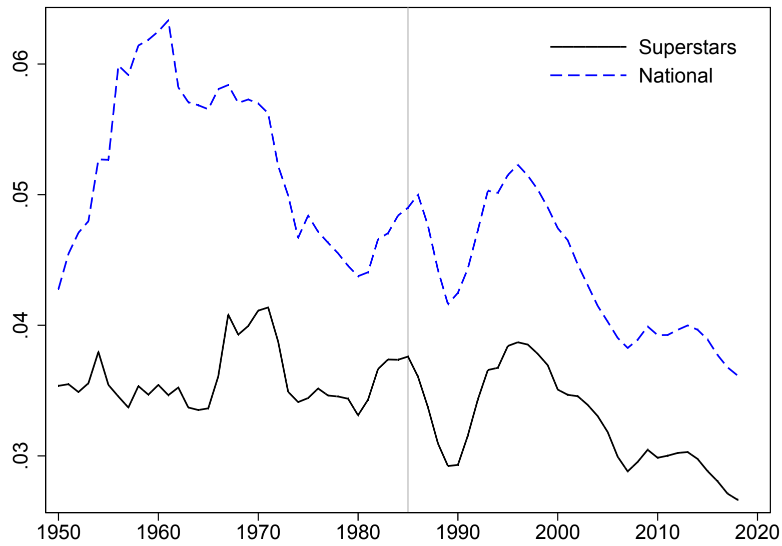

Figure 2.4. Rent–price ratios in the data

Notes: The solid black line is the non-weighted average rent–price ratio of 27 national superstar cities from Amaral et al. (2021). The dashed blue line is the average of the national rent–price ratio from Jordà et al. (2019) weighted by the number of sample cities in the respective country.

Figure 2.4 plots the average rent-price ratios in the 27 superstar cities in Amaral et al. (2021) and in the national rent-price ratios Two observations are important. First, rent–price ratios have been lower in the superstar cities over the entire period since 1950. This evidence validates the assumption regarding the initial differences in rent–price ratios.

Second, the rent–price ratios in the superstar cities and at the national level have moved in parallel trajectories since 1985 (abstracting from the cyclical variation), suggesting a common downward trend. Rent-price ratios fell by around 1.2 percentage points from 1985 to 2018 in the superstar cities and at the national level. The equally large fall in rent–price ratios in the superstar cities and at the national level is equivalent to the parallel fall in rent–price ratios predicted by the model. Note that alternative mechanisms that attempt to explain the increase in price dispersion based on factors that solely affect the superstar cities, would predict a divergence in rent–price ratios between the superstar cities and the rest.

2.4.3 Calibration

To simulate the increase in price dispersion in response to a fall in r in our model, we calibrate the model to the following values. We set the expected rent growth in the superstar and the reservation city equal to 1.75 %, $g_t^B = g_t^A = 0.0175 \ \ \forall t$, which is close to long-run rent growth rates observed in the data.[8] Next, we assume that the discount rate in the superstar city is 1 percentage point lower than in the reservation city; $r_t^B = r_t^A - 0.01 \ \ \forall t$. This is equivalent to the difference in total housing returns of around 1 percentage point found in Amaral et al. (2021). For simplification we assume that rents in the superstar city and in the reservation city are equal to one, $Rent_t^B = Rent_t^A = 1 \ \ \forall t$, such that the resulting housing prices are equal to the price–rent ratios.

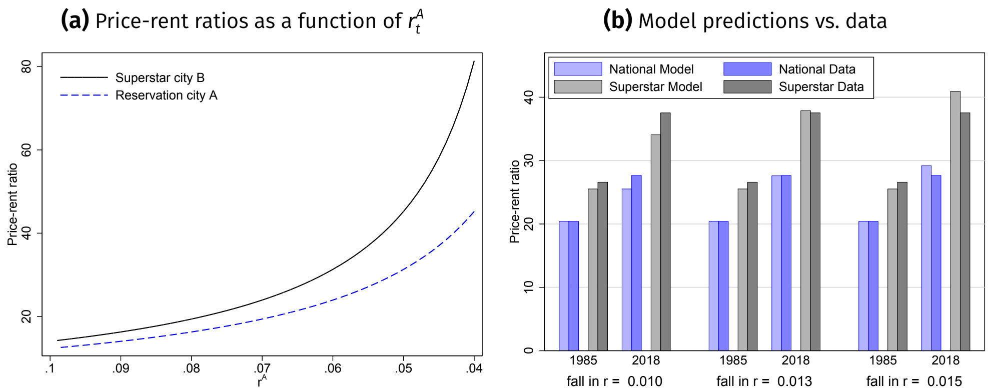

Figure 2.5. Simulated price-rent ratios in response to a fall in r

Notes: Panel (a) shows price–rent ratios for the superstar city and the reservation city in the model relative to the discount rate in the reservation city. Panel (b) compares the model to the data for different assumed values of the fall in the discount rate r. For both exercises, we assume that $g = 0.0175$ and $r^B = r^A - 0.01$.

Figure 2.5 panel (a) plots the resulting price–rent ratios in the superstar and reservation cities as a function of r_t^A, demonstrating that the dispersion in price–rent ratios increases when discount rates fall. Although the initial difference between the cities is small for high discount rates, the difference becomes substantial when discount rates are low.

8. Between 1950 and 2018, rents grew on average by 1.86 % in the 27 superstar cities and by 1.65 % at the national level.

The next step is to assess whether our model matches the increasing level and dispersion in price–rent ratios in the data. To do so requires estimates for the discount rates in 1985 and 2018. Unfortunately, to the best of our knowledge, there are no reliable estimates of housing discount rates to calibrate our model. Therefore, we take a reverse approach and back out the discount rate for the national average in 1985 from the price–rent ratio we observe in the data using our model. The resulting discount rate is $r^A_{1985} = 0.067$.[9]

Next, we compare our model to the data assuming different scenarios for the fall in discount rates between 1985 and 2018. The estimated decline in the real risk-free rate ranges from 2.5 to 5 percentage points depending on the estimation method (Negro et al., 2019; Rachel and Summers, 2019). At the same time, there is considerable evidence that risk-premia have risen during this period (Caballero, Farhi, and Gourinchas, 2017; Kuvshinov and Zimmermann, 2020). This partially offsets the effect of the fall in the risk-free rate on discount rates for risky assets.

Figure 2.5 panel (b) compares the price–rent ratios predicted by our model to the actual price–rent ratios in the data for the years 1985 and 2018. We represent three scenarios for the fall in discount rates. On the left, discount rates fell by 1 p.p., in the middle by 1.3 p.p. and on the right by 1.5 p.p. Overall, the model slightly overshoots the price–rent ratio in the superstar cities in 1985.[10] This indicates that the difference in risk-premia between the superstars and the national average was either smaller than 1 percentage point or the rent-growth expectations have been slightly higher in the superstar cities.

In the scenario where discount rates fall by 1 percentage point, our model cannot fully account for the rise in levels and dispersion of the price–rent ratio. It does, however, generate a substantial portion of the increase in levels and dispersion we observe in the data. Assuming a fall in r of 1.5 percentage points instead, our model does overshoot both the level and the dispersion in housing prices we observe in the data. To perfectly match the increase in levels and dispersion in the data, requires a fall in discount rates of around 1.3 percentage points.

Our model also matches the increase in levels and dispersion of price–rent ratios under different scenarios for g Most notably, if we assume that expected rent growth was 1 p.p. higher in the superstar city, our model produces very similar results without assuming any difference in discount rates between superstar cities and the national average. Given the small difference in observed rent growth and the stable return difference between superstar cities and the national average, we assert that a constant difference in discount rates is more realistic than a constant differ-

9. Estimates of ex-ante real risk-free rates range from 2.5 % to 5 % for 1985 (Negro et al., 2019; Rachel and Summers, 2019) and estimates for the risk-premium on the equity market portfolio range between 2.5 % and 4.3 % for the period between 1950 and 2000 in the US (Fama and French, 2002). This implies a discount rate on risky assets between 5 % and 9 %.

10. Note that the model exactly matches the national price–rent ratio in 1985 by construction.

ence in expected rent growth. There might be differences in both variables, and the specific combination of r and g might even be investor-specific. A large-scale simulation of many different combinations of different model variables (Appendix 2.B) demonstrates that falling discount rates robustly lead to increasing housing price dispersion for most value combinations for r and g. We also use the US MSA-level data introduced in section 2.3 to compare the full distribution of price–rent ratios in 2018 with our model prediction. Although our model does not perfectly fit the individual MSA-level data, it matches the overall distribution (Appendix 2.C).

2.5 Conclusion

In this paper, we build a spatial Gordon growth model, that explains the increase in housing price dispersion that cannot be explained by increasing rent dispersion, where existing models fall short. For reasonable values regarding the fall in discount rates, our model produces the same rise in levels and dispersion of price–rent ratios observed in the data without assuming any changes in fundamentals or expectations.

The key novelty of the model is that a uniform fall in interest rates can have heterogeneous spatial effects. This implies that arguments against the large influence of interest rates on the current surge in housing prices are probably misleading. Given the central role of housing in urban economics, future research is expected to adopt this mechanism into more complex spatial models.

Appendix 2.A Superstar cities revisited

2.A.1 Rent growth

Gyourko, Mayer, and Sinai (2013) derive a set of propositions, that directly imply that superstar cities should have experienced stronger rent growth than the rest of the country. Proposition 1 states that superstar cities have higher rental values than the rest of the country. Proposition 3 states that an increase in aggregate income leads to stronger rental increases in the superstar cities than in the rest.[11] These two propositions are tested in Tables 2 and 3 of the paper, using log house value as the dependent variable. Here, we replicate the analysis focusing on the effects on house value growth and rent growth. Table 2.A.1 presents our regression output. There are two primary results. First, the coefficients for rent values are significant and positive, just as the coefficients for house values. Second, the coefficients for rent values are slightly less than half those of house values. This indicates that the effects on rents are much smaller than on prices, which raises the question of whether we can fully explain the strong divergence in prices with the divergence in rents.

Table 2.A.1. Replicating Panel A from Tables 2 and 3 in Gyourko, Mayer, and Sinai (2013)

	log house value	log rent value	log house value	log rent value
Superstar	0.605	0.291		
	(0.0729)	(0.0377)		
Superstar x Rich			0.394	0.172
			(0.0356)	(0.0193)
N	1116	1116	1116	1116
adj. R^2	0.414	0.308	0.856	0.861

Notes: This table replicates Panel A from Tables 2 and 3 in Gyourko, Mayer, and Sinai (2013). In addition to the regression on log house value, we perform the same regression on rent log value. Columns 1 and 2 present the results of a regression of the left hand-side variable on a superstar dummy and year fixed effects. Columns 3 and 4 present the OLS coefficients of a regression on an interaction effect of a superstar dummy and the log number of rich families in the U.S. and time and superstar fixed effects. Standard errors are in parentheses and are clustered at the MSA-level.

2.A.2 Price-rent ratios

In this subsection, we present evidence that the divergence in price-rent ratios between superstar cities and the rest has strongly increased since the 1980s, extending the data set presented in Gyourko, Mayer, and Sinai (2013) to 2010 and 2018. We then use the definition of superstar cities to categorize the cities into superstars

11. Propositions 2 and 4 relate to income growth in the superstar cities.

group and non-superstars groups, which we call the rest of the country. We estimate an equally weighted average of price–rent ratios for both groups by year. Figure 2.A.1 presents the results. The Figure shows that price–rent ratios have been increasing over time in superstar areas and in the rest of the country. However, in the superstar cities, price–rent ratios have increased much more, leading to a growing regional divergence in price–rent ratios.

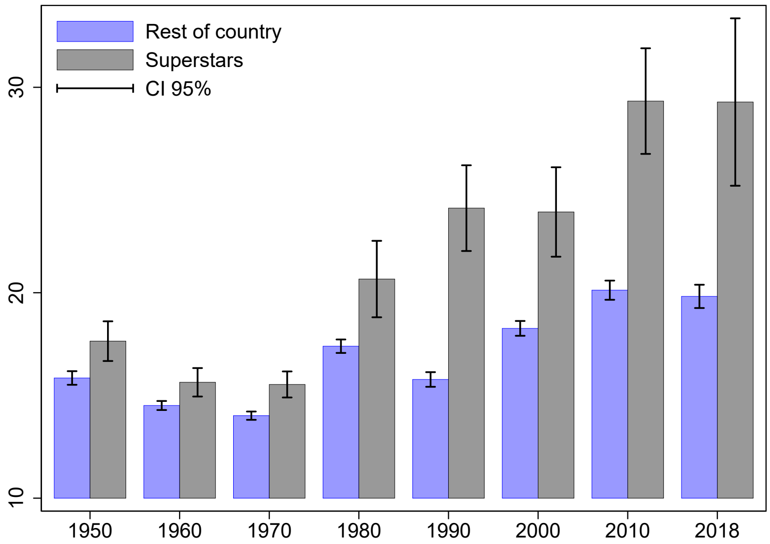

Figure 2.A.1. Price–rent ratios in the U.S., 1950-2018

Notes: We define superstar cities as cities that were at least once a superstar city between 1950 and 2000 according to the superstar definition in Gyourko, Mayer, and Sinai (2013). We extended the data from Gyourko, Mayer, and Sinai (2013) to 2010 and 2018. Each bar represents an unweighted average by year for the specific group. 95% confidence bands are shown in black.

The model developed by Gyourko, Mayer, and Sinai (2013) predicts that price–rent ratios are higher in superstar cities, but it does not account for the growing gap between superstars and non-superstars over time.

Appendix 2.B Model simulation of risk–free rate fall on house price divergence

To examine the scope conditions under which a falling discount rate leads to house price divergence between superstar and non-superstar (reservation) cities, we simulate our spatial Gordon model for a range of potential, and not always realistic, values. The result displays the house price divergence (in log) as a function of falling discount rates (in %) and is broken down for all possible combinations of differences in rent and discount rate growth rates between superstar and reservation city (2.B.1). The figure demonstrates that house price divergence occurs under a majority of calibrations, as long as the superstar rent growth excess and the reservation city excess discount rate is sufficiently high.

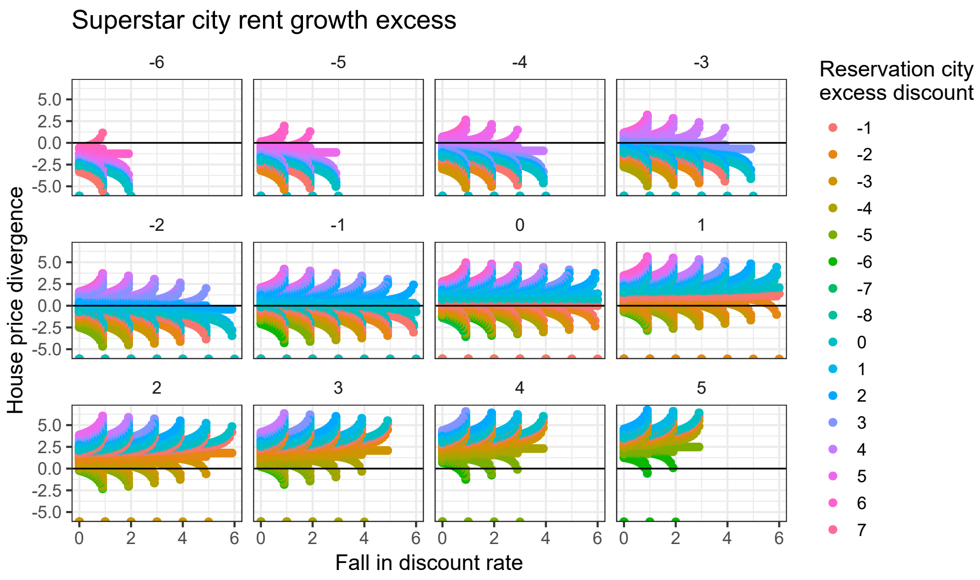

Figure 2.B.1. Simulation results by superstar rent growth excess

Notes: Facets show the percentage points by which the superstar citys' rent growth exceeds that of the reservation city. Colors indicate the percentage points by which the reservation citys' discount rate exceeds that of the superstar city.

Appendix 2.C Model evidence & simulation using US MSA-level data

We also use the US MSA-level data from Gyourko, Mayer, and Sinai (2013) was extended to 2018 in Amaral et al. (2021) to test our mechanism empirically. We want to replicate Figure 3 in the main paper empirically. Our mechanism predicts a one-to-one relation between rental yields in 1980 and in 2018, with a linear shift due to the fall in discount rates (compare Figure 3 panel (a) in the main paper). It also predicts a non-linear relation between rental yields in 1980 and price–rent ratios in 2018, with initially lower rental yield MSAs subsequently having disproportionately higher price–rent ratios (compare Figure 3 panel (b) in the main paper). As demonstrated below, these predictions hold to a great extent in the data.

Figure 2.C.1 panel (a) plots the rent–price ratios for all MSAs in 2018 relative to the rent–price ratios in 1980. It also shows a linear fit with the resulting regression coefficients. Rent–price ratios in 2018 can indeed be predicted by rent–price ratios in 1980 but have fallen uniformly by approximately 85 basis points. Of course, MSA-level rent–price ratios do not perfectly align with the regression line. This implies

that rent–price ratios have also been affected by city–specific shocks. Not all variation in rent–price ratios can be explained by a fall in discount rates alone, however, the linear fit can explain 48 % of the variation in the data.

(a) Rent-price ratio 2018 **(b)** Price-rent ratio 2018

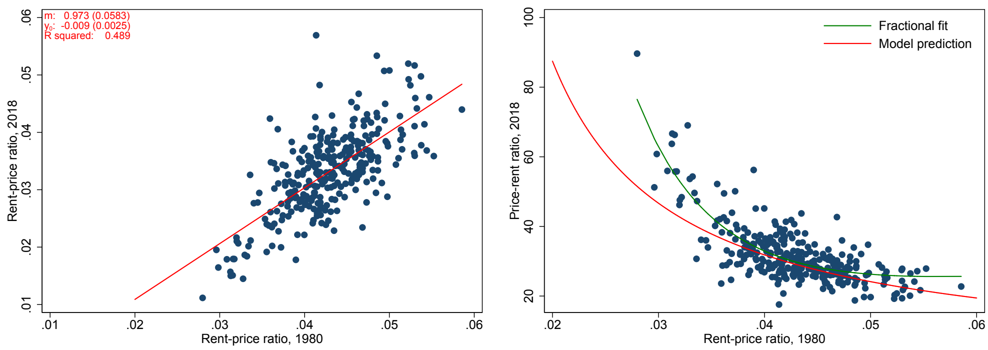

Figure 2.C.1. Comparison model and US MSA-level data

Notes: Panel (a) shows the rent–price ratios in 2018 relative to the rent–price ratios in 1980 together with a linear fit and the resulting regression coefficients (standard errors in parentheses). Panel (b) shows the price–rent ratio in 2018 relative to the rent–price ratio in 1980 together with a fractional fit and the predictions of our model resulting from the linear fit in Panel (a). The data is taken from Gyourko, Mayer, and Sinai (2013) and extended by Amaral et al. (2021).

Panel (b) of Figure 2.C.1 plots price–rent ratios in 2018, also presenting a fractional fit to the data (green line). The red line depicts the price–rent ratios that the model would predict for 2018, given the rent–price ratio in 1980 and the uniform fall in rent–price ratios estimated in panel (a). Again, the model does not fit the data perfectly, however, agree with the overall picture of the data and predicts higher price–rent ratios for cities that already had low rent–price ratios in 1980. The fact that price–rent ratios in cities with the lowest rental yields initially are even higher than predicted by the model leaves some room for alternative explanations. One example would be increasingly more optimistic rent expectations (g) in the superstar cities relative to the rest of the country. Another would be a tightening of supply constraints in superstar cities.

References

Adler, David, and Ben Ansell. 2019. "Housing and Populism." *West European Politics* 43 (2): 344–65. DOI: 10.1080/01402382.2019.1615322. [223]

Amaral, Francisco, Martin Dohmen, Sebastian Kohl, and Moritz Schularick. 2021. "Superstar Returns." *FRB of New York Staff Report*, (999): [224, 225, 228–232, 234, 235, 240, 241]

Ansell, Ben W. 2019. "The Politics of Housing." *Annual Review of Political Science* 22 (1): 165–85. DOI: 10.1146/annurev-polisci-050317-071146. [223]

Begley, Jaclene, Lara Loewenstein, and Paul S. Willen. 2021. "The Rent-price Ratio during the 2000s Housing Boom and Bust: Measurement and Implications." *Work in progress*, [224, 227]

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* 107 (5): 614–20. DOI: 10.1257/aer.p20171036. [236]

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. DOI: 10.1093/rfs/1.3.195. [225]

Demers, Andrew, and Andrea L. Eisfeldt. 2021. "Total Returns to Single-Family Rentals." *Real Estate Economics*, (May): DOI: 10.1111/1540-6229.12353. [224, 227, 228]

Fama, Eugene F., and Kenneth R. French. 2002. "The Equity Premium." *Journal of Finance* 57 (2): 637–59. DOI: 10.1111/1540-6261.00437. [236]

Garriga, Carlos, Rodolfo Manuelli, and Adrian Peralta-Alva. 2019. "A Macroeconomic Model of Price Swings in the Housing Market." *American Economic Review* 109 (6): 2036–72. DOI: 10.1257/aer.20140193. [225]

Glaeser, Edward L., and Joseph Gyourko. 2007. "Arbitrage in Housing Markets." [227]

Gordon, Myron J. 1962. "The savings investment and valuation of a corporation." *Review of Economics and Statistics*, 37–51. [224, 232]

Gupta, Arpit, Vrinda Mittal, Jonas Peeters, and Stijn Van Nieuwerburgh. 2021. "Flattening the curve: pandemic-induced revaluation of urban real estate." *Journal of Financial Economics*, [223]

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar Cities." *American Economic Journal: Economic Policy* 5 (4): 167–99. DOI: 10.1257/pol.5.4.167. [224, 227, 231, 232, 238–241]

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott. 2018. "Tarnishing the golden and empire states: Land-use restrictions and the US economic slowdown." *Journal of Monetary Economics* 93: 89–109. [223]

Hilber, Christian, and Andreas Mense. 2021. "Why Have House Prices Risen so Much More than Rents in Superstar Cities?" [224, 228, 231, 232]

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions." *Journal of Economic Perspectives* 19 (4): 67–92. DOI: 10.1257/089533005775196769. [225]

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2): 1–39. DOI: 10.1257/mac.20170388. [223]

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): 1225–98. DOI: 10.1093/qje/qjz012. [229, 230, 234]

Kroen, Thomas, Ernest Liu, Atif R Mian, and Amir Sufi. 2021. "Falling Rates and Rising Superstars." [225]

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2020. "The Expected Return on Risky Assets: International Long-run Evidence." *SSRN 3546005*, [225, 236]

Miles, David Kenneth, and Victoria Monro. 2019. "UK House Prices and Three Decades of Decline in the Risk-Free Real Interest Rate." *SSRN Electronic Journal*, DOI: 10.2139/ssrn.3508653. [225]

Negro, Marco Del, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2019. "Global Trends in Interest Rates." *Journal of International Economics* 118 (May): 248–62. DOI: 10.1016/j.jinteco.2019.01.010. [225, 236]

Nieuwerburgh, Stijn van, and Pierre-Olivier Weill. 2010. "Why Has House Price Dispersion Gone Up?" *Review of Economic Studies* 77 (4): 1567–606. DOI: 10.1111/j.1467-937x.2010.00611.x. [224, 226]

Piazzesi, Monika, and Martin Schneider. 2016. "Housing and Macroeconomics." In *Handbook of Macroeconomics*. Elsevier, 1547–640. DOI: 10.1016/bs.hesmac.2016.06.003. [228, 231]

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel. 2020. "Segmented Housing Search." *American Economic Review* 110 (3): 720–59. [231]

Poterba, James M. 1984. "Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach." *Quarterly Journal of Economics* 99 (4): 729. DOI: 10.2307/1883123. [223]

Rachel, Lukasz, and Lawrence Summers. 2019. "On Secular Stagnation in the Industrialized World." Working paper. DOI: 10.3386/w26198. [225, 236]

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78. DOI: 10.1086/261120. [226]

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." *Current Issues in Urban Economics*, 74–104. [226]

Sagi, Jacob S. 2021. "Asset-Level Risk and Return in Real Estate Investments." *Review of Financial Studies* 34 (8): edited by Stijn Van Nieuwerburgh, 3647–94. DOI: 10.1093/rfs/hhaa122. [231]

Chapter 3

Freedom of Enterprise and Economic Development in the German Industrial Take-Off*

3.1 Introduction

What role did institutions play in inducing the Industrial Revolution? What was the effect of the enlargement of markets and an increase in aggregate demand? The importance of institutions in modern economic growth has been emphasised since the work of Douglas North and, more recently, by Acemoglu et al.[1] These authors claim that inclusive institutions are necessary to create the incentives required to invest in and industrialise production processes. The role of aggregate demand, on the other hand, has been stressed by Murphy, Shleifer, and Vishny (1989) and in reference to trade by Pascali (2017). The underlying mechanism is that demand must be sufficiently high to enable increasing returns to scale (IRS) technologies to break even. The decrease in transportation costs attributable to railways and steamships as well as political unification prompted the expansion of trade and the enlargement of markets. This boosted aggregate demand and might explain the Industrial Revolution.

In this paper, I analyse the industrial take-off in Germany. I show that the interaction of an institutional reform and an increase in market size triggered industrialisa-

* I would like to thank Francisco Amaral, Moritz Schularick and Kaspar Zimmermann for their extensive feedback. I am grateful to Benjamin Born, Dmitry Kuvshinov, Sheilagh Ogilvie, Paul Schäfer, Lingwei Wu and the participants of seminars in Bonn for their valuable comments. I gratefully acknowledge financial support by the DFG Research Training Group 2281, 'The Macroeconomics of Inequality'. An early version of this research project entitled "Entrepreneurial Freedom as a driving Institution during the German Industrial Take Off" was submitted as a master's thesis by Martin Dohmen.

1. See e.g. North and Thomas (1973), North and Weingast (1989), Acemoglu, Johnson, and Robinson (2005a) and Acemoglu (2012)

tion and economic growth. On their own, neither an increase in aggregate demand nor a reform towards better institutions was sufficient to launch industrialisation.

Proponents of institutional theories argue that a stimulating institutional change - for example, the introduction of property rights - improves economic incentives. This, in turn, induces agents to introduce IRS technologies in industrial production. An implicit assumption is that IRS technologies are more efficient than prior production processes; this is only true, however, if economic conditions are favourable and aggregate demand is sufficiently high. As a consequence, institutional innovations might only have been effective when combined with major reforms in economic conditions such as tariff laws and market regulations.

Investigation of the interaction between institutional innovation and economic reform encounters one major difficulty. Significant changes in institutions and economic conditions typically happen on the same geographical level. Earlier empirical studies (such as the seminal work by Acemoglu et al.[2]) have analysed the effect of institutions on country-level. These studies struggled to disentangle the effects of an institutional improvement from a changing economic environment because of the lack of an adequate control group. Using regional variation within rather than between European countries, Tabellini (2010) demonstrates that economic development varied across regions featuring the same formal institutions related to cultural or historical conditions. He, thereby, illustrated that a constant environment is needed to estimate the causal influence of institutional reforms.

In this paper, I show that an institutional innovation combined with an economic reform could trigger industrialisation in an otherwise constant economic environment. To do so, I use three distinctive features of German history: one dealing with the institutional change, one with the economic reform and one with the constant environment.

At the onset of Germany's Industrial Revolution, the country was divided into independent states. They differed institutionally but shared a single culture, language and history. This variation was also investigated by Acemoglu, Cantoni, Johnson, and Robinson (2011), who found evidence that German states that underwent radical institutional reforms during the French occupation subsequently grew at a more rapid pace. I examine a geographically narrower setting to rigorously separate the effect of institutions from other factors influencing economic development and to analyse the interaction with a change in economic conditions. In contrast to the aforementioned authors, I focus on a particular institutional innovation, the introduction of the so-called *Gewerbefreiheit*, which consisted of the abolition of guilds and the unconditional freedom of enterprise. The *Gewerbefreiheit* created new incentives to industrialise manufacturing production, whereas the previous guild system

2. See e.g. Acemoglu, Johnson, and Robinson (2001) and Acemoglu, Johnson, and Robinson (2005a)

hindered the transformation from craftsman's workshop to factory. The *Gewerbefreiheit* therefore was directly connected to the Industrial Revolution.

I demonstrate that this institutional innovation alone did not initiate industrialisation. To make the industrialisation of production processes profitable, the market for manufacturing goods needed to be sufficiently large, such that IRS technologies were able to break even. This is when the second distinct feature of German history comes into play. Whereas in other countries political integration and decreasing transportation costs gradually led to the establishment of larger markets, the political separation of Germany prevented this process. Only when German states unified their markets upon the foundation of the German *Zollverein* (German customs union) did the market size increase considerably. This created the necessary prerequisite for industrial production to become efficient.

German history thus makes it possible to empirically test whether the interaction of an institutional innovation and an increase in market size was able to induce economic growth. The political separation enables me to compare a state with freedom of enterprise to a state that featured restrictive guilds, both before and after the foundation of the German *Zollverein*. To ensure that the treatment and control units shared a constant economic environment, I exploit the third singularity of German history, the division of the Kingdom of Westphalia. The allocation of territories between Prussia and the Electorate of Hesse (the 'Electorate') works as a suitable natural experiment since the division did not follow any economic logic. It was decided at the Congress of Vienna and was part of a package deal linked to the division of the Kingdom of Saxony. Additionally, the division of Westphalia was determined by population numbers with no attention being paid to economic potential or factors correlated with industrial development. Therefore, it can credibly be assumed to have been exogenous to factors associated with later economic growth. This natural experiment allows me to rely on the identification strategy used by Michalopoulos and Papaioannou (2013). These authors utilized the fact that drawing a random border through a culturally homogeneous territory ensures the cultural similarity of the treatment and control units. Geographic proximity accounts for other potential drivers of economic growth.

I conduct a difference-in-differences analysis to compare economic development in the Prussian and Hessian counties, which were both part of the former Kingdom of Westphalia. The identification assumption requires that, in the absence of treatment, counties in both states would have developed in a similar manner. To validate this assumption, I perform three tasks. First, in a historical discussion, I provide evidence that new border locations were exogenous and assignment to the treatment or control group was therefore close to random. Second, a pre-treatment regression demonstrates that, after the division of the Kingdom of Westphalia but before the foundation of the *Zollverein*, counties developed similarly in both states. The introduction of the *Gewerbefreiheit* alone was thus not able to induce economic development. Lastly, I use adequate control variables to ensure that other driving forces of industrialisation suggested by the literature did not bias my estimates.

The results indicate that after the foundation of the Zollverein, counties with *Gewerbefreiheit* had a considerable developmental advantage compared to counties featuring the guild system. As regards population growth, my major proxy for economic development, Prussian counties grew, on average, 19-23 log points more over the 27-year sample period; this effect is economically large and statistically highly significant. Alternative proxies of economic development support this interpretation.

To confirm that this relationship is indeed causal, I perform two robustness checks. First, I zoom in on one newly defined border between Prussia and the Electorate to utilise a spatial regression discontinuity design. The results demonstrate that the growth advantage of Prussian counties was truly caused by state affiliation. Second, I zoom out and include the two other large Hessian states in my analysis. This shows that the effect revealed was due to the difference in institutions and was not attributable to other peculiarities of Prussia or the Electorate.

In the Appendix, I discuss other institutional differences between the states in question and draw the conclusion that the disparity regarding freedom of enterprise was the only first-order difference influencing economic growth. To validate that the abolition of guilds was indeed the main driver of the uncovered growth differential I analyse two additional outcome variables. According to the channel I propose, counties that had abolished guilds already when the *Zollverein* was founded should, first, feature larger manufacturing shares later on as they had a comparative advantage in manufacturing production. Second, within manufacturing, the share of self-employed workers like independent craftsmen should be lower, as more manufacturing sectors should already have introduced IRS technologies and, thus, small workshops should have been replaced by larger factories. I examine occupational shares surveyed in the Prussian census in 1882 and confirm both predictions empirically.

My results have direct implications for three strands of the economic literature. First, they relate to the literature investigating institutions as a fundamental cause of growth.[3] The interaction of institutional innovation and the enlargement of markets demonstrates that it is necessary to analyse which institutions are compatible with certain social and economic conditions. Unlike Acemoglu, Cantoni, et al. (2011), my findings suggest that a superior institutional setting, which simply needs to be implemented to induce prolonged economic growth, might not exist. Apart from having the right institutions, social and economic conditions must be favourable to supporting development.

My results are comparable to the results of two previous papers showing that the benefits from international trade depend on the institutional setting. Acemoglu,

3. Overviews may be found e.g. in Acemoglu, Johnson, and Robinson (2005b), Ogilvie and Carus (2014) and Fuchs-Schündeln and Hassan (2016)

Johnson, and Robinson (2005a) showed that the rise of the Atlantic trade generated a significant positive impact on development only in European countries with favourable institutional settings. Pascali (2017) demonstrated that the short-run effect of the first globalisation of trade on economic growth was positive only for countries characterised by more inclusive institutions. My results provide further evidence that the effect of trade or, more generally, aggregate demand, depends on the institutional setting. Moreover, I demonstrate that both elements - favourable institutions and high aggregate demand - are required. Favourable institutions alone are not sufficient to induce economic development.

Second, my paper contributes to the literature on customs unions. Several papers analyse the German *Zollverein* as the "pioneer and by far the most important customs union" (Viner (1950) p. 97). Voth (2001) surveyed the existing literature challenging the common view that the *Zollverein* was important for the economic or political development of Germany and asked for more evidence. Subsequent literature mainly examined the foundation of the *Zollverein*. Ploeckl (2015) argued that the sequential accession of states can be explained in a bargaining model with Prussia as the agenda setter. Wolf and Huning (2019) demonstrated that the foundation of the *Zollverein* was a consequence of physical trade costs and the allocation of territory towards Prussia at the Congress of Vienna. According to the authors, the new borders implied that revenue-maximizing states were incentivized to join a trade union with Prussia. Their model also explains why the Hessian states in my analysis joined the *Zollverein* after the Southern German customs union was formed and reassures that this decision was not primarily driven by expected development advantages. Comparable to my analysis, the authors also argue that Prussia's new borders have been exogenous to trade. They do, however, not look at the economic consequences of joining the *Zollverein*. Regarding the effects of joining the *Zollverein* on the countries economy, Keller and Shiue (2014) showed that it causally fostered market integration. The authors examined the convergence of wheat prices between member states controlling for the endogenous relationship between trade and membership and find that bilateral price gaps between member cities fell by about one-third with the implementation of the *Zollverein*. This provides evidence that the foundation of the *Zollverein* indeed increased market size as separate markets became integrated subsequently. Ploeckl (2013) additionally demonstrated that, in the state of Baden, entry to the *Zollverein* increased the size of the manufacturing sector and shifted the occupational structure towards higher income occupations within the craft sector. I contribute to this literature by showing that economic gains from joining the *Zollverein* were unequally distributed. States with a liberal institutional framework were able to utilize the potential of increased market integration, whereas the economic gains for states with impeding institutions were minor. Hence, if states aim at fostering economic development by joining a customs union next to maximizing revenue, they would probably benefit from implementing a liberal institutional setting.

A related paper is Keller and Shiue (2016). The authors analysed the effects of institutional reforms and market integration on city growth also in 19th century Germany. They found that market integration had a significant effect on city population growth and that the independent effect of institutional differences was small. They also provided evidence that the length of French occupation during the Napoleonic Wars increased market integration later on and argued that more liberal institutions increased market integration that in turn fostered development. In accordance with the authors, I argue that the effect of institutional reforms without an increase in market size or market integration was small. I do, however, suggest an alternative mechanism for how institutions and market integration interact in supporting economic development. I show that counties with favourable institutions featured higher economic growth after an equal and exogenous increase in market integration through the foundation of the *Zollverein*.[4] In Keller and Shiue (2016) probably both mechanisms were at play, which would explain the large effects of market integrations on city growth that the authors find.

Finally, my results relate to the literature concerning craft guilds. This literature addresses whether guilds fostered or hindered economic development (for overviews see Ogilvie (2004), Ogilvie (2014) and Epstein (2010)). One side argues that guilds have been an efficient solution to overcoming market failures in various domains. Their arguments have been criticised by Ogilvie and others, who argued that guilds have been primarily rent-seeking organisations exploiting market power and hindering competition. In a recent paper, De la Croix, Doepke, and Mokyr (2017) developed a theoretical model to demonstrate that guilds might have been beneficial for the intergenerational transmission of skills. I contribute to that literature by empirically assessing the causal impact of guilds on economic development. It is, however, not the aim of my analysis to take a stand in the debate regarding whether guilds hindered or fostered economic growth in general. It might still be true that, long before the Industrial Revolution, guilds constituted an appropriate institution to facilitate economic progress. The advancing enlargement of markets is first required to create a favourable environment for institutions supporting craftsmen to industrialise production processes. In the model world of De la Croix, Doepke, and Mokyr (2017), the introduction of the *Gewerbefreiheit* could also be interpreted as the transition from the guild equilibrium to the even superior market equilibrium.

The remainder of this paper is organised as follows. Section 2 discusses the historical background and mechanisms at play; Section 3 describes the data. Section 4

4. In my spatial regression discontinuity design specification, after the *Zollverein* was established, market integration was arguably very similar for municipalities on both sides of the border. Keller and Shiue (2016) measure market integration as price gaps in wheat prices between cities. In my setting, price gaps between close-by villages were probably very low after tariff barriers vanished, because a no-arbitrage condition would imply that price gaps were not larger than transportation costs. Still, the municipalities with favourable institutions were able to benefit disproportionally from the increase in market integration.

explains my empirical strategy; Section 5 presents the main results and Section 6 adds two additional Hessian states to the analysis. The last section concludes the paper.

3.2 German history as a natural experiment

In this section, I discuss the three distinct features of German history that lay the foundation for my empirical analysis. I first describe the introduction of the *Gewerbefreiheit*, contrasting it with the old guild system. I then illustrate the enlargement of markets induced by the foundation of the German *Zollverein*. Lastly, I discuss the separation of the Kingdom of Westphalia at the Congress of Vienna, which is fundamental to my identification strategy and ensures that the treatment and control units are comparable culturally as well as in other relevant dimensions.

However, I would like to illustrate two points to contextualise the following discussion. First, I give a brief overview of the timing of the various events I discuss later. Table 1 shows a timeline to structure this overview. A more detailed discussion, including references to the relevant historical literature, can be found in Appendices A1 and A2. The first focuses on the history of the Kingdom of Westphalia and the second on the spread of the *Gewerbefreiheit* in Germany.

Table 3.1. Timeline of historical events

1792	Start of the Coalition Wars
1807	Treaties of Tilsit and founding of the Kingdom of Westphalia
1808-11	*Gewerbefreiheit* established in the Kingdom of Westphalia and Prussia
1814/15	The Congress of Vienna
1816	Restoration of the guild system in the Electorate of Hesse
1834	Foundation of the German *Zollverein*
1864	Start of the Unification Wars

In 1792, the Coalition Wars between revolutionary France and other European great powers began. Under Napoleon's leadership, France conquered Germany and overturned the old order. In 1807, he dictated the Treaties of Tilsit and founded the Kingdom of Westphalia. The new ruler implemented far-reaching economic and social reforms, including the introduction of the *Gewerbefreiheit*. Prussia, which was defeated by Napoleon, triggered similar reforms around 1810. In 1813, however, Napoleon was defeated by the allies and a new German order was defined at the Congress of Vienna in 1814-1815. There, the Kingdom of Westphalia was divided and the German Confederation, an association of 39 German-speaking states in Central Europe, was established. This event defines the environment of my empirical analysis; it remained constant until the German unification wars started in 1864. This marks the end of my sample period. In 1866, Prussia annexed the Electorate of Hesse.

Two events took place in-between those dates that are crucial for my later analysis. In 1816, the Electorate of Hesse reintroduced the old guild system in the counties of the former Kingdom of Westphalia now under its governance, thereby creating the institutional differences between the treatment and control units. The second prerequisite for industrialisation was the foundation of the German *Zollverein* in 1834, which stimulated aggregate demand. To make clear why (only) the combination of the *Gewerbefreiheit* and the *Zollverein* might have triggered industrialisation, as a second overview point, I will present a simple mechanism to illustrate a possible channel.

The mechanism is based on Murphy, Shleifer, and Vishny (1989), who developed a closed economy model in which aggregate demand had to exceed some margin such that the country industrialised. To keep the discussion as simple as possible, I focus on manufacturing production and ignore the rest of the model economy. The authors assumed that each manufacturing good is produced in a separate sector that is small relative to the overall economy. Two technologies are available for producing a manufacturing good: manual labour featuring constant returns to scale and an IRS technology. Manual labour enables one unit of output q to be produced at a constant cost of $\alpha > 1$. The second technology, industrial production, requires a fixed investment cost of C and variable costs of 1 per unit of output. In the model, a sector industrialises if and only if demand is sufficiently high for the IRS technology to break even, meaning the following:

$$\alpha * q \quad > \quad C + 1 * q \tag{3.1}$$

In German states, before the founding of the *Zollverein*, most sectors were not industrialised, such that $\alpha * q_{State} \; < \; C + 1 * q_{State}$ must have been true in those sectors. I make two additional historical assumptions, which I justify below. First, I assume that the foundation of the *Zollverein* increased aggregate demand by some margin, meaning $q_{Zollv.} \; >> \; q_{State}$. Second, the introduction of the *Gewerbefreiheit* reduced fixed costs to industrialise such that $C_{Gewerbefr.} \; < \; C_{Guilds}$. Therefore, for some sectors in the economy, it will be the case that:

$$\alpha * q_{Zollv.} < C_{Guilds} + 1 * q_{Zollv.} \quad \& \quad \alpha * q_{Zollv.} > C_{Gewerbefr.} + 1 * q_{Zollv.} \tag{3.2}$$

If the number of manufacturing sectors in the economy for which both inequalities hold is sufficiently large, there will be a difference in industrialisation and, consequently, economic growth between the states that implemented the *Gewerbefreiheit* and the states that featured the old guild system.

In the following discussion outlining the first two distinct features of German history, I argue that both of these assumptions are credible: First, the fixed costs of setting up an enterprise under the guild system were larger than those under the *Gewerbefreiheit* and, second, the foundation of the *Zollverein* increased aggregate demand noticeably.

3.2.1 The introduction of the *Gewerbefreiheit* and abolition of guilds

Around the time of the industrial take-off, fundamental institutional changes occurred in Germany. One was the abolition of the guild system and the introduction of an institutional setting called *Gewerbefreiheit* (freedom of enterprise). In its most advanced form, the new institutional setting not only eliminated all guilds but established the unconditional freedom of enterprise. In the historical literature, however, the term *Gewerbefreiheit* is mainly used to describe the abolition of craft guilds. I will also constrain the following analysis to this particular part. This does not, however, cause a relevant restriction because at the onset of the industrial take-off, most non-agricultural production was still conducted by craftsmen. Moreover, the production processes formerly conducted by craftsmen were essentially those to be industrialised. Therefore, to understand why the introduction of the *Gewerbefreiheit* probably led to considerably lower fixed costs in establishing industrial production, it is crucial to analyse the barriers to industrialisation that were due to craft guilds instead. For details about the dissemination of the *Gewerbefreiheit* throughout Germany as well as the exact laws in Prussia, please refer to Appendix A2.

The two key channels through which guilds might have impeded the switch from manual labour to IRS technology are the reduction of incentives and the limitation of the potential of craftsmen to innovate. The following discussion will focus on these two points. It is based on a contemporary analysis by Bovensiepen (1909) of the guild laws in the Electorate introduced in 1816. However, the former situation in most other parts of continental Europe was comparable.[5]

Before the Industrial Revolution, most non-agricultural goods were produced using manual labour by craftsmen. These craftsmen tended to settle down in cities, where they formed guilds. Guilds were unions of self-employed members aimed at improving the production and living condition of craftsmen.[6] The guilds also ensured the equality of their members by managing the joint purchasing of resources and allocation to members of equal prices and proportionally equal shares, setting

5. For a broader analysis of the guild system in Europe between the 16th and the 19th centuries and its influences on and barriers to economic growth, consider, for example, Ogilvie (2014). Comparing the following discussion to this paper does reveal that the lines of argumentation used by Bovensiepen and Ogilvie are remarkably similar and demonstrates that the guild system in the Electorate of Hesse was not exceptional - apart from its longevity.

6. For example, by assuring the quality of their products, defining rules for apprenticeship and journeymanship and offering pensions and insurances.

rules defining the maximum number of journeymen and assistants per member, the prohibition of trading with products not self-produced and the obligation of having only one point of sale.[7]

Over the centuries, guilds organised and controlled industrial production and thus became very powerful. In most European countries, political power protected the guilds and ensured that everyone conducting a craft was required to be a member of the respective guild (*Zunftzwang*). This meant that only members were allowed to perform a particular craft and that all members had to obey the guild rules. In many cases, the guild system required craftsmen to settle in cities. To protect subsistence and urban prosperity, it was forbidden to execute most crafts in villages or the urban hinterland.[8] Therefore, the guilds established a quasi-monopoly on producing and selling certain products.[9]

At the onset of the Industrial Revolution, decreasing transportation costs and political integration led to an enlargement of markets and induced a profound structural transformation. At this point, guilds became more costly for economic development because they restrained the tendency to industrialise. Bovensiepen argued that on one hand, the illusory security guilds provided to their members as part of their monopoly hindered innovation because craftsmen were not incentivised to alter their production technologies. On the other hand, the guilds hindered the transformation of a crafts enterprise that produced for a local market to a factory that used new machinery and the division of labour to satisfy national demand; this blockade resulted from guilds limiting the work patterns of their members to the specific craft, limiting their sales market to their own workshop and capping their number of employees. Although factories were, in principle, excluded from guild rules, guilds still prevented the appearance of new factories by requiring that they needed to be built from scratch rather than developed out of the workshop of an entrepreneurial craftsman.[10] This way, huge amounts of capital were required to open a factory and the associated risk was large, making the barriers to entry very high. Ogilvie (2014) additionally argued that the rent-seeking behaviour of craft guilds had negative consequences on innovation. By regulating production processes, fixing minimum prices, excluding entry, requesting heavy human capital investment in a specific production technology through apprenticeship and imposing demarca-

7. See Bovensiepen (1909), page 6-7. The literature in favour of guilds argues that guilds were established because they helped to overcome market failures in various domains. Their quality controls reduced information asymmetries between producers and customers concerning the quality of goods, their rules for apprenticeship solved imperfections in markets for skilled training owing to moral hazard and their social capital and collective punishment helped to enforce contracts. See Ogilvie (2004) and De la Croix, Doepke, and Mokyr (2017).

8. Exceptions in the Electorate were blacksmiths, nailers, wheelwrights, carpenters, bricklayers, roofers, potters and weavers.

9. See Bovensiepen (1909) pp.7 and 17.

10. See Bovensiepen (1909) pp. 76-77.

tions between various craft guilds, they reduced incentives and increased the costs that are associated with innovation.[11]

To summarise, craft guilds impaired incentives for craftsmen to become entrepreneurs and blocked the transformation from workshops to factories. This prevented an efficient allocation of capital and considerably increased the costs of opening a factory. Consequently, it is probable that the manufacturing sectors regulated by guilds faced considerably higher fixed costs to establish industrial production, so $C_{Gewerbefr.} < C_{Guilds}$.

3.2.2 The foundation of the German *Zollverein*

As demonstrated in my later empirical analysis, the decrease in fixed costs after the introduction of the *Gewerbefreiheit* seemed to have been insufficient to make industrial production profitable. According to the described mechanism, an increase in aggregate demand was another necessary condition. The underlying model by Murphy, Shleifer, and Vishny (1989) relies on the assumption of costly world trade such that domestic demand becomes relevant. Next to transportation costs and the difficulty of penetrating foreign markets, they emphasised protectionism as a reason to expect costly world trade.[12]

This was especially relevant for German states at the beginning of the 19th century. Many potential trade partners, for example most European 'Great Powers', used high tariffs to protect their internal markets from foreign competition. In addition, potential German industrial production was unable to compete with the far advanced English industry. The Continental Blockade under Napoleon first destroyed German's overseas trade connection, but its removal then exposed German products to English competition.[13] International demand for German manufacturing goods was, therefore, almost non-existent.

The conditions inside the German territory were similarly bad. Economic policy was widely dominated by mercantilist ideas and protectionism. Around 1800, this fact, combined with the territorial separation, resulted in the presence of approximately 1,800 tariff borders being erected between and inside German states.[14] After the Congress of Vienna, the situation did not improve much. Although territorial reorganisation led to the enlargement of states and, therefore, of markets and many German states began eliminating their internal tariff borders, inner-German protectionism still suppressed aggregate demand. In 1818, Prussia implemented regulations without any internal tariffs but with external tariff borders and high transit duties. Although this system was comparably modern, without any export duties and only moderate import duties on manufacturers, these import duties still functioned

11. For the full arguments, see Ogilvie (2014) p. 185.
12. See Murphy, Shleifer, and Vishny (1989) p. 540.
13. See Hahn (1984) pp. 15-16 and 21 and Ohnishi (1973) pp. 7-9.
14. See Hahn (1984) page 10.

as protective barriers. Because tariffs were defined in nominal terms, decreasing prices for manufacturing goods led to increasing tariffs in relative terms.[15] These, combined with the high transit duties and strict border controls, had severely negative consequences for trade relations of Prussia's smaller German neighbour states, as a wave of complaints demonstrated. In reaction, other German states installed similar systems, as the Electorate of Hesse did in 1824.[16]

Because of this situation, the German states made several attempts to form tariff unions in different constellations. Prussia had a special interest in convincing the Electorate of Hesse and other German states to join its tariff system because it also suffered from the separation of its western territories from the eastern core area.[17] In the beginning, Prussian attempts were unsuccessful; however, the applied pressure provoked a tariff war and the efforts of several smaller German states to form a tariff union of their own. As these failed, the separation of markets persisted. At the end of the 1820s, Prussia intensified its efforts, mainly for political reasons, and finally convinced the Grand Duchy of Hesse to join a tariff alliance in 1828 by offering financial advantages. This reduced the resistance of the other German states. In 1831, the Electorate of Hesse joined the Prussian tariff union and in 1834 the German *Zollverein* (German customs union) was founded. Overnight, all tariffs and import restrictions between member states vanished. In 1835 and 1836, more states (e.g. the Grand Duchy of Nassau) joined; by 1842, 28 of the 39 German states became members. Keller and Shiue (2014) show that the new customs union substantially and causally increased market integration thereby unifying previously separated markets. The German *Zollverein*, therefore, finally led to the establishment of a large common tariff and economic area, which drastically increased total demand.[18]

For Prussia, the new common market did not only finally connect its two parts but also approximately doubled its market size in terms of population.[19] The Hessian

15. See Hahn (1984) pp. 16 and 21-23; when the reform was set up in 1818, duties on manufacturing were meant to not exceed 10% of the goods' value; however, duties were defined per item or per weight or size of the good, instead of relative to the value. Even in 1818, however, the underlying prices were in many cases already far from reality such that duties exceeded the ceiling of 10% by far. During the following years, decreasing values and prices increased relative tariffs further, but the Prussian government did not significantly adjust the tariffs; see also Ohnishi (1973) pages 77-91.

16. See Hahn (1984) pp. 24-25 and 38 as well as Seier: 'Modernisierung und Integration in Kurhessen 1803-1866' in Heinemeyer (1986) p. 451.

17. See Figure 1 as well as Hahn (1984) p. 18 and Nipperdey (1983) pp. 182 and 358-360.

18. See Nipperdey (1983) pp. 183 and 358-361.

19. In 1834, all of Prussia had 13,509,927 inhabitants, 3,685,309 of which lived in the western part and 9,824,618 in the eastern core country; the other states joining the *Zollverein* until 1836 together 11,413,021 inhabitants in 1834; using population numbers from HGIS Germany, 2006-2013, http://www.hgis-germany.de/ (last accessed: 29 January 2019).

states also largely benefited from access to the new tariff union,[20] which drastically increased their market size.[21]

In total, the foundation of the German *Zollverein* finally created a large common market replacing protective tariffs with free trade flows and common regulations. The size of the *Zollverein* grew even more over time as new member states entered. The assumption that $q_{Zollv.} >> q_{State}$ seems, therefore, to be highly credible for all member states.

3.2.3 The separation of the Kingdom of Westphalia

My identification strategy relies on the assumption that the assignment to treatment was quasi-random and that the treatment and control units were comparable in terms of all other relevant dimensions; this implies that the separation of the Kingdom of Westphalia and the location of new borders are assumed to have been independent of any economic considerations and exogenous to factors associated with later economic growth. This assumption is justified by historical circumstances.

After Napoleon's defeat, the new German order was defined at the Congress of Vienna, which was a meeting of ambassadors of European states led by the great powers of Prussia, Russia, Austria, Great Britain and France. Its primary goals were to secure peace in Europe through the balance of power and to allocate the conquered territories. However, it was dominated by power politics and single states trying to maximise their military influence. Territorial negotiations around the Prussian monarchy were driven by the promise of the allies provided during the war to re-establish Prussia's power of 1806. In practice, this meant to ensure that Prussia obtained the same number of inhabitants as it had before it was defeated by Napoleon. As Russia demanded to get all of Poland, the Prussian interests focused on annexing the entire Kingdom of Saxony. Both great powers formed a coalition to protect their claims. The other great powers, especially France and Austria, did not accept this allocation of territories and intended to secure at least some parts of the Kingdom of Saxony. The conflict got out of control and Europe again found itself very close to war. It could only be resolved when Great Britain, through an external shock - a peace agreement with the United States (US) - suddenly became more confident and started to mediate the conflict. Britain suggested that the Kingdom of Saxony should be divided and Prussia be compensated instead with large territorial gains in central and western Germany. This compensation included large parts of the Kingdom of Westphalia to obtain the promised number of inhabitants. This

20. According to Hahn regarding the Hessian states, the *Zollverein* increased trade, offered protection from international competition and supported the growth of existing and the creation of new industries by increasing demand. See Hahn, 'Der hessische Wirtschaftsraum im 19. Jahrhundert' in Heinemeyer (1986) pp. 398-399.

21. In 1834, the Electorate of Hesse had 700,583 inhabitants, which was below 3% of total population of all member states from 1836.

suggestion was motivated by the plan to form a protective barrier against France in western Germany. After further negotiations, Prussia had to accept this allocation. Consequently, and against its own will, it expanded to the west and annexed a considerable share of the former Kingdom of Westphalia (see Figure 1).

The compensation of Prussia in central and western Germany meant that there were fewer shares left to allocate among the smaller German states, which demanded rewards for joining the war against France. Consequently, although the Elector of Hesse had been a loyal ally against Napoleon, he was not able to enforce all of his territorial claims at the Congress, which included territorial expansions inside the former Kingdom of Westphalia.[22] Still, the Electorate of Hesse was refounded out of a sizeable part of the Kingdom of Westphalia as well as a smaller part that had previously belonged to the Grand Duchy of Frankfurt in 1812 (see Figure 1).[23]

In summary, the definition of the common borders between the Electorate of Hesse and Prussia depended directly on the decision about the division of Saxony, which was arranged by Great Britain and triggered by an exogenous shock. Their exact location was, therefore, not influenced by any economic considerations and was independent of the former history or geography of the territories in question; it was simply meant to attain the promised number of inhabitants for Prussia. Consequently, the partitioning can readily be assumed to be exogenous to later economic growth. For this reason, the division of the Kingdom of Westphalia is a clean natural experiment to study the effect of the various institutions of Prussia and the Electorate of Hesse. Moreover, both countries grew substantially compared to before the French Revolution and the new borders neither exactly lined up with the borders in the former Holy Roman Empire nor did they follow any significant geographical structure.

After the Congress of Vienna had taken place and all remaining territorial conflicts had been resolved in 1820, the territorial situation was as follows. The Electorate of Hesse was located in the middle of the German Confederation separating the western and eastern parts of the Prussian monarchy, meaning that it shared borders with Prussia in the west as well as the east (see Figure 1). In the west, the Electorate was adjacent to the newly founded Prussian Province of Westphalia, which comprised a part of the former Kingdom of Westphalia as well as parts of other states formerly under French influence.[24] In the east, the adjacent Prussian province was the Province of Saxony, which also contained sizeable parts of the former Kingdom of Westphalia.

22. The Elector of Hesse planned to annex larger parts in the west of the former Kingdom of Westphalia around Paderborn and Corvey, but these parts were needed to compensate Prussia.

23. For historical facts about the Congress of Vienna, see Stauber (2014) and Lentz (2014); for the territorial negotiations concerning the Electorate of Hesse, see Hundt (1996) p. 234.

24. Considering the borders of 1812.

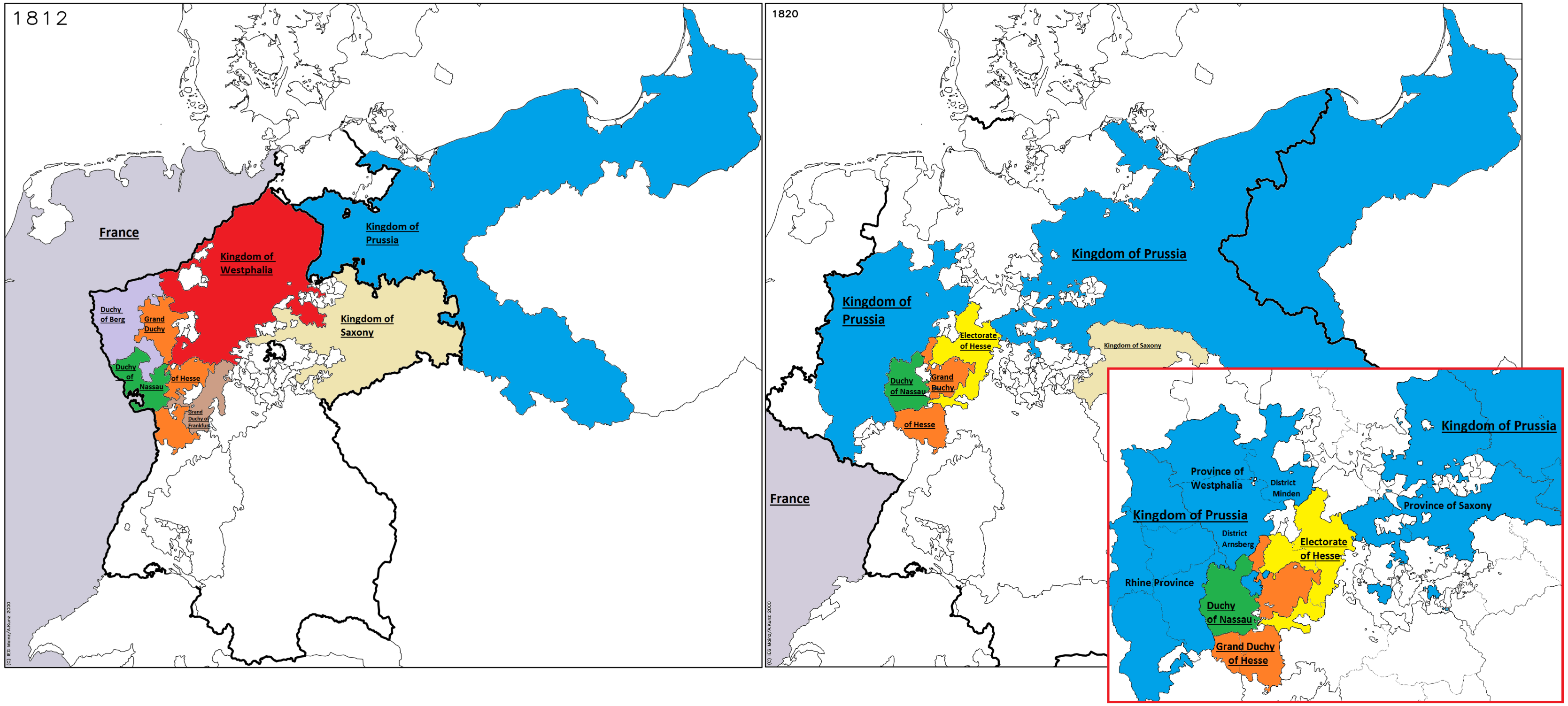

Figure 3.1. Map of Germany in 1812 and 1820

Notes: Marked areas are the relevant states mentioned in the text. The map in the red box zooms in on the Hessian states and shows province and district borders. Source: Hundt and Moschl (2004 a,b) and HGIS Germany (Kunz and Dietze, 2007).

The territorial composition implies that both adjacent Prussian provinces, as well as the Electorate itself, were strongly influenced by Napoleon and the French Revolution. This holds especially true for the territories in the former Kingdom of Westphalia. The extensive and radical institutional reforms rooted in the Kingdom of Westphalia combined with the very similar institutional and cultural background of the territories before Napoleon's conquest most probably resulted in a comparable cultural relationship with the new institutions, which ensures the same functioning of the latter in both states.[25]

After the Kingdom of Westphalia was divided at the Congress of Vienna, the parts belonging to Prussia and those to the Electorate experienced comparable institutional development. Both states showed attempts to return to the old order that Napoleon had overturned. This was, however, of limited success and most institutional changes made during Napoleon's occupation survived the restorative tendencies. The only major exception was the re-establishment of the guild system in the Electorate of Hesse. It can readily be assumed to be the only first-order institutional difference between counties of the former Kingdom of Westphalia later belonging to Prussia or the Electorate of Hesse during my sample period.[26] A detailed comparison of other institutional dimensions can be found in Appendix A3, which demonstrates that the only other institutional dimension that is suspected to have had a noticeable influence on economic development and might potentially have shown perceivable differences is education. For this reason, I control for the observable differences in educational outcomes in my later analysis. The other institutional reforms often considered in economic literature are the implementation of a modern legal order, the abolition of serfdom and agrarian reforms.[27] All have been comparable between Prussia and the Electorate as demonstrated in Appendix A3. Moreover, a modern legal order would probably have benefited the entire economy and the abolition of serfdom and agrarian reforms targeted the agrarian sector. Below, however, I show that after the treatment period the manufacturing share was larger in counties that belonged to Prussia and had abolished guilds already when the *Zollverein* was founded. This indicates that Prussian counties had a comparative advantage in manufacturing production compared to counties in the Electorate. The abolition of guilds is the reform that has the largest potential to explain this finding as it specifically affected the manufacturing sector.

25. See Acemoglu, Cantoni, et al. (2011), who make a similar argument for the reforms brought by Napoleon in general. For a more profound discussion of the history of the Kingdom of Westphalia, which explains why it is plausible to expect the same culture and cultural relation to new institutions please refer to Appendix A1.

26. This might be especially true after the Electorate of Hesse got a new constitution in 1832, see Acemoglu, Cantoni, et al. (2011).

27. See e.g. Acemoglu, Cantoni, et al. (2011) and Kopsidis and Bromley (2015).

3.3 Data

This section introduces the data I use and discusses how I approximate economic development on a regional level during a time when economic activity was not yet systematically measured. For the various data sources and matching procedures please refer to Appendix B.

Membership in the German *Zollverein* precipitated the obligation to hold regular censuses following common rules, to which the existence of good and comparable population data for all member states can be attributed. I use this population data to assess the effect of the institutional difference between Prussia and the Electorate of Hesse on economic growth. My treatment period starts in 1837 as this is the closest year to the foundation of the *Zollverein* that features good data availability. It ends in 1864 when the German unification wars started. I add 1849 as an intermediate year, as it roughly halves the time period and the necessary data is available, and I use population data from 1821 to examine the pre-treatment effect.

The units of observation in most of my empirical analysis are counties (German *Kreise*), which had an average of around 40,000 inhabitants. These still constitute a high level of aggregation and covered several cities and/or villages.[28] This results in a relatively low number of observations, although a large number of local economies is covered. Municipality level data is, however, only available for parts of the time period and sample area, and necessary control variables are missing. To attenuate concerns regarding the sample size, I broaden the sampled area in some specifications to include more counties. Furthermore, I use municipality-level data in section 3.5.3, which boosts the number of observations considerably.

I approximate economic development using population growth because regional data on GDP does not exist for this time period. I measure population growth as the log difference in population to reduce skew in the data and make results easy to interpret. Other proxies regularly used in the literature such as urbanisation rates and city population growth cannot be used on a county level or on a at least similarly finely grained grid.[29] County-level observations are necessary for my empirical strategy because they constitute comparable administrative units, allowing me to focus on a homogeneous sample and analyse only those regions that were geographically close to each other. However, population growth captures the same idea as these proxies, namely, that the population distribution is an indicator of the geographical distribution of economic activity. Consequently, when comparing proximate regions, a higher population growth indicates stronger economic development. On an aggregate level, population growth was also used by Clark (2014).

28. Most counties included around 50 to 100 municipalities.

29. Urbanization rates are, for example, used in Acemoglu, Cantoni, et al. (2011), Acemoglu, Johnson, and Robinson (2005a) and Acemoglu, Johnson, and Robinson (2002a) and city population growth in Fernihough and O'Rourke (2014), Cantoni (2015) and Keller and Shiue (2016).

Population growth approximates economic progress at the beginning of the Industrial Revolution for two reasons. First, a higher level of economic development led to longer life expectancies i.a. because of lower infant mortality rates owing to better nutrition. As birth rates were remarkably similar between the Electorate of Hesse and the adjacent Prussian provinces in the first part of the considered period and only differed slightly afterwards,[30] this translated directly into higher population growth rates. Moreover, the nearly identical birth rates during the first half of the sample period ensures that demographic transition processes do not distort my results. Second, during the considered time period, severe poverty led to high emigration, predominantly to America.[31] This phenomenon became important from 1831 onwards and was intensified by crop failures in 1846 such that in 1847, 80,000 people emigrated from the German Confederation. In 1854, this number increased to 239,000.[32] Since economic problems were the main motivation for leaving the German Confederation,[33] emigration was higher in poorer and less industrialised regions. Therefore, this led to a higher reduction in the number of inhabitants in those regions that experienced slower economic development. Statistically relevant internal migration, in contrast, happened only in form of urbanisation, whereby people from surrounding or nearby regions migrated to larger cities.[34]

Next to population growth, population density is often used as a proxy for economic development, especially for the time prior to the Industrial Revolution because in pre-modern economies, the Malthusian trap ensured that GDP per capita was constant over the long run.[35] During this time, land was an important factor of production. Therefore, higher productivity and a higher GDP gave rise to higher population density, which, in consequence, again reduced per worker productivity. Therefore, I proxy initial economic development using population density;[36] this is especially important as counties were designed in such a way that they had roughly the same number of inhabitants. The direct relationship between population density

30. From 1837 to 1848, the birth rate in the Electorate of Hesse was on average 3.6475% per year and averaged 3.65125% over both Prussian Provinces Westphalia and Saxony during the same time; from 1849 to 1864, it was 3.4% in the Electorate and 3.6954% in the Prussian Provinces, whereby the Province of Westphalia was closer to the Electorate with 3.5375% (Köllmann (1980)). Whether the later difference in birth rates was due to heterogeneous economic development is only speculative, but Nipperdey (1983) suggests that economic conditions positively influenced birth rates in the considered time (p. 109). For the Grand Duchy of Hesse and the Duchy of Nassau, data about birth rates is incomplete, but for the years it exists, it is close to the ones for the Electorate.

31. More than 90% of German emigrants went to the United States.

32. See Siemann (1995) pp. 90-93.

33. Comparing prices for staple food like potatoes and rye with emigration numbers reveals that in times of large price increases emigration numbers peaked; see Siemann (1995) pp. 90-92.

34. See Nipperdey (1983) p. 112.

35. See, for example, Acemoglu, Johnson, and Robinson (2002a); this paper also documents a high correlation between population density and urbanisation rates.

36. The positive relation between productivity and population density results from the Malthusian model and was empirically shown by Ashraf and Galor (2011).

and productivity is another reason to proxy economic development using population growth, because in my analysis I hold observational areas fixed. Population growth rates are, therefore, identical to growth rates in population density.[37] I additionally use population density directly as an alternative outcome variable in a robustness check.

As a third indicator for economic development, I follow the work of Becker and Woessmann (2009) and use the income taxes per capita in 1877-1878 as a proxy for county income during that year. This might be the most reliable indicator of relative economic progressiveness of the counties in Prussia but lies outside my sample period. However, comparing this measure with my main measure, population growth, yields a correlation of 0.56 using the population growth between 1864 and 1877 and a correlation of 0.61 using that between 1837 and 1877.[38] This suggests that population growth is a reliable proxy for economic growth, especially during the first years of my sample period, when industrialisation in Germany had just begun and the Malthusian trap was still active. To shed more light on the source of economic development, I additionally use the share of the workforce working in manufacturing and the share of self-employed workers within manufacturing on county level as alternative outcome variables. The necessary data were only collected in 1882, which is also outside my sample period. Still, these variables arguably provide information on developments in the preceding decades.

I use additional county-level control variables in my regressions; in section 5.3, I switch to municipality-level data to zoom in on one border between Prussia and the Electorate. For this exercise, I change my sample period to 1837 through 1849, as municipality-level data for Prussia were unavailable for 1864.[39]

3.4 Empirical strategy

In this section, I outline the regression technique used in my empirical analysis and discuss the identifying assumption. I exploit the division of the Kingdom of Westphalia and the foundation of the German *Zollverein* as a natural experiment to test the hypothesis that the institution of the *Gewerbefreiheit*, when combined with a considerable increase in market size, induced economic growth. To do so, I focus on

37. Population density is defined as population divided by area; therefore, growth in population density in a constant area equals growth in population. In logs: $log(\frac{pop_{i,t}}{area_i}) - log(\frac{pop_{i,t-1}}{area_i}) = (log(pop_{i,t}) - log(area_i)) - (log(pop_{i,t-1}) - log(area_i)) = log(pop_{i,t}) - log(pop_{i,t-1})$.

38. Using my baseline sample, all counties of Prussia and the Electorate of Hesse, which belonged to the former Kingdom of Westphalia, except city-counties, as the necessary data are missing (see Appendix B).

39. The data sources for municipality data also contain the year 1858, but for the sake of comparability with the county-level data, I use the year 1849 as the outcome year. All results also hold for using 1858 as the end of the sample period.

counties in Prussia and the Electorate of Hesse, which belonged to the former Kingdom of Westphalia,[40] and use a difference-in-differences regression of the following form:

$$y_{i,t} = \alpha_i + \delta_t + \beta_t * \delta_t * Gewerbefreiheit_i \\ + \gamma_t * \delta_t * Pop.\ density_{i,1821} + X_i'\zeta_t * \delta_t + \epsilon_{i,t} \quad (3.3)$$

Here, $y_{i,t}$ is population in county i at time t in log points.[41] I include county fixed effects, α_i, to control for population differences between counties that are constant over time and time fixed effects, δ_t, to capture the common growth trends. $Gewerbefreiheit_i$ is a dummy variable that takes the value of one if the county is located in Prussia and, thus, featured the *Gewerbefreiheit* at time t and zero if it is located in the Electorate. Therefore, the β_ts are the coefficients of interest, the growth advantage in log points of counties with *Gewerbefreiheit* at time t. From the second specification onward, I control for population density in 1821, the start of my sample period, interacted with the time fixed effects as a proxy for time varying effects of already existing economic development differences and for whether a county is rural or urban. X_i contains a set of additional, time-invariant, county-level control variables used in later regressions. As Moran's I points towards spatial autocorrelation, I use the spatial correction proposed by Conley (1999) to adjust standart errors.[42] Standart errors are additionally adjusted to account for serial autocorrelation and heteroscedasticity following Newey and West (1983).[43]

The parallel trend assumption - the identifying assumption in a difference-in-differences regression - requests that, in absence of treatment, counties in both states would have grown at the same rate. For this to be plausible, they need to show the same growth trend before treatment, that is, before the *Zollverein* was founded. This implies that the coefficient of the Gewerbefreiheit in 1837, β_{1837}, should be close to zero as the *Zollverein* had only just been established. Below I show that this is indeed the case.

However, to ensure that the counties in both states would have grown at the same rate after the foundation of the *Zollverein* had it not been for institutional differences,

40. I include Minden and Lübbecke from the district Minden in Prussia, although they only belonged in Westphalia until 1810 and then to France until 1813/14, because the introduction of the *Gewerbefreiheit* in Westphalia already took place in 1808-1810 and because France had similar laws. The results remain the same and, if anything, become stronger if these two counties are excluded.

41. I multiply log population by 100, such that the coefficients of interest measure the difference in population in log points.

42. I use a distance cut-off of 50 km for county-level data, beyond which I assume that the correlation between the error terms of two observations is zero. All results are robust to using different cut-off values. Results for the main specification can be found in Appendix 3.C.1.

43. I use the Stata command *acreg* developed by Colella, Lalive, Sakalli, and Thoenig (2019) with the *spatial*- and the *hac*-option to compute standard errors.

it is crucial that they are comparable in all dimensions except for the treatment. Table 2 contains summary statistics for all of the counties in the former Kingdom of Westphalia in 1837. It shows that, although the counties in Prussia were larger, on average, in terms of population, their size lies approximately in the same range as those in the Electorate of Hesse. In contrast, Prussian counties were much more heterogeneous as regards area. This is due to the way each state defined their counties. They were, in principle, designed to each include an approximately constant number of inhabitants, but Prussia had already established the so-called city-counties (*Stadtkreise*) that covered the area of larger cities separated from their hinterlands. In the Electorate, conversely, counties always also covered the surrounding areas of cities. There are two such city-counties in my sample that have comparably small areas and consequently a high population density. For this reason, population density is much more heterogeneous in the Prussian sample and - averaged over all of the counties - also higher. This does not mean that the overall population density was higher in Prussia. In fact, considering only areas of the former Kingdom of Westphalia, 69.7 inhabitants lived per square kilometre (km^2) in the Prussian parts and 72.1 in the parts belonging to the Electorate of Hesse. Therefore, initial development measured by overall population density was quite similar between both states and, if anything, higher in the parts belonging to the Electorate.[44] To ensure that this difference in the design of counties does not influence my results, I complementarily ran all regressions using a dummy for the city-counties or alternatively excluded them from the sample. In both cases, the treatment effect remained virtually unchanged.[45] Furthermore, using municipality-level data, in section 6.4, I was able to control more sharply for the structure of urban development than using county-level population density.

Apart from county size and population density, there might have been other systematic differences between Prussian and Hessian counties that were relevant to economic growth. I am going to discuss potentially important influential factors one after another and control for them in my regressions.

At first, one of the most relevant driving forces of economic development during industrialisation was the availability of natural resources, especially coal.[46] In Germany, hard coal as well as soft coal mining developed rapidly during the industrial

44. The higher heterogeneity in population density in the Prussian parts of the Kingdom of Westphalia also does not necessarily point at a higher urbanization rate. According to the data of Acemoglu, Cantoni, et al. (2011), compared to the whole Electorate, the urbanization rate was higher in the Province of Saxony covering the eastern part of the former Kingdom of Westphalia in Prussia, but lower in the Province Westphalia excluding the Mark, which covered the western part, in 1800 as well as in 1850. The available data do, however, not allow one to calculate urbanisation rates solely in the parts that have belonged to the former Kingdom of Westphalia.

45. Results are available upon request.

46. Fernihough and O'Rourke (2014) show that the availability of coal had a strong positive influence on city population size in Europe during the Industrial Revolution.

Table 3.2. Summary statistics

	Electorate of Hesse				Prussia			
	Mean	StdDev	Min	Max	Mean	StdDev	Min	Max
Population 1837	32899	10237	21291	62727	39234	8371	17291	59284
Area in km^2	455.99	98.31	320.97	614.85	562.83	280.51	24.90	1309.13
Pop. density 1837	73.81	25.08	49.81	154.36	133.61	231.71	29.88	1061.95
Working mining 1882	0.19	0.22	0.00	0.77	1.09	2.79	0.00	13.60
Literacy rate 1871	93.87	1.31	92.04	97.11	94.93	2.51	87.50	98.13
Share Protestants 1871	91.46	11.70	53.51	97.48	76.45	36.22	1.67	99.68
Fortification	0.00	0.00	0.00	0.00	0.06	0.25	0.00	1.00
Observations	14				31			

take-off, as coal was a necessary energy source for considerable aspects of industrial production. Another crucial input factor for industrial production was iron ore, which was needed to produce steel. Nevertheless, the presence of mineable natural resources in the former Kingdom of Westphalia was comparably low. Unfortunately, county-level data about mining do not exist for the sample period. Table 3 presents per capita figures for the mining of hard coal, soft coal and iron ore, in 1850,[47] considering the entire Electorate of Hesse compared to the area composed of the Prussian Province of Saxony and district of Minden, which together covered all parts of the former Kingdom of Westphalia in Prussia.[48] It can be seen that no region was mining a significant amount of hard coal or iron ore.[49] Moreover, the Electorate of Hesse mined more as regards tonnes per capita of both natural resources, so any resulting bias would work against finding a positive and significant treatment effect.

Still, the mining of soft coal was substantially higher in the Prussian territories than in the Electorate. As hard coal was being used more and more as an energy source, especially for the production of iron and steel during the industrial take-off, soft coal deposits might not have played a major role in economic progress. This intuition is supported when comparing the district of Minden with the Electorate of Hesse. Column 3 shows that mining of all three natural resources was considerably lower in Minden. In section 5.3 below, which uses municipality-level data to

47. 1850 is the first year with data available for all units showed here; from HGIS Germany, 2006-2013, http://www.hgis-germany.de (last accessed: 11 December 2018).

48. Approximately 74% of counties in the Electorate of Hesse belonged to the former Kingdom of Westphalia and 61% of counties covered by the Province of Saxony and the district of Minden. This is the most comprehensive coverage possible with the available data.

49. As comparison, in the district of Arnsberg, which covered the Ruhr area, the major coal-producing area in Germany, the mining of hard coal in 1850 was approximately 1.55 tonnes per capita and of iron ore was 0.11, at least one order of magnitude higher than in the considered regions.

examine the growth effect of the *Gewerbefreiheit* directly at the north-western border between the Electorate and Prussia, only Prussian municipalities are included, which are located in the district of Minden. This demonstrates that the results are robust to soft coal mining.

Table 3.3. Mining of natural resources in 1850

	Electorate	Prov. Saxony & district Minden	District Minden
Hard coal	0.1100	0.0174	0.0211
Soft coal	0.1141	0.4920	0.0002
Iron ore	0.0167	0.0098	0.0161

Notes: The table shows the mining of natural resources in tonnes per capita in 1850. Column 1 includes all of the counties in the Electorate of Hesse. Column 2 combines the Prussian Province of Saxony and the district of Minden, which together include all parts of the former Kingdom of Westphalia in Prussia. Column 3 includes the district of Minden. *Source:* HGIS-Germany, 2006-2013.

Another way to adjust for the availability of natural resources is to directly control for the relevance of mining on the county level. A drawback is that the necessary data only exist from 1882 onward, after Prussia annexed the Electorate of Hesse in 1866 and started collecting occupational data. This year lies far outside my sample period. However, as industrialisation advanced, natural resources became even more important and mining increased.[50] Therefore, mining in 1882 proxies the availability of mineable natural resources during the sample period. I thus include the share of the labour force working in mining in 1882 interacted with time fixed effects as control variables in my regressions. These variables capture the potentially time-varying effect of the availability of natural resources on economic development.

Next to the availability of natural resources, widespread education is considered to have contributed to Germany's rapid pace of industrialisation. Prussia was especially well-known for its well-functioning primary education system during the second half of the 19th century.[51] Therefore, controlling for education is potentially important because there might have been a systematic difference between Prussia and the Electorate of Hesse. Following Becker and Woessmann (2009), I approximate education by the literate share of the population. Literacy was first surveyed in Germany in the Prussian census of 1871. The census data contain the number of people aged 10 or older who were or were not able to read and write. Thus, it

50. For example, in 1850, hard coal mining averaged over all of Prussia was approximately 0.266 tonnes per capita; in 1882 it was already 1.712 tonnes per capita in Prussia, which at this time also included the Electorate of Hesse and other German states.

51. Compare Becker and Woessmann (2009) p. 548; Prussia's educational system is viewed as being partly responsible for the shift of industrial leadership from Great Britain to Germany.

should provide a reasonably clear picture of the success of the two educational systems in the years prior to the census and characterise the educational background of the population.[52] Using this data, I include the literate share of the population on the county level interacted with time fixed effects in my regressions. Note that controlling for education after the end of the sample period is an especially difficult test for the effect of treatment on economic growth because not only does education plausibly drive economic development but also economic prosperity might have led to a higher level of education. The control variable therefore might also capture developmental advantages realised during the sample period.

I also control for the share of Protestants living in a county in 1871 interacted with time fixed effects.[53] As denomination was already predetermined centuries before and arguably remained constant thereafter,[54] the fact that the census year occurs after the sample period does not play a significant role.

Another potential concern is that Prussia increased its military presence in the newly annexed territories, especially close to the new borders. This would have not only led to an increase in the population of these areas, but potentially also to considerable investments in infrastructure, which might have further influenced economic development. However, Prussian defence policy focused on larger cities and the further western regions, especially the Rhine valley, to protect Germany against France, which was situated far west of the former Kingdom of Westphalia.[55] Nonetheless, I use a dummy variable to control for the presence of Prussian fortresses in my sample. As Prussia established only a few, albeit large ones, only two such fortresses were located in the former Kingdom of Westphalia.[56] Excluding the respective counties from my regressions leads to similar results. The Electorate of Hesse did not possess any fortresses.

52. Prussia also surveyed primary school enrolment in 1849, but the necessary data for the Electorate is missing. For Prussian counties, which belonged to the Kingdom of Westphalia, the correlation between primary school enrolment in 1849 and literacy in 1871 is positive (correlation coefficient 0.23); Source: iPEHD (see Appendix B). I do, however, believe that the literacy rate in 1871 provides a more comprehensive approximation of the success of the two educational systems as it comprises the entire population older than 10 instead of looking at one specific cohort.

53. The reason is that Becker and Woessmann (2009) cannot completely rule out substantial effects of Protestantism on economic growth.

54. See Becker and Woessmann (2009) p. 555.

55. In fact, in Prussian counties formerly belonging to the Kingdom of Westphalia only on average 0.96% of the population were counted as military population in 1864. This is below the average share of all Prussian counties (1.17%); Source: iPEHD (see Appendix B). Data on the military population is missing for the Electorate.

56. The seven counties that I coded as possessing a fortress featured an average share of military population in 1864 of 7.24%. This is far above the average share of all Prussian counties (1.17%); Source: iPEHD (see Appendix B). The two counties possessing fortresses located in the former Kingdom of Westphalia had an average share of military population of 6.72%. These are the only two outliers in this sample as all other Prussian counties formerly belonging to the Kingdom of Westphalia had a share of military population below 3%.

3.5 Results

This section contains the empirical analysis of the growth differences between counties still having the *Gewerbefreiheit* and counties again featuring the guild system after the foundation of the German *Zollverein*. I exploit the partitioning of the Kingdom of Westphalia between Prussia and the Electorate and the new borders defined at the Congress of Vienna.

3.5.1 Baseline results

Table 4 illustrates the results of the difference-in-differences regression presented above using population in log points between 1821 and 1864 as the outcome variable. Column 1 states that the population in a county of the former Kingdom of Westphalia, which later belonged to Prussia and kept the *Gewerbefreiheit*, did not grow more between 1821 and 1837, on average, compared to the average population growth in counties belonging to the Electorate of Hesse. After 1837, however, counties featuring the *Gewerbefreiheit* grew significantly more than counties again featuring the guild system. This applies to both the period between 1837 and 1849 as well as the period between 1849 and 1864. In 1864, population in counties featuring the *Gewerbefreiheit* had grown by 21.17 log point more since 1821. This implies that, between 1837 and 1864, Prussian counties had a growth advantage of around 23 log points. In contrast, counties in the Electorate only grew by 2.11 log points between 1837 and 1864 according to the time fixed effects.[57] The difference is economically large and statistically highly significant. Controlling for a time varying effect of population density in 1821 only slightly reduces the estimate, but reduces standard errors perceptibly. The coefficient on population density for 1864 is positive and highly significant, showing that more densely populated counties, such as larger cities, grew significantly faster after the foundation of the *Zollverein*. This might indicate that already more developed areas had a further developmental advantage during the industrial take-off. However, it will primarily reflect the strong urbanisation during Germany's Industrial Revolution, which started at the beginning of the 19th century and intensified around 1850.[58]

Column 3 additionally controls for time varying effects of the share of Protestants in each county, the literacy rate in 1871 and the share of the workforce working in mining in 1882. Column 4 furthermore adds the fortification dummy. Including these control variables barely changes the size of the treatment effect, yet again re-

57. The time fixed effects in this regression have been approximately 18.23 log points in 1837, 23.80 log points in 1849 and 20.34 log points in 1864.

58. See Nipperdey (1983) p. 112.

Table 3.4. Difference-in-difference regression: Log population between 1821 and 1864

	(1)	(2)	(3)	(4)	(5)
Gewerbefreiheit * 1837	-1.815	-1.392	-0.215	-0.735	-4.348
	(3.223)	(2.957)	(2.750)	(2.916)	(3.616)
Gewerbefreiheit * 1849	6.302**	5.782**	6.739***	5.794**	2.309
	(3.070)	(2.727)	(2.575)	(2.649)	(3.570)
Gewerbefreiheit * 1864	21.17***	19.24***	19.57***	18.36***	13.14***
	(4.293)	(3.661)	(2.469)	(2.673)	(3.801)
Population density * 1837		-0.00788	-0.00858	-0.0126**	0.000472
		(0.00614)	(0.00603)	(0.00513)	(0.000401)
Population density * 1849		0.00968	0.00592	-0.00146	0.00230***
		(0.00841)	(0.00696)	(0.00377)	(0.000388)
Population density * 1864		0.0360***	0.0264***	0.0169***	0.00348***
		(0.00998)	(0.00817)	(0.00362)	(0.000416)
Constant	3.41e-15	3.45e-15	3.47e-15	3.51e-15	-4.49e-15
	(0.636)	(0.537)	(0.326)	(0.305)	(0.387)
Observations	180	180	180	180	380
R^2	0.918	0.925	0.931	0.934	0.947
County & time FEs	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	P&L&M	All	All

Notes: The table shows difference-in-difference estimates at the county level. The dependent variable is population in log points. Only counties belonging to the former Kingdom of Westphalia are included, except for Column 5, which includes all of the counties in the Electorate of Hesse and the adjacent Prussian provinces. All regressions include county and time fixed effects. Column 3,4 and 5 additionally include controls for the share of Protestants and the literacy rate in 1871 and share of the labour force working in mining in 1882 interacted with time fixed effects. Column 4 and 5 also include the fortification dummy interacted with time fixed effects. Standard errors (in parenthesis) are Conley-adjusted with a distance cut-off of 50 km and corrected for heteroscedasticity and autocorrelation (HAC). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

assuringly reduces standard errors and increases the R^2.[59] For 1864, all coefficients of the additional control variables are positive, but the one belonging to the share of the labour force working in mining is statistically insignificant.[60] This accords with the minor importance of mining in the analysed regions showed above.

As the sample size, when focusing on the counties belonging to the former Kingdom of Westphalia, is limited, Column 5 shows the same regression including all of the counties of the Electorate of Hesse and the two adjacent Prussian provinces.

59. Quantile regressions demonstrate that the estimate for the Gewerbefreiheit is large and highly significant over the entire distribution instead of being driven by specific outliers, see Appendix Table 3.C.4.

60. See Appendix Table C.8.

This more than doubles the number of observations. Although this specification is not as clean as using the division of Westphalia as a natural experiment and the considered counties are more heterogeneous, it is reassuring that the effect in 1864 remains sizeable and highly significant.[61]

To summarise, between 1821 and 1837, Prussian counties, which kept the *Gewerbefreiheit*, did not have any population growth advantage relative to comparable counties in the Electorate again featuring the guild system. This result supports the parallel trend assumption and indicates that the institutional difference did not have any positive effect prior to the existence of the German *Zollverein*. After the German *Zollverein* was founded, however, Prussian counties grew significantly faster in terms of population than comparable counties in the Electorate. The next section demonstrates that the advantage in population growth indeed seems to indicate faster economic development.

3.5.2 Alternative outcome variables

In this subsection, I use cross-sectional regression for counties that formerly belonged to the Kingdom of Westphalia to support my main results. First, I focus on alternative measures of economic development as outcome variables to demonstrate that the difference in population growth uncovered beforehand indeed captures a development advantage of Prussian counties. Second, I focus on occupational shares to provide additional empirical support for the mechanism I suggest.

To analyse alternate proxies for economic development I rely on population density in 1864 and income taxes per capita in 1878. To use these variables, I had to exclude the Prussian city-counties, because using population density as an outcome variable, these exceedingly large outliers render the present variation unimportant. Note that, if anything, this will bias the treatment effect towards zero, as large cities were especially densely populated and also grew faster owing to urbanisation. For income taxes as the dependent variable, the necessary data are not available for city-counties.[62] Of course, both outcome variables do not constitute perfect indicators of economic development. The fact that income taxes per capita are measured outside my sample period in 1878 is a valid concern, as this year lies after the German unification wars and unification itself, which had a large influence on economic performance. Using these additional outcome variables can nevertheless help to val-

61. The smaller and insignificant estimate in 1849 compared to the specifications focusing on the former Kingdom of Westphalia might stem from the fact that the guilds have not been abolished totally in some of the other parts of Prussia during the first half of the sample period; see Appendix A2.

62. For the specifications using income taxes, this also applies to the county of Cassel in the Electorate of Hesse because it became a city-county after the Electorate was annexed by Prussia and before the data on income taxes were collected. Excluding Cassel also in specifications using population density increases the estimates by some margin; however, in 1864, Cassel was still joint with its hinterlands. Compare Appendix B and Becker and Woessmann (2009) Appendix I B.

idate that the higher population growth in counties with *Gewerbefreiheit* resulted from a higher level of economic progress.

The results can be found in Table 5, which indicate that, after controlling for the initial distribution of population, population density in 1864 was significantly higher in the Prussian counties with *Gewerbefreiheit*. Column 1 indicates that in the Prussian counties, 16 more persons resided per km^2, on average, compared to the overall population density in the Prussian parts of the former Kingdom of Westphalia of 91 inhabitants per km^2 and 75 in the parts belonging to the Electorate of Hesse. So, in less than 30 years, the Prussian parts overtook the Hessian parts by far. After controlling for the other drivers of economic development, the treatment effect changes only slightly and the standard errors decrease.

Table 3.5. Alternative outcome variables

	Pop. density 64		Income tax pc 77/78		Share manuf. 82		Self-employed in manuf.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gewerbefreiheit	16.02***	14.13***	0.546***	0.529***	2.328**	3.236***	-6.288*	-7.247**
	(3.160)	(2.334)	(0.208)	(0.126)	(1.172)	(1.249)	(3.667)	(3.118)
Pop. density 1837	1.233***	1.243***	0.00235	0.00563*	0.00895***	0.00866***	-0.0113***	-0.00641
	(0.0476)	(0.0216)	(0.00309)	(0.00289)	(0.00181)	(0.00319)	(0.00395)	(0.00464)
Share Protestants		0.0455		0.00537**		0.0331**		-0.0990***
		(0.0435)		(0.00220)		(0.0167)		(0.0174)
Literacy rate		1.862***		0.178***		0.0608		-0.812
		(0.689)		(0.0658)		(.)		(0.569)
Working in mining		0.659		-0.120***		-0.348**		0.0373
		(0.528)		(0.0252)		(0.166)		(0.374)
Fortification		12.08***		0.0816		-2.245		0.116
		(2.828)		(0.197)		(1.760)		(1.292)
Constant	-14.23***	-194.0***	1.881***	-15.48**	9.466***	0.813	44.64***	129.6**
	(3.488)	(65.24)	(0.213)	(6.282)	(0.872)	(.)	(2.639)	(53.17)
Observations	43	43	42	42	45	45	45	45
R^2	0.923	0.948	0.160	0.611	0.282	0.365	0.224	0.454

Notes: The table shows OLS estimates at the county level. In Columns 1-2, the dependent variable is population density in 1864 in inhabitants per km^2, in Columns 3-4 income tax per capita in 1877-1878 in marks, in Columns 5-6 the share of the labour force working in manufacturing in 1882 by county in percent and in Columns 7-8 the share of self-employed workers within the workers in manufacturing in 1882 by county in percent. Only counties belonging to the former Kingdom of Westphalia are included and in Columns 1-4 all city-counties are excluded. Standard errors (in parenthesis) are adjusted for spatial autocorrelation following Conley (1999) using a distance cut-off of 50 km. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

Columns 3 and 4 contain the results using income taxes per capita in 1877-1878 as the outcome variable. They show that, even after the German unification and after 12 years the Electorate already belonged to Prussia and had the *Gewerbefreiheit*, income taxes per capita were still significantly higher in the counties that already had the *Gewerbefreiheit* during the sample period. After controlling for other influential factors, income taxes per capita were, on average, 0.53 marks higher in the counties that already belonged to Prussia when the *Zollverein* was established. This corresponds to approximately 22% of average income taxes per capita in the sample area.

Next, I provide empirical support for the hypothesis that the lasting abolition of craft guilds was indeed a major driver of the growth advantage of Prussian counties after the *Zollverein* was founded. If the mechanism I outlined in Section 2 holds, then, at the end of my sample period, the manufacturing share in Prussian counties should be higher than in counties that belonged to the Electorate. The reason is that Prussian counties would have developed a comparative advantage in the sectors in which production already switched to the IRS technology in these counties, but stayed at manual labour in the Electorate. This difference would probably persist even after Prussia annexed the Electorate. Moreover, a switch to the IRS technology would imply that more (larger) factories should have arisen in these manufacturing sectors and the share of self-employed craftsmen should have decreased.

I can test both hypotheses using occupational data from the Prussian census in 1882. Columns 5 and 6 of Table 5 show that the share of the labour force working in manufacturing was indeed higher in counties that already belonged to Prussia beforehand even 16 years after the Electorate was annexed by Prussia. According to Column 6, after controlling for other influential factors, the manufacturing share was three percentage points higher in counties that featured the *Gewerbefreiheit* already during my sample period. This estimate is highly significant and economically large compared to an average manufacturing share of all counties belonging to the former Kingdom of Westphalia of around 12 percent. Columns 7 and 8 demonstrate that within the manufacturing sectors the share of self-employed workers was significantly lower in 1882 in counties that already featured the *Gewerbefreiheit* during the sample period. This indicates that more or larger factories have arisen in these counties and the share of self-employed craftsmen has decreased. Again, with a difference of around seven percentage points the effect is economically large compared to an average share of self-employed workers of 39 percent and statistically significant.

In total, the results above show that the counties of the former Kingdom of Westphalia that belonged to Prussia and maintained the *Gewerbefreiheit* after the Congress of Vienna industrialised faster after the foundation of the *Zollverein*. They had a significant and persistent developmental advantage relative to comparable counties in the Electorate of Hesse wherein the old guild system was reintroduced. This development advantage came along with a larger manufacturing sector and a lower share of self-employed craftsmen within this sector. This is what would be expected if the *Gewerbefreiheit* was the main reason for the development advantage.

3.5.3 Regression discontinuity using municipality-level data

In this subsection, I 'zoom in' on the border to Prussia in the north-west of the Electorate of Hesse to reinforce the causal interpretation of the estimates presented above.[63] To do so, I change my unit of observation to the municipality (*Gemeinde*), which is the smallest administrative unit in Germany. This enables me to focus on regions close to the border. This way, I can use a spatial regression discontinuity approach. The idea is that other factors influencing economic growth, e.g. geographical variation or climate conditions, change continuously over space, whereas institutions change discontinuously at the state border. By focusing on a small bandwidth around the border, geographical proximity captures other drivers of economic growth that will be approximately equal on both sides of the new, exogenous border. Additionally, I include the distance to the new border as a control variable in my regressions that I multiply by minus one for all municipalities that are located in the Electorate. This control variable captures all remaining variation in omitted variables that change continuously over space (from the Electorate towards Prussia). For this reason, the fact that I am unable to use the control variables applied so far, which do not exist on the municipality-level, might not have a significant impact on my results.[64] Because of data limitations, I focus on the time period between 1837 and 1849.

The border in question separated the Electorate of Hesse from the Prussian district of Minden in the Province of Westphalia. Thus, it was located in the middle of the former Kingdom of Westphalia. For the reasons outlined above, its exact location was quasi-random and the border cut through a homogeneous territory. Importantly, the border did not follow a river or any other significant geographical structure and did not directly match with any previous border of former administrative units. Although it was located close to a district border inside the Kingdom of Westphalia, which, in turn, followed older borders, enough municipalities changed the side of the border to make the inclusion of a fixed effect for the old district possible (see Figure 2). This way, I can control for any relevant factors influencing economic growth that differed between both districts of the former Kingdom of Westphalia as well as cultural differences determined by earlier borders. The common history and culture assumption is therefore highly credible when focusing only on municipalities very close to the border. This implies that any significant effect of the border location

63. The border in the south-west of the Electorate is less suitable for a robustness check, as it is not located inside the former Kingdom of Westphalia and much shorter; for the border in the east, the necessary municipality-level data are not available.

64. The only variable that might have differed systematically between counties on both sides of the border is education because Prussia might have had a superior educational system. However, literacy rates in 1871 in the border counties have been considerably similar (in the two ones in the Electorate, 0.9380% and 0.9347%; in the Prussian border counties, 0.9429% and 0.9351%, respectively). Prussian fortresses have been far away from the considered border (the only one in my baseline sample in the west of the Electorate was in Minden, which was located on the other side of the district of Minden).

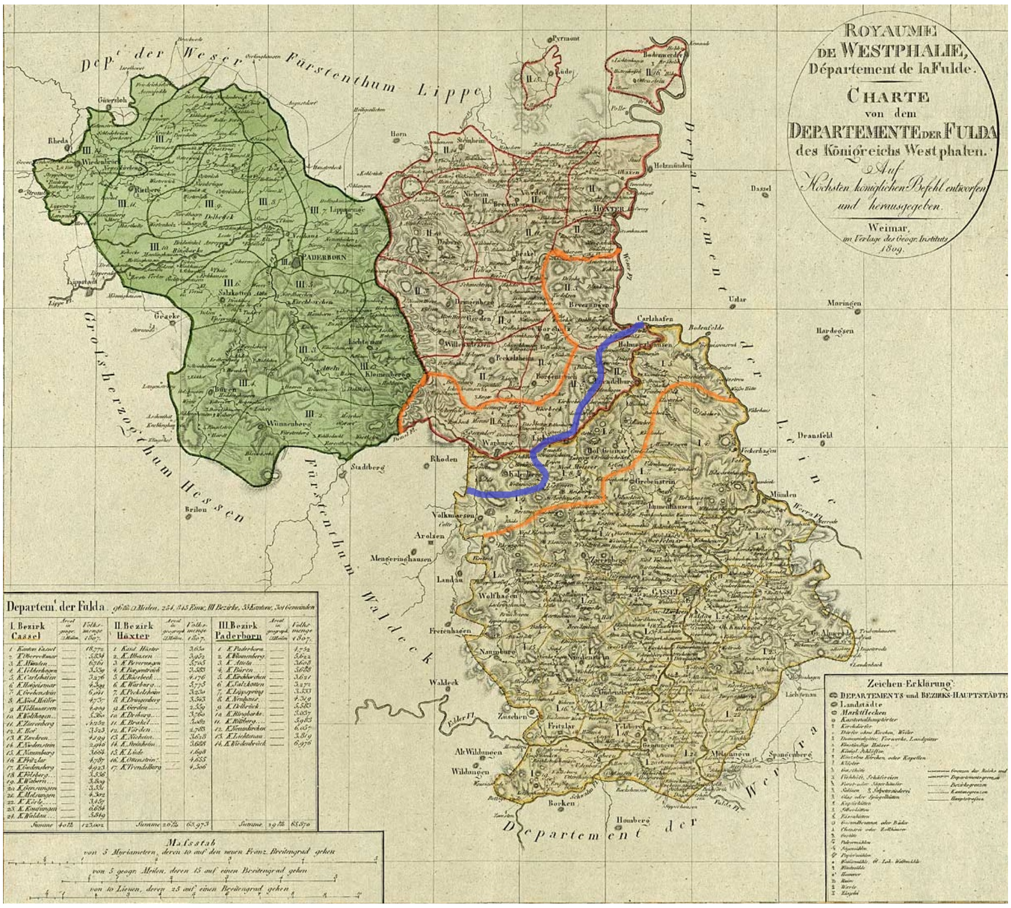

Figure 3.2. Map of the department Fulda of the Kingdom of Westphalia and the new border

Notes: The map shows the department Fulda, which was a part of the Kingdom of Westphalia. The yellow area is the district Kassel and the red one the district Höxter, which were administrative subdivisions of the department. The blue line shows the new border between the Electorate of Hesse and Prussia. The orange lines depict the sample of the Kanton specification in Table 6. Source: Smalian (1809).

found can only be driven by the different, new state affiliation and consequently the difference in institutions between Prussia and the Electorate of Hesse.

My baseline sample includes all municipalities that were located in the counties either next to the border or adjacent to counties that were, in turn, next to the border. In contrast to counties, municipalities were not designed to have approximately the same number of inhabitants. Instead, they were the smallest organisational units in existence that had their own administrations. They thus constituted a logical classification of a coherent group of people and could range from a small rural village to a large city. In my sample, they ranged from 19 to 31,000 inhabitants in 1837, with a mean of 749 inhabitants and a standard deviation of 1,658.

To assess the effect of the *Gewerbefreiheit* directly at the border, I use the following regression:

$$y_{i,1849} - y_{i,1837} = \alpha + \beta * Gewerbefreiheit_i + \gamma * Pop._{i,1837} + \zeta * Border\ dist._i + \epsilon_i \quad (3.4)$$

Again, $y_{i,t}$ is population in county i at time t in log points, such that the outcome variable is the difference in log population between 1837 and 1849. The constant α captures the common growth trend and β is the coefficient of interest. Instead of controlling for population density, I directly control for population size in 1837. The reason for doing so is that area data for the Prussian municipalities are missing and, as indicated above, for municipalities, the number of inhabitants is a reasonable proxy for initial development and for whether a municipality is urban or rural. The smaller observational units enabled me to control for the influence of urban development more sharply than possible with county-level data because urban areas are strictly separated from rural observations and clearly distinguished by their initial populations. I report the results weighted by the number of inhabitants in 1837 and unweighted results for some specifications. Using the municipality-level data, the difference between weighted and unweighted regressions is relevant, as the size of the observational unit is very heterogeneous. This implies that weighted regressions focus on growth in large municipalities such as cities and therefore emphasize urban development. Unweighted regressions, by contrast, are mainly influenced by rural areas. As done before, I use the spatial correction proposed by Conley (1999) to adjust standard errors.[65]

The results for the municipality-level regressions are shown in Table 6. Columns 1 and 2 include all of the municipalities in the sample, which results in 394 observations. In Column 1, observations are weighted by the number of inhabitants in 1837; in Column 2 observations are unweighted. In both regressions,

65. I use a distance cut-off of 10 km for municipality-level data, as the average distance between municipalities close to the border is considerably smaller compared to the average distance between counties in my main regression. All results are robust to using different cut-off values.

the estimate belonging to the dummy for *Gewerbefreiheit* is positive and significant. This demonstrates that close to the new border, the institutional difference between Prussia and the Electorate had a significant impact on economic development. The estimates for the unweighted regression are very similar to the weighted ones. This illustrates that smaller villages and rural areas likewise profited from the freedom of enterprise although larger municipalities grew more on average. This pattern can be explained by the fact that under the guild system, it was forbidden to conduct most crafts in rural areas. With *Gewerbefreiheit*, in contrast, also rural areas were allowed to develop manufacturing production.

Table 3.6. RDD at municipality-level: Difference in log population in 1837 and 1849 in log points

	(1) All	(2) All	(3) Adj	(4) Kantons	(5) Kantons	(6) Kantons
Gewerbefreiheit	5.631***	5.314**	5.078**	5.607**	5.047***	5.988***
	(1.910)	(2.128)	(2.355)	(2.231)	(1.845)	(1.777)
Pop. 1837 in 1000	0.237***	0.144*	0.0955	0.591*	0.603	0.396
	(0.0320)	(0.0802)	(0.577)	(0.355)	(0.380)	(0.689)
Border distance	-0.0308	-0.0110	0.0246	0.0597	0.00963	0.0717
	(0.0375)	(0.0339)	(0.0617)	(0.0810)	(0.124)	(0.131)
Old district Kassel					-1.561	0.664
					(1.095)	(1.479)
Constant	5.987***	6.533***	6.294***	6.635***	7.625***	6.392
	(1.037)	(1.109)	(1.410)	(1.144)	(0.588)	(.)
Observations	394	394	214	68	68	68
R^2	0.133	0.072	0.168	0.279	0.289	0.203
Weighted by pop.	Yes	No	Yes	Yes	Yes	No
Share Prussia	0.53	0.53	0.60	0.51	0.51	0.51
Mean border dist.	23.41	23.41	13.90	4.93	4.93	4.93
Max border dist.	55.00	55.00	36.30	14.47	14.47	14.47

Notes: The table shows OLS estimates at the municipality level. The dependent variable is log population in 1849 minus log population in 1837 in log points. Columns 1 and 2 include all of the municipalities in counties adjacent to the border or adjacent to adjacent counties. Column 3 includes only municipalities in counties adjacent to the border and Columns 4-6 include only municipalities in kantons from the former Westphalia that were divided by or adjacent to the new border. Columns 1 and 3-5 are weighted by the municipality population in 1837. The third to last row states the share of municipalities that are on the Prussian side of the border within the various samples. The two last rows contain the mean and the maximal distance to the border in km of all municipalities in the respective sample. Standard errors (in parenthesis) are adjusted for spatial autocorrelation following Conley (1999) using a distance cut-off of 10 km. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

One drawback is that the baseline sample still includes municipalities that are farther away from the border and quite heterogeneous. In particular, it still contains larger cities.[66] To focus on more homogeneous observations closer to the border, Column 3 shows the results of a regression only including municipalities that are located in counties that are directly adjacent to the border. This reduces the mean border distance within the sample from 23 to 14 km. In this specification, large cities are automatically excluded and the included region can be considered as rural.[67] Using this more homogeneous sample, the effect of the *Gewerbefreiheit* is very similar and the R^2 increases, indicating that the institutional difference explains more of the variation in growth rates. By contrast, as observations are more similar, the control for the initial population is closer to zero and insignificant. The coefficient of the border distance is also small and insignificant indicating that omitted variables that change continuously towards Prussia do not play a large role.

The next three columns present the results based on a sample of municipalities, which were located in the former kantons of the Kingdom of Westphalia adjacent to or divided by the new border, as depicted in Figure 2. Kantons were small territorial subdivisions of the French administrative system containing between 6 and 13 municipalities in my sample. They were established together with the Kingdom of Westphalia and did not correspond to any former administrative unit. The sample thus focuses on very similar municipalities grouped by the French occupation and very close to the new border, which, in turn, broke with the French administrative system. This results in a distance to the border of all of the observations of less than 15 km. Column 4 shows the results of a weighted regression based on this sample. The effect of the *Gewerbefreiheit* is remarkably stable and significant. Moreover, the R^2 increases further, as observations become even more similar owing to geographical proximity. Column 5 includes a dummy variable that takes the value of 1 for municipalities that belonged to the old district of Kassel. It captures any potential difference attributable to a different administrative arrangement during and before the existence of the Kingdom of Westphalia. However, the coefficient turns out to be insignificant and the estimate for the *Gewerbefreiheit* does not change remarkably. Column 6 shows the results of the same regression not weighted by population. Focusing more on rural areas does not seem to have a large influence, but somewhat increases the estimate.

To summarise, this section demonstrates that focusing on a small bandwidth and exploiting the discontinuity created by the new, exogenous border leads to significant results. Reassuringly, all estimates are around the same size as in the county-level regressions including all of the counties of the former Kingdom of Westphalia for the year 1849. The county-level estimates consequently reflect the causal effect of state affiliation between Prussia and the Electorate of Hesse on population growth.

66. In particular, the cities of Cassel and Paderborn are included, which were the only municipalities with more than 5.000 inhabitants in my sample.

67. Municipalities range from 19 to only 3,459 inhabitants in 1837, with a mean of 648 and a standard deviation of 596.

This effect was likely due to the difference in institutions between the *Gewerbefreiheit* and the guild system, as it was arguably the only first-order difference between both states influencing economic development during the sample period. The next section demonstrates that this conclusion is valid by examining additional states.

3.6 External validity: The inclusion of additional states

3.6.1 Choice of states

Although the investigation of the institutional differences between the Electorate of Hesse and Prussia constitutes the cleanest natural experiment to analyse the effects of the *Gewerbefreiheit* compared to the guild system, my results do not hinge on examining these two states. To demonstrate this fact, in this section, I include the Grand Duchy of Hesse and the Duchy of Nassau in my analysis. These additional states are suitable comparison units for multiple reasons.

The first state, the Grand Duchy of Hesse, was a representative state of the old guild system. While it was slightly more liberal concerning freedom of enterprise and institutionally more heterogeneous than the Electorate of Hesse, the guilds were still powerful throughout the entire sample period.[68] Additionally, the Grand Duchy also was a founding member of the German *Zollverein*.

Moreover, the Grand Duchy of Hesse constitutes a suitable comparison state because of its geographical proximity, large and exogenous border movements and equivalent historical influences. It was located in-between the Electorate of Hesse, Nassau and the western part of Prussia. During the Congress of Vienna, its borders changed fundamentally (see Figure 1). Together with the Duchy of Nassau, it even belonged to the states that had to accept the largest territorial movements relative to their size and population. The reasons were its late switch from Napoleon's side to that of the allies during the war and unintelligent negotiations at the Congress. To attain the promised compensation for Prussia after the decisions about Saxony, the Grand Duchy had to hand over large territories in its north-west region. In exchange, it received new territories in its south and west.[69] The Grand Duchy itself was a loyal ally of Napoleon and a member state of the Confederation of the Rhine, such that the French influence was large. In other institutional dimensions, the Grand Duchy resembled the other considered states, as discussed in Appendix A4. In summary, the Grand Duchy of Hesse lends itself to being used as an alternative control unit to illustrate that the difference in the freedom of enterprise drives the growth differential between the Electorate of Hesse and Prussia as opposed to other features exclusive to the Electorate.

68. For details on the laws concerning the guild system and the *Gewerbefreiheit* in the Grand Duchy and Nassau, please refer to Appendix A4.

69. See Hundt (1996) pp. 240-245.

The second state, the Duchy of Nassau was, next to Prussia, one of the most liberal states concerning the freedom of enterprise. For this reason, it can be used as an alternative treatment unit. This illustrates that the previously discovered effect is not due to a developmental advantage of Prussia, but can also be found when examining another, smaller state. Although Nassau was not a founding member of the German *Zollverein*, it joined the tariff union in 1836, two years after its foundation and before the beginning of my sample period. As Nassau belonged to the Hessian states and shared common borders with the Prussian Province Westphalia, the Rhine Province, the Electorate and the Grand Duchy of Hesse, geographical proximity is given. Its territory changed fundamentally over the course of the Congress of Vienna, because it also had to hand over large regions to Prussia and received territories of the former France and the Grand Duchy of Berg in exchange. This was due to decisions made during the war and its late switch to the side of the allies, which next to the decision about Saxony ensured that the border locations were exogenous to later growth.

There are, however, two noticeable differences between Nassau and the other Hessian states. First, next to freedom of enterprise, the only other relevant institutional difference was the late start of agrarian reforms in Nassau.[70] This might have constituted a disadvantage for economic development during the sample period. Second, Nassau had sizeable iron ore deposits at its disposal and its mining industry played a considerable role.[71] The mined iron ore was, however, mostly transported to the Ruhr area and not processed in Nassau such that the effect on overall economic development was probably small.[72] Overall, the Duchy of Nassau showed some minor differences to the other Hessian states, but it is still well suited to function as a robustness check to validate that the results found above are not due to a 'Prussia-effect'.

3.6.2 Results including the additional Hessian states

In this subsection, I analyse the effect of the *Gewerbefreiheit* using the Duchy of Nassau and the Grand Duchy of Hesse as alternative treatment and control units. I run a similar regression as Equation 3.4 but again using counties as observational units and controlling for population density instead of population size and border

70. See Appendix A4

71. From 1848 to 1857, around one per cent of the population in Nassau worked in iron ore mining.

72. The iron industry in Nassau was missing the access to cheap hard coal, so it could not compete with the Ruhr area, and it was cheaper to transport the mined iron there. For this reason, the iron industry in Nassau even lost market shares during the industrial take-off and was only able to develop after 1865, when a new railway line was built. See Hahn: "Der hessische Wirtschaftsraum im 19. Jahrhundert" in Heinemeyer (1986), pp. 418-419.

distance.[73] I alter the sample to include various states. One disadvantage is that I cannot use the control variables previously used in this setting as the necessary data do not exist.

The results can be found in Table 7. For Column 1, I use as treatment units all of the counties located in the Duchy of Nassau instead of Prussia and as control units all of the counties in the Electorate of Hesse. The effect of the institutional difference between Nassau and the Electorate is large and highly significant. It is comparable in size to the estimates when using Prussia as the treatment unit. This indicates that the effect found previously was caused by the institutional difference between the Electorate, which featured the guild system, and states with the *Gewerbefreiheit*, as opposed to a pure Prussia-effect.

Column 2 represents a placebo regression because it includes all of the counties in the Electorate and the counties in the Grand Duchy of Hesse that featured the guild system.[74] Instead of a dummy for the *Gewerbefreiheit*, I include a dummy that takes the value of one if the county was located in the Grand Duchy of Hesse and zero otherwise. Although both states differed in size and the institutional differences abstracted from the freedom of enterprise were similar to those between the other states - importantly the Electorate and Prussia - the estimate of the placebo dummy is insignificant and very close to zero. This is another indicator that the large treatment effect found above indeed captures developmental advantages owing to having the *Gewerbefreiheit* after the *Zollverein* was founded.

Columns 3 and 4 show the respective estimates when comparing all the counties in the Duchy of Hesse that had the guild system to either the adjacent Prussian districts, excluding the coal-producing Ruhr area,[75] (Column 3) or Nassau (Column 4). Both estimates are highly significant and comparable in size to the effects found using the Electorate as the control unit. Therefore, no distinctive features of the Electorate of Hesse seem to be responsible for the large effects presented thus far.

The last two columns contain estimates from a regression including all four states, Nassau and Prussia with the *Gewerbefreiheit* and the Electorate and all of the counties in the Grand Duchy that featured the guild system. For Column 5, I include all counties in the three adjacent Prussian provinces. The estimate of the dummy for *Gewerbefreiheit* is large and highly significant, so the overall picture con-

73. Again, I correct standard errors for spatial autocorrelation using the procedure suggested by Conley (1999) and a cut-off value of 50 km. As in my main specification, regressions are not weighted by population. Results for regressions weighted by population in 1837 are virtually the same.

74. All regressions containing the Grand Duchy of Hesse exclude all counties in the Province Rheinhessen, as these had the *Gewerbefreiheit*, and the county of Offenbach, which acquired the *Gewerbefreiheit* as a privilege (see Appendix A4).

75. I only use the adjacent districts and exclude the Ruhr area, which developed rapidly owing to large coal fields, to ensure that the counties are comparable aside from the institutional setting. Including the Ruhr area or using the whole adjacent Prussian provinces, the effect becomes much stronger (24.30***(6.77) or 24.39***(6.85)).

Table 3.7. Additional states: Difference in log population in 1837 and 1864 in log points

	(1)	(2)	(3)	(4)	(5)	(6)
	Elec-Nas	Elec-GrDu	GrDu-Pru	GrDu-Nas	Pooled all	Pooled distr.
Gewerbefreiheit	14.67***		16.64***	15.32***	20.60***	15.07***
	(1.171)		(3.825)	(2.807)	(3.341)	(1.816)
Pop. density 1837	0.281***	0.199***	0.175***	0.257	0.00323*	0.191***
	(0.0299)	(0.0248)	(0.0411)	(.)	(0.00132)	(0.0432)
Grand Duchy		0.0363				
		(2.945)				
Constant	-19.73***	-13.40***	-11.30	-18.33***	2.532	-12.68***
	(2.725)	(2.485)	(5.772)	(4.170)	(2.490)	(3.810)
Observations	46	37	38	45	199	103
R^2	0.649	0.452	0.541	0.544	0.258	0.535

Notes: This table shows OLS estimates at the county level. The dependent variable is log population in 1864 minus log population in 1837 in log points. Column 1 includes all of the counties in the Electorate of Hesse and Nassau. Column 2 includes all of the counties in the Electorate and the Grand Duchy of Hesse, except for Rheinhessen and Offenbach. Grand Duchy is a dummy variable that takes the value of 1 if a county is located in the Grand Duchy of Hesse. Column 3 includes the same counties in the Grand Duchy and additionally all of the counties in adjacent Prussian districts excluding the Ruhr area; Column 4 instead additionally includes all counties in Nassau. Columns 5 and 6 pool counties in all four states. Regarding Prussia, Column 5 includes all of the counties in all three adjacent Prussian provinces, Westphalia, Saxony and the Rhine Province, whereas Column 6 only includes the counties in the adjacent Prussian districts excluding the Ruhr area. Standard errors (in parenthesis) are adjusted for spatial autocorrelation following Conley (1999) using a distance cut-off of 50 km. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

firms the results found thus far. However, the R^2 is considerably smaller than those for the other regressions. This reflects the fact that, by including all of the counties of the three adjacent Prussian provinces, counties are being included that are located far away from the Hessian states and that are much more heterogeneous in other dimensions influencing economic growth.[76] To obtain a more homogeneous sample, the sixth specification only includes Prussian counties that are located in districts adjacent to at least one of the Hessian states and, again, excludes the Ruhr area. The treatment effect in this specification is, again, comparable in size those found before and the R^2 recovers.

To sum up, this section demonstrates that the treatment effects found above are not unique to the comparison between Prussia and the Electorate of Hesse. In

76. Part of the low R^2 is driven by the existence of city-counties in Prussia, which have considerably small areas, resulting in a substantially high density. This can be seen in the regression as the estimate for the density in 1837 decreases drastically. Excluding the five affected cities drives the estimate for the density up again but importantly reduces the effect of the *Gewerbefreiheit* only slightly (19.51***(2.55)). The R^2 recovers to 0.373. The following specification circumvents this problem, as none of the affected cities is located in an adjacent district.

fact, including the Duchy of Nassau and the Grand Duchy of Hesse as alternative treatment and control units leads to similar estimates. The difference in institutions between the guild system and the *Gewerbefreiheit* appears to have caused significantly different levels of economic growth throughout all of the considered states, after an increase in aggregate demand induced by the foundation of the German *Zollverein* created the necessary conditions.

3.7 Conclusion

This paper examines the interaction between institutions and market size as drivers of economic development during the German industrial take-off. I suggest a mechanism by which an increase in aggregate domestic demand combined with a decrease in fixed costs to establish IRS technology led to industrialisation and economic progress. This mechanism is tested using three unique features of German history as a quasi-natural experiment. First, the abolition of guilds and the introduction of the *Gewerbefreiheit* in some German states arguably decreased the fixed costs required to industrialise manufacturing production. Second, the foundation of the German *Zollverein* at once increased aggregate demand considerably in the member states. Third, the new German borders defined at the Congress of Vienna and the division of the Kingdom of Westphalia enabled me to separate these effects from cultural differences and other potential drivers of economic development.

In a difference-in-differences framework, I empirically demonstrate that Prussian counties that had the *Gewerbefreiheit* were able to realise large and highly significant developmental advantages compared to counties in the Electorate of Hesse but only *after* the foundation of the *Zollverein*. This developmental advantages were accompanied by a larger share of the labour force working in manufacturing and less self-employed workers within manufacturing after the treatment period. Both findings conform to a switch to IRS technologies within manufacturing sectors in Prussian counties. To validate that the relationship is causal, I perform two additional empirical exercises. First, I use a spatial regression discontinuity design around one newly defined, exogenous border. These results show that state affiliation is indeed responsible for the large difference in economic development uncovered before. Second, I demonstrate that the treatment effect is stable when including two additional Hessian states. This suggests that my results are indeed driven by the institutional differences between the guild system and the *Gewerbefreiheit* and valid beyond the two states initially investigated.

My analysis reveals that a modification in institutions adjusted to a changing economic structure can indeed foster economic development. This finding has two major implications. First, it is relevant for understanding the Industrial Revolution as it proves that institutional innovation interacting with a significant structural change was able to catalyse industrialisation in a country that subsequently became an industrial leader. This provides strong evidence for the hypothesis that institutional transformation constitutes a key aspect of understanding the Industrial Revolution. Moreover, my analysis demonstrates that it is indispensable to determine the institutions that are suitable for supporting the development of a particular economy. Relying solely on standard measures of institutional quality disregards the interaction of institutions with cultural and economic conditions, which might yield misleading results. Future research should focus more closely on these kinds of interactions.

Appendix 3.A Historical Appendix

3.A.1 The history of the Kingdom of Westphalia

'In the beginning was Napoleon.'[77] Thomas Nipperdey, a famous German historian, starts his exhaustive trilogy on German history from 1800 to 1918 with this sentence. It reflects the origin of modern German history. Prior to Napoleon's arrival in Germany, the Holy Roman Empire existed in that place. It was an old territorial unit fragmented into a large number of independent states. Institutions and legal order in the Holy Roman Empire did not differ noticeably between the singular states: they were all run in an autocratic fashion[78] and shaped by legal and social inequality.[79] This fact ensures a common institutional history of all parts of the later Kingdom of Westphalia and the German Confederation more generally, which is important to guarantee the same cultural relation to new institutions.

When Napoleon won the War of the Fourth Coalition, conquered Germany and dictated the Treaties of Tilsit in 1807, the old order was overturned, and the Holy Roman Empire ceased to exist. Prussia lost half its territory; the Electorate of Hesse, as well as many other, smaller states, was wiped off the map. In their place, Napoleon founded the Kingdom of Westphalia as a vassal state of the French Empire and as a model state to demonstrate the advantages and superiority of the French institutional and social system. It was governed by Napoleon's brother, who implemented far-reaching economic and social reforms on Napoleon's demand. The Kingdom of Westphalia was the first German state to get a constitution - the French institutions, as well as the Code Civil, were implemented in a 'revolution from above'. This resulted in a total reorganisation of administration, law and social order. The drastic reforms also were meant to unify the different territories on which the Kingdom was built and therefore included a new administrative subdivision, which was determined by practical reasons and broke with historical borders. The unification also implied the creation of a common single market including the abolishment of internal tariff borders and the implementation of uniform regulations. Conversely, the Kingdom was still a purely agrarian country without notable industrial production. Napoleon demanded high financial war contributions, strong military support for his numerous wars and the cession of profitable public demesnes. These and the trade policy under Napoleon's Continental Blockade led to financial distress, heavy

77. *'Am Anfang war Napoleon.'*, Nipperdey (1983) p. 11.

78. See Acemoglu, Johnson, and Robinson (2002b) Appendix, p. 59. Tabellini (2010) and Acemoglu, Johnson, and Robinson (2002b) define constraints on the executive as their institutional measure to be 1 ('no regular limitations on the executive's actions') throughout all German states until 1800. The only exemptions were some of the so-called free cities and Württemberg, which were all far from the Kingdom of Westphalia and the later Hessian states.

79. See Acemoglu, Cantoni, et al. (2011) p. 3288

taxes, economic problems and mass poverty inside the Kingdom of Westphalia.[80] The radical reforms, which aimed at overthrowing the old social order and proclaimed a substantially different institutional setting, in combined effect with the consequences of nationalisation, heavy taxes, economic distress and the war in general, drastically shaped the new state. Despite the short time period of the Kingdom's existence, it therefore can be assumed that it turned into a homogeneous territorial unit. This ensured that, even after Napoleon had been defeated, the Kingdom had been occupied by the allies in 1813 and finally divided at the Congress of Vienna; the separate parts were still comparable in a historical and cultural dimension.

3.A.2 The spread of the *Gewerbefreiheit*

Around the time of the industrial take-off in Germany, a fundamental institutional change took place away from the guild system to an institutional setting called *Gewerbefreiheit* (freedom of enterprise). In its most advanced form, the new institutional setting not only led to the total abolishment of all guilds but also to unconditional freedom of enterprise. This included freedom of establishment such that everyone was allowed to found any given enterprise everywhere in the respective state.[81] Although the term *Gewerbefreiheit* originally did not only relate to crafts but also to all kinds of enterprises, the connection to the abolishment of guilds means that the focus in the historical literature lies on craftsmen. For this reason and as it allows for a better comparison to the guild system, I will focus in the following on the regulations concerning crafts.

The transformation from the guild system to *Gewerbefreiheit* in German states was triggered by an external shock. During the French Revolution, France abolished all guilds and, in 1791, introduced the *Gewerbefreiheit*. This institutional change was brought to Germany by Napoleon, who first introduced the *Gewerbefreiheit* in the regions left of the river Rhine annexed by France in 1795/97. During the reforms in the Kingdom of Westphalia dictated by Napoleon, the *Gewerbefreiheit* was implemented there between 1808 and 1810.[82]

After Prussia was defeated by Napoleon in 1806, it lost half its territory and was forced to make massive tribute payments. In the aftermath, the Prussian government realised the underdevelopment of the Prussian monarchy and triggered economic and social reforms to be able to raise the payments and to regain their status as a great power. These were influenced by the French Revolution and became known as

80. For the history of the Kingdom of Westphalia see Berding: 'Das Königreich Westphalen als napoleonischer Modell- und Satellitenstaat (1807-1813)' in Gerd Dethlefs (2008) pp. 15-29.

81. In most cases, this still required to obtain a trade certificate; however, in states with *Gewerbefreiheit* this was nothing more than a formality.

82. The full installation of the *Gewerbefreiheit* happened in three distinct laws in 1808, 1809 and 1810. See Bovensiepen (1909) p. 11 and Kiesewetter (2004) p. 57. It was also introduced in other states controlled by Napoleon as, for example, the Duchy of Berg.

the Stein-Hardenberg Reforms. In the course of these reforms, Prussia introduced the *Gewerbefreiheit* in 1810/11.

However, after the war against Napoleon was finally won in 1815 and the reconquered German territories allocated among the allies, the Restauration began. During this time period, known as the Congress System, most changes made by Napoleon, rooted in the French Revolution, were withdrawn. In the case of the *Gewerbefreiheit* this was carried out in most of the German states but in different ways and intensities. Although the reformers around Hardenberg had lost some of their power in Prussia, they were still able to withstand the pressure of the opponents of the *Gewerbefreiheit*. This and the king's promise to protect the former laws of the newly annexed territories led to a heterogeneous situation throughout the country. In the parts of Prussia that already had the *Gewerbefreiheit*, like the territories belonging to the former Kingdom of Westphalia, it persisted. In other regions, like the Prussian parts of former Saxony, the guilds were able to keep their power.[83] Although new laws in 1820 and 1831 removed some of the few local barriers that still existed owing to the leftovers of the guild system, only as late as 1845 was a common law, which established freedom of enterprise in all of Prussia, introduced. This law constituted a compromise by defending the central points of the *Gewerbefreiheit* but limiting it slightly by requiring a certificate of competence for some occupations[84] and allowing the formation of free guilds without limiting market access to its members.[85] Still, Prussia was the country with the least restrictive industrial code among the larger German states.

In the other German states, the *Gewerbefreiheit* was partly or totally withdrawn at some point after the Congress of Vienna or had never before been implemented. Among the most restrictive states was the Electorate of Hesse. William I., Elector of Hesse, was a restorative ruler and tried to extinguish the institutional changes introduced during the French occupation. On that account, he re-established the guild system as it was before the French invasion with new guild laws in March 1816. Despite various attempts to reform the guild system, these laws persisted with only minor changes until 1867, after Prussia annexed the Electorate of Hesse.[86]

The comparison of Prussia to the Electorate of Hesse does therefore constitute a perfect setting to examine the effects of the new freedom of enterprise compared

83. See Vogel (1983) p. 41.

84. These certificates of competence were justified by safety requirements and granted after state examination. They were required i.a. for doctors and pharmacists, master builders, sea captains, bricklayers and booksellers. The occupations that required such a certificate are specified in the 'Allgemeine Gewerbeordnung' from 1845, §42-58; see Richter (2007) p. 46.

85. See Steindl (1986) pp. 3562-3565 and Richter (2007) pp. 46-48; in 1849, the *Gewerbefreiheit* in Prussia was de jure limited even more, but as the law was not enforced by the state authorities, the new limitations were not relevant.

86. See Bovensiepen (1909) pp. 11-16 and 177.

to the old guild system because both states stuck to the chosen institution in the strictest way present in the German Confederation.

3.A.3 Institutional comparison between the Electorate and Prussia

In this section, I will analyse other relevant institutional features implemented after the Congress of Vienna and demonstrate that they were remarkably similar when comparing Prussia and the Electorate of Hesse. At first, it is important to mention that, with the introduction of the German *Zollverein* at latest, all institutional regulations concerning factor and commodity markets were the same in all member states.[87] Furthermore, the historical and economic literature identifies considerable institutional changes around the onset of the German industrial take-off, which might have had a significant impact on economic development.

In an extensive chapter about the industrialisation in Germany, Thomas Nipperdey discusses the role that the state played for supporting industrial development. He explicates that the different states did not create or initialise the Industrial Revolution, as, for example, direct public investments were an exception. The states only created some of the framework conditions. Next to the introduction of the *Gewerbefreiheit*, Nipperdey mentions the agrarian reforms, which liberated peasants, and the implementation of a modern legal order such as the Code Civil.[88] These institutional dimensions directly match up with the institutional changes Acemoglu, Cantoni, et al. (2011) or rather Kopsidis and Bromley (2015) focus on to estimate institutional influences of French occupation on German economic development.[89] Additionally, Nipperdey indicates that, during the time of interest, liberal reformers understood political participation as a motivation for progressiveness.[90] This idea is prominent in economic literature in which constraints on the executive are regularly used as a proxy for institutional quality before and around the Industrial Revolution.[91] Lastly, Nipperdey states that education took centre stage in the governments' attempts to support economic development. Although, according to his assessment, educational advantages only had a real impact on industrialisation starting in the 1850s/60s, educational differences might have been of importance for economic performance.[92] This idea is emphasised by Becker and Woessmann (2009). Following these references, I will in turn compare four institutional dimensions between

87. See Kopsidis and Bromley (2015) p. 169.

88. See Nipperdey (1983) pp. 181-183.

89. Acemoglu, Cantoni, et al. (2011) consider four institutional changes: the introduction of the civil code, the abolition of serfdom, agrarian reforms and the abolition of guilds. Kopsidis and Bromley (2015) explain that abolition of serfdom took place in all of Germany prior to Napoleon's arrival and therefore excluded it.

90. See Nipperdey (1983) p. 183.

91. Consider, for example, Acemoglu, Johnson, and Robinson (2005a) and Tabellini (2010).

92. See Nipperdey (1983) pp. 183-184.

Prussia and the Electorate of Hesse: the legal order, agrarian reforms, constraints on the executive and the educational system.

At first, legal order in the Prussian parts of the former Kingdom of Westphalia did not change considerably after Prussia took over control. The French reform acts continued to exist, at least partially, owing to the above-mentioned promise of the king to protect the former laws. Of course, the general Prussian law[93] was introduced in the new territories, but it implemented a similar legal order based on the French Revolution.[94] According to Kopsidis and Bromley (2015), it had comparable effects on economic development as the Code Civil.[95] Still, the Prussian monarchy was a classical state of the Restauration, oppressing liberal ideas and showing attempts to return to the old order.[96] So was the Electorate of Hesse. In particular, during the first years after its re-foundation, the Elector undid many of the former reforms conducted in the Kingdom of Westphalia. He did, however, keep some relevant parts; most importantly, he did not reintroduce the privileges of the nobility. Moreover, in 1821, administrative institutions in the Electorate were transformed back to the status of the former Kingdom, for example, by separating judiciary and administration.[97] Eventually, the new constitution of the Electorate of Hesse, implemented in 1831, fully restored a legal order comparable to the Code Civil.[98] In conclusion, both states' legal systems inherited essential elements from the Kingdom of Westphalia. Additionally, they had already returned entirely to the legal order of the Code Civil before the *Zollverein* was founded.

Second, the situation for peasants and premises for agricultural production were similar between the two states. Serfdom itself had already ceased to exist throughout Germany prior to Napoleon's arrival.[99] Agrarian reforms went along the same lines. Peasants obtained the opportunity to become independent and gain full property rights over the land they were cultivating; however, to achieve this, they had to pay a high initial transfer to their landlords. In the later stage of the reforms, the peasants were supported by mortgage banks, which were set up to help them finance this lump sum transfer. Even the timing of the reforms did not differ ex-

93. *'Allgemeines Landrecht für die Preußischen Staaten'*

94. For example, it also introduced the uniform rule of law due to the Prussian *Oktoberedikt* from 1807. In general, the Stein-Hardenberg reforms, which are considered considerably important for Prussia's development, mainly acquired inspiration from the reforms in the Kingdom of Westphalia; see Berding: 'Das Königreich Westphalen als napoleonischer Modell- und Satellitenstaat (1807-1813)' in Gerd Dethlefs (2008) pp. 19 and 29

95. See Kopsidis and Bromley (2015) page 169 and its online appendix p. 1.

96. See Nipperdey (1983) p. 334.

97. See Seier: 'Modernisierung und Integration in Kurhessen 1803-1866' in Heinemeyer (1986), pp. 448-451.

98. Kopsidis and Bromley (2015) date the introduction of a legal order comparable to the civil code in the Electorate to 1831, which is still before the start of my sample period, but propose a continuing existence in the Prussian parts after 1815.

99. See Kopsidis and Bromley (2015) p. 166.

cessively. They had started already in the Kingdom of Westphalia; however, in both countries, the reforms were not completed until the German revolution.[100]

Third, the executives have been largely unconstrained during the time of interest. Prussia and the Electorate were both governed by absolute rulers, and modern constitutions either did not exist or were ignored. Only in 1848 did the Prussian king fulfil his promise and install a constitution that possessed some modern features while still leaving vast power to the crown.[101] The Electorate, in contrast, already had a considerably more liberal constitution in 1831. Being common in these times, the monarch still held a stronger position than the parliament, but the constitution already placed significant constraints on him. In the following years, however, the Elector circumvented the constitution several times; in 1852, it was repealed and replaced by a considerably more conservative one.[102] This indicates that constraints on executives have been extremely limited in both states and that there was no significant difference between them.[103] If anything, the Electorate of Hesse had better institutions in the sense of a more constrained executive, which would work against finding an effect of Prussia's advance in freedom of enterprise.

Lastly, I focus on the development of a new educational system. In Prussia, the Stein-Hardenberg Reforms are known to have been accompanied by educational reforms initiated by Wilhelm von Humboldt. Humboldt aimed at a general education for all citizens, which was meant to exceed training targeting skills necessary for later work. He pursued the installation of a comprehensive network of 'Volksschulen', Prussian primary schools, to achieve the complete realisation of compulsory education.[104] The Prussian educational system was a prototype and model for all other German states in general as well as for the different types of schools.[105] Said differently, the Electorate of Hesse copied the Prussian educational system. Still, as Prussia did work as the role model, it is not entirely ensured that education in the Electorate achieved the same quantity or quality. For this reason, I follow Becker

100. See Nipperdey (1983) pp. 158-169; Acemoglu, Cantoni, et al. (2011) and Kopsidis and Bromley (2015) date the agrarian reform in the Electorate on 1832 and in the Province of Westphalia of Prussia on 1825 but in the Province of Saxony on 1808. However, these reforms had been long and unstable processes, so it is questionable to define an exact timing. The sample split at the time of the German revolution conducted in my empirical analysis (see Appendix C.1) reveals that the advancement of the agrarian reforms did not affect my estimates because, after the completion of the reforms, they are still similar.

101. See Nipperdey (1983) pp. 647-651.

102. See Seier, 'Modernisierung und Integration in Kurhessen 1803-1866' in Heinemeyer (1986), pp. 453-460.

103. Tabellini (2010) and Acemoglu, Johnson, and Robinson (2002b) give all of Hesse a value of 1 (on a scale from 1, unlimited authority, to 7, executive parity or subordination) in constraints on the executive in 1850 and the Polity IV database, which is used in both papers, rates constraints on the executive in Prussia as 1 or 2 between 1816 and 1858.

104. See Nipperdey (1983) pp. 56-65.

105. 'Volksschule' (primary schools), 'Gymnasium' (secondary schools) and the universities; see Nipperdey (1983) pp. 452, 454, 463 and 471.

and Woessmann (2009) and control, in my empirical analysis, for literacy rates as a proxy for education and particularly for the success of Prussia's famous primary schooling system. Thereby, I rule out that a possibly systematic difference in public education biases my results.

In summary, at least when controlling for education, the fundamental difference between the guild system in the Electorate of Hesse and the *Gewerbefreiheit* in Prussia can readily be assumed to be the only first-order institutional difference between the two states influencing economic development.

3.A.4 Institutions in the Grand Duchy of Hesse and the Duchy of Nassau

In this section, I will first discuss the laws and regulations concerning craft guilds and freedom of enterprise in the Grand Duchy of Hesse and the Duchy of Nassau. Additionally, I will briefly describe other relevant institutional differences to Prussia and the Electorate of Hesse.

In my analysis, I classify the Grand Duchy of Hesse as still having the guild system and the Duchy of Nassau as having the *Gewerbefreiheit*. Although the classification in both states is not as clear as it is in Prussia or the Electorate of Hesse, they still had a strong tendency in the corresponding direction. Compared to other states of the German Confederation, they stuck to the chosen institution relatively clearly. In the following, I will, in turn, discuss the laws and implementation in the Grand Duchy and Nassau.

In the Grand Duchy of Hesse, a leading decision about the abolishment of guilds and the introduction of freedom of enterprise was not made until 1866, so the guild system was not fundamentally liberalised beforehand. However, some restrictions had been removed up to this point. Most importantly, this included the constraints on masters to only work in the territory wherein their guild was active, which were dropped in 1821. Still, freedom of establishment for craftsmen was strongly limited, and guilds continued to exist for many crafts. Moreover, the situation was heterogeneous throughout the country. In the Province Rheinhessen, which belonged to France before 1814, the Code Civil was introduced under Napoleon and never repealed afterwards such that the freedom of enterprise existed. In the city of Offenbach, the *Gewerbefreiheit* was introduced alongside other liberal reforms in 1819 to support economic development and to privilege the city. I exclude all counties in both regions from my analysis. Apart from these two regions, the guild system continued to exist in the Grand Duchy of Hesse until 1866.[106] In total, this implies that craftsmen were slightly less restricted in the Duchy of Hesse than in the Electorate of

106. See Hahn, 'Der hessische Wirtschaftsraum im 19. Jahrhundert' in Heinemeyer (1986) pp. 394-395; and Dipper, 'Gewerbefreiheit in Hessen' in Dipper (1996) pp. 251-261 as well as Mascher (1866) pp. 645-646.

Hesse, but both states did not implement the *Gewerbefreiheit* as a new institutional setting. Consequently, I classify the Grand Duchy as a state having the guild system.

By contrast, the government of the Duchy of Nassau introduced the *Gewerbefreiheit* in 1819 by passing a law to abolish all guilds.[107] Thirty years later, however, during the German revolution and owing to massive turmoil of craftsmen throughout all of Germany, freedom of enterprise was restricted again. To ensure quality and limit market access, in 1849, the government introduced a system in which craftsmen again needed to obtain a 'master' certificate before conducting a craft. The examination to obtain such a title was similar compared to the old guild system but conducted under governmental supervision. Other masters of the same craft and from the same district had to test the candidate; if he passed the examination, the new master was still restricted to conducting his craft in the municipality he lived in. If he instead wanted to found a crafts enterprise in another municipality, he had to get permission of the corresponding local government. This implies that the new law was stricter than the guild system in the sense that everybody, independent of the specific craft, had to obtain a master certificate before conducting a craft, compared to a former system with many exemptions.[108] Still, as soon as a craftsman had become a master, he was not restricted on how to conduct his craft such that he was able to utilise productivity gains from new technologies and division of labour. In total, the new law did restrict entry to the crafts but did not hinder innovations and entrepreneurial activities of practising craftsmen. Moreover, the new law was repealed already in 1860, and full freedom of enterprise was reintroduced. The stricter law was, after all, only in place for 11 years after 30 years of unconditional freedom of enterprise. For this reason, and as the law of 1849 had not been an attempt to re-establish the guild system, I consider the Duchy of Nassau to have the *Gewerbefreiheit* throughout the sample period.

The institutional setting in the Grand Duchy of Hesse was similar compared to the other considered states. It was a constitutional monarchy with a strong and inviolable position of the grand duke such that constraints on the executive were low. Its constitution of 1820 marks the introduction of a legal order similar to the Code Civil.[109] Agrarian reforms went along the same lines as in the other states, although they were conducted comparatively early.[110]

107. Even though the examination for the master and journeyman craftsman's certificate kept existing, this leftover from the old system did not play any role for further development, as the permission to practice a craft did not depend on getting such a certificate.

108. See Heinemeyer (1986) p. 393; and Dipper, 'Gewerbefreiheit in Hessen' in Dipper (1996) p. 256 as well as Mascher (1866) pp. 605-607 and Spielmann (1926) pp. 143-144.

109. See Kopsidis and Bromley (2015) p. 165 and its online appendix p. 2.

110. Kopsidis and Bromley (2015) date them to 1816, earlier than for the Province of Westphalia (1825) and the Electorate of Hesse (1832). This was still later than in the Rhine Province of Prussia (1804), which next to the Province of Westphalia was the adjacent Prussian province to the Grand Duchy. According to Nipperdey, they have not been exceptional; see Nipperdey (1983) pp. 167-170.

The regime in Nassau was a constitutional monarchy and in terms of institutions similar to the ones in the other states. In contrast to the others, however, Nassau was one of the latest states in the German Confederation to have started on agrarian reforms. They began as recently as 1840, but then, owing to the turmoil around the German revolution, they were carried out faster, more radically and with more favourable terms for the peasants.[111] Still, the late start might have led to a disadvantage in development.

Appendix 3.B Data Appendix

For my analysis, I mainly rely on census data. For the period from 1837 to 1871, these censuses were conducted following common rules of the German *Zollverein*; therefore, the resulting data is of good quality and comparable even between different states. The data for the years after 1871 are only taken from Prussian censuses; thus, the data collection procedures were the same. One drawback is that I can only use this data source for states annexed by Prussia and Prussia itself, which is the reason I do not use the control variables when including the Duchy of Nassau and the Grand Duchy of Hesse. Data sources were the following.

For the counties belonging to Prussia, I use population data for the years 1821, 1849, 1864 and 1877 from the ifo Prussian Economic History Database (iPEHD; Becker, Cinnirella, Hornung, and Woessmann (2014)). The database also contains surface area data for these counties. I use total county area from the 1858 census, as this is the only year with good data available during the sample period. To ensure constant observational units, I merge the data as recommended by the authors of iPEHD using their variable *kreiskey1800* for merging. I combine these datasets with data from the census in 1837 from historical records.[112] County borders did not change between 1837 and 1849, so I directly merge this data to the data of 1849 and then follow the merging procedure described above.

Population as well as area data for counties in the Electorate of Hesse and the Grand Duchy of Hesse are taken from records of the statistical office of Hesse.[113] As county borders changed several times in these two states during the sample period, I reconstruct counties in the borders of 1852 using the contained municipality-level data and treat these borders as constant for the entire sample period.[114] This makes

111. See Nipperdey (1983) p. 169.

112. From 'Die Bevölkerung des preussischen Staats: nach dem Ergebnisse der zu Ende des Jahres 1837 amtlich aufgenommenen Nachrichten in staatswirthschaftlicher, gewerblicher und sittlicher Beziehung' (1839).

113. From 'Historisches Gemeindeverzeichnis für Hessen Heft 1: Bevölkerung der Gemeinden 1834 bis 1967' (1968).

114. I use the records of the statistical office about changes in counties from 'Historisches Gemeindeverzeichnis für Hessen Heft 2: Gebietsänderungen der hessischen Gemeinden und Kreise 1834 bis 1967' (1968) to reconstruct the original counties containing the same municipalities as in 1852.

it possible to observe population changes in constant geographical units. I interpolate the years 1837 and 1849 as geometric means of the census data from the censuses three years earlier and three years later, respectively, because data concerning exactly these years do not exist for the necessary grid. For both states, I exclude all small exclaves that are not connected to the main territory, for example, the county of Vöhl for the Duchy of Hesse, which is surrounded by the Principality of Waldeck and Pyrmont and the Electorate of Hesse as well as the county of Schmalkalden belonging to the Electorate of Hesse, which was located inside the Thuringian states. Pre-treatment county-level data in 1821 for the Electorate of Hesse are taken from historical records.[115] I use municipality-level data from this source and another historical record[116] to reconstruct counties in borders of 1852 as well.

In the Duchy of Nassau, county borders did not change during the sample period.[117] This fact allows me to rely directly on historical records.[118]

The county-level data on Protestantism, literacy and workforce share in mining, which I use as control variables, are also taken from iPEHD. One restriction is that such data on the necessary grid are available only for Prussia and only for later years. I have therefore had to rely on data from years after Prussia annexed the Electorate of Hesse. The fact that the respective censuses took place after the sample period does, however, most probably not affect the results because either these variables can be assumed to have been approximately constant during as well as after the sample period or any resulting bias would work against me. Still, as the Grand Duchy of Hesse was not annexed by Prussia, I can only use these control variables for my main analysis comparing the Electorate of Hesse to Prussia.

The data for income taxes per capita in 1877-78 on county level is taken from iPEHD as well. I use the way described in Becker and Woessmann (2009) Appendix I B to construct the necessary variable. For details about the data sources and how the Prussian tax system worked, please refer to that paper. Please note that data for the Prussian city-counties are not available in the data containing the income taxes and population in 1877, so city-counties need to be excluded from all analyses working with this data. In my baseline sample, this includes two cities in Prussia, which belonged to Westphalia beforehand, Halle and Magdeburg, and the city of Kassel in the former Electorate of Hesse.

To construct the share of the workforce working in manufacturing and the share of self-employed workers within manufacturing on county level, I again rely on iPEHD. The data was taken from the occupational census in Prussia in 1882, which

115. From 'Sammlung von Gesetzen, Verordnungen, Ausschreiben und anderen allgemeinen Verfügungen für Kurhessen' (1822).

116. 'Kurhessisches Staats- und Address-Handbuch' (1823)

117. This excludes an administrative reform in 1849 during the German revolution that was repealed in 1854 and did not influence data collection during the censuses.

118. I use data for the years 1837 and 1864 from the 'Staats- und Adreß-Handbuch des Herzogthums Nassau' of 1839 and 1865. Area data are taken from the publication of 1865.

collected information on employment and self-employment across two-digit sectors. I again follow Becker and Woessmann (2009) to construct the share of the total labour force working in the manufacturing sector, who in turn followed the classification by the Prussian Statistical Office to classify the sectors.[119] In contrast to the authors, I do, however, exclude the mining industry, which in modern sector classifications is typically not included as a manufacturing sector. All results are robust to including the mining industry as a manufacturing sector. To calculate the share of self-employed workers within manufacturing, I sum the number of self-employed workers over all manufacturing sectors within each county and then divide by the total number of workers within all manufacturing sectors in the same county.

I match the data for the control variables as well as the population data in 1877, income taxes per capita in 1877-1878 and the worker shares in 1882 for the Electorate of Hesse with the earlier data for the Electorate using county names. As county borders and names did not change after Prussia annexed the Electorate, the matching is straightforward. The only thing that did change was that Kassel became a city county and was divided from its hinterlands, which became a separate county ('Landkreis Kassel') in 1866. Therefore, by constructing the control variables for Kassel, I matched the city county Kassel with its hinterlands again to reconstruct the old county of the Electorate. For the total population in 1877 and income taxes in 1877-1878, the data for city-counties are not available. Thus, in cases where I use this data, I exclude the county of Kassel.

The fortification dummy, which I use as an additional control variable, is constructed using the 'List of fortifications in Germany' from the German version of Wikipedia.[120] This list provides all fortifications in Germany with the date of construction and demolition. As I am only interested in lager fortifications, which had an impact on the number of inhabitants or infrastructure, this list can be assumed to be complete. In ambiguous cases, I conducted further research to find out if a fortification still existed and was in use during the sample period.

The municipality-level data used in section 3.5.3 are taken from publications of the respective statistical offices. For the Electorate of Hesse, I use the same source as for the county-level data described above; for the municipalities in Prussia, I take the data from a publication of the statistical office of North Rhine-Westphalia.[121] As the data for Prussian municipalities do not contain the year 1864, I only use the years 1837 and 1849 and focus on a shorter sample period in the respective section.[122]

119. See Becker and Woessmann (2009) Appendix I D.

120. 'Liste der Festungen in Deutschland', https://de.wikipedia.org/wiki/Liste_der_Festungen_in_Deutschland (last accessed 16 January 2019).

121. From 'Gemeindestatistik des Landes Nordrhein-Westfalen: Bevölkerungsentwicklung 1816-1871' (1966).

122. Both data sources also contain the year 1858; however, for comparability with the county-level data, I use the year 1949 as the outcome year. All results also hold for using 1858 as the end of the sample period.

Data for municipalities that currently belong to the federal state Saxony do not exist from the respective statistical office.

I geocode the data to be able to use Conley adjusted standard errors. I use the Stata package *opencagegeo* that relies on *OpenCage Geocoder API*. I rely on the county name for geocoding, which in most cases coincides with the name of the administrative centre of the county. Only in unclear cases, when either the name of the county changed over time or does not point to a specific city, I instead use the name of the administrative centre of the former county. To geocode municipality data, I use the name of the municipality and additionally add the county name. I correct obvious mistakes using Google Maps.

Appendix 3.C Robustness and additional empirical evidence

3.C.1 Full difference-in-difference results with different distance cut-offs

Table 3.C.1. Difference-in-difference regression: Log population between 1821 and 1864

	(1)	(2)	(3)	(4)	(5)
Gewerbefreiheit * 1837	-1.815	-1.392	-0.215	-0.735	-4.348
	(3.223)	(2.957)	(2.750)	(2.916)	(3.616)
Gewerbefreiheit * 1849	6.302**	5.782**	6.739***	5.794**	2.309
	(3.070)	(2.727)	(2.575)	(2.649)	(3.570)
Gewerbefreiheit * 1864	21.17***	19.24***	19.57***	18.36***	13.14***
	(4.293)	(3.661)	(2.469)	(2.673)	(3.801)
Population density * 1837		-0.00788	-0.00858	-0.0126**	0.000472
		(0.00614)	(0.00603)	(0.00513)	(0.000401)
Population density * 1849		0.00968	0.00592	-0.00146	0.00230***
		(0.00841)	(0.00696)	(0.00377)	(0.000388)
Population density * 1864		0.0360***	0.0264***	0.0169***	0.00348***
		(0.00998)	(0.00817)	(0.00362)	(0.000416)
Share Protestants * 1837			0.0440	0.0390	0.0440
			(0.0417)	(0.0431)	(0.0310)
Share Protestants * 1849			0.0512	0.0421	0.0990***
			(0.0369)	(0.0378)	(0.0324)
Share Protestants * 1864			0.119**	0.108**	0.160***
			(0.0509)	(0.0517)	(0.0311)
Literacy rate * 1837			-0.203	-0.233	-0.256
			(0.831)	(0.810)	(0.548)
Literacy rate * 1849			0.470	0.415	-0.0999
			(0.703)	(0.666)	(0.542)
Literacy rate * 1864			1.745**	1.674**	0.371
			(0.718)	(0.688)	(0.703)
Share mining * 1837			-0.293	-0.218	0.576
			(0.431)	(0.435)	(0.632)
Share mining * 1849			-0.540	-0.404	0.925
			(0.454)	(0.459)	(0.715)
Share mining * 1864			0.133	0.308	2.932***
			(0.594)	(0.615)	(1.081)
Fortification * 1837				9.731*	6.897***
				(5.195)	(2.535)
Fortification * 1849				17.67***	17.58***
				(5.543)	(3.405)
Fortification * 1864				22.70***	28.34***
				(5.001)	(5.297)
Constant	3.41e-15	3.45e-15	3.47e-15	3.51e-15	-4.49e-15
	(0.636)	(0.537)	(0.326)	(0.305)	(0.387)
Observations	180	180	180	180	380
R^2	0.918	0.925	0.931	0.934	0.947
County & time FEs	Yes	Yes	Yes	Yes	Yes

Notes: The table shows difference-in-difference estimates at the county level. The dependent variable is population in log points. Only counties belonging to the former Kingdom of Westphalia are included, except for Column 5, which includes all of the counties in the Electorate of Hesse and the adjacent Prussian provinces. All regressions include county and time fixed effects. Standard errors (in parenthesis) are Conley-adjusted with a distance cut-off of 50 km and corrected for heteroscedasticity and autocorrelation (HAC). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 3.C.2. Difference-in-difference regression: Log population between 1821 and 1864 (Distance cut-off 30 km)

	(1)	(2)	(3)	(4)	(5)
Gewerbefreiheit * 1837	-1.815	-1.392	-0.215	-0.735	-4.348
	(2.159)	(1.898)	(2.141)	(2.290)	(3.200)
Gewerbefreiheit * 1849	6.302***	5.782***	6.739***	5.794***	2.309
	(1.861)	(1.548)	(1.893)	(1.966)	(3.164)
Gewerbefreiheit * 1864	21.17***	19.24***	19.57***	18.36***	13.14***
	(2.925)	(2.255)	(2.026)	(2.150)	(3.418)
Population density * 1837		-0.00788	-0.00858	-0.0126**	0.000472
		(0.00843)	(0.00855)	(0.00515)	(0.000444)
Population density * 1849		0.00968	0.00592	-0.00146	0.00230***
		(0.0105)	(0.00947)	(0.00338)	(0.000402)
Population density * 1864		0.0360***	0.0264***	0.0169***	0.00348***
		(0.0119)	(0.0101)	(0.00252)	(0.000536)
Share Protestants * 1837			0.0440	0.0390	0.0440
			(0.0367)	(0.0381)	(0.0318)
Share Protestants * 1849			0.0512	0.0421	0.0990***
			(0.0328)	(0.0336)	(0.0331)
Share Protestants * 1864			0.119***	0.108**	0.160***
			(0.0421)	(0.0430)	(0.0354)
Literacy rate * 1837			-0.203	-0.233	-0.256
			(0.449)	(0.463)	(0.552)
Literacy rate * 1849			0.470	0.415	-0.0999
			(.)	(.)	(0.569)
Literacy rate * 1864			1.745***	1.674***	0.371
			(0.206)	(0.236)	(0.767)
Share mining * 1837			-0.293	-0.218	0.576
			(0.343)	(0.342)	(0.559)
Share mining * 1849			-0.540***	-0.404**	0.925
			(0.209)	(0.199)	(0.646)
Share mining * 1864			0.133	0.308	2.932***
			(0.252)	(0.267)	(0.998)
Fortification * 1837				9.731*	6.897**
				(5.244)	(2.987)
Fortification * 1849				17.67***	17.58***
				(5.580)	(3.736)
Fortification * 1864				22.70***	28.34***
				(5.295)	(6.004)
Constant	3.41e-15	3.45e-15	3.47e-15	3.51e-15	-4.49e-15
	(0.403)	(0.235)	(.)	(.)	(0.364)
Observations	180	180	180	180	380
R^2	0.918	0.925	0.931	0.934	0.947
County & time FEs	Yes	Yes	Yes	Yes	Yes

Notes: The table shows difference-in-difference estimates at the county level. The dependent variable is population in log points. Only counties belonging to the former Kingdom of Westphalia are included, except for Column 5, which includes all of the counties in the Electorate of Hesse and the adjacent Prussian provinces. All regressions include county and time fixed effects. Standard errors (in parenthesis) are Conley-adjusted with a distance cut-off of 30 km and corrected for heteroscedasticity and autocorrelation (HAC). $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 3.C.3. Difference-in-difference regression: Log population between 1821 and 1864 (Distance cut-off 100 km)

	(1)	(2)	(3)	(4)	(5)
Gewerbefreiheit * 1837	-1.815	-1.392	-0.215	-0.735	-4.348
	(3.114)	(2.872)	(2.049)	(2.275)	(2.957)
Gewerbefreiheit * 1849	6.302**	5.782**	6.739***	5.794***	2.309
	(2.524)	(2.386)	(1.637)	(1.775)	(2.988)
Gewerbefreiheit * 1864	21.17***	19.24***	19.57***	18.36***	13.14***
	(4.680)	(4.125)	(1.253)	(1.518)	(3.451)
Population density * 1837		-0.00788	-0.00858***	-0.0126***	0.000472
		(.)	(0.00142)	(0.00479)	(0.000410)
Population density * 1849		0.00968***	0.00592	-0.00146	0.00230***
		(0.00287)	(0.00413)	(0.00331)	(0.000372)
Population density * 1864		0.0360	0.0264	0.0169***	0.00348***
		(.)	(.)	(0.00410)	(0.000442)
Share Protestants * 1837			0.0440**	0.0390*	0.0440
			(0.0201)	(0.0210)	(0.0402)
Share Protestants * 1849			0.0512***	0.0421***	0.0990**
			(0.0164)	(0.0149)	(0.0411)
Share Protestants * 1864			0.119	0.108	0.160***
			(.)	(.)	(0.0460)
Literacy rate * 1837			-0.203	-0.233	-0.256
			(0.676)	(0.684)	(0.574)
Literacy rate * 1849			0.470	0.415	-0.0999
			(0.566)	(0.570)	(0.578)
Literacy rate * 1864			1.745***	1.674***	0.371
			(0.323)	(0.339)	(0.842)
Share mining * 1837			-0.293	-0.218	0.576
			(0.348)	(0.365)	(0.423)
Share mining * 1849			-0.540*	-0.404	0.925*
			(0.283)	(0.293)	(0.542)
Share mining * 1864			0.133	0.308	2.932***
			(0.240)	(0.281)	(0.817)
Fortification * 1837				9.731	6.897***
				(5.957)	(2.511)
Fortification * 1849				17.67***	17.58***
				(6.293)	(3.726)
Fortification * 1864				22.70***	28.34***
				(5.742)	(4.894)
Constant	3.41e-15	3.45e-15	3.47e-15	3.51e-15	-4.49e-15
	(0.785)	(0.651)	(.)	(0.145)	(0.174)
Observations	180	180	180	180	380
R^2	0.918	0.925	0.931	0.934	0.947
County & time FEs	Yes	Yes	Yes	Yes	Yes

Notes: The table shows difference-in-difference estimates at the county level. The dependent variable is population in log points. Only counties belonging to the former Kingdom of Westphalia are included, except for Column 5, which includes all of the counties in the Electorate of Hesse and the adjacent Prussian provinces. All regressions include county and time fixed effects. Standard errors (in parenthesis) are Conley-adjusted with a distance cut-off of 100 km and corrected for heteroscedasticity and autocorrelation (HAC). $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

3.C.2 Quantile regressions

Table 3.C.4. Quantile regression: Log population growth in counties formerly belonging to the Kingdom of Westphalia

	(1)	(2)	(3)	(4)
	OLS	q0.25	q0.5	q0.75
Gewerbefreiheit	19.20***	16.37***	20.45***	21.32***
	(2.890)	(3.392)	(2.797)	(3.861)
Pop. density 1837	0.0274***	0.0358	0.0258	0.0298
	(0.00328)	(0.214)	(0.0141)	(0.0266)
Share Protestants	0.0692	0.0845	0.0382	0.0757
	(0.0486)	(0.0470)	(0.0481)	(0.109)
Literacy rate	1.890*	0.897	2.074***	2.422
	(0.775)	(1.442)	(0.457)	(1.274)
Share working mining	0.529	0.138	2.012	1.106
	(0.939)	(1.930)	(1.487)	(9.257)
Fortification	10.94***	15.32	9.612***	3.592
	(2.423)	(12.23)	(2.666)	(.)
Constant	-183.8*	-95.29	-200.2***	-231.7
	(70.81)	(122.6)	(39.62)	(116.6)
Observations	45	45	45	45

Notes: The table shows quantile regressions at the county level. The dependent variable is the difference in log population between 1864 and 1837 in log points and the regression is estimated in spirit of equation 3.4. Only counties belonging to the former Kingdom of Westphalia are included. Column 1 shows the OLS results for comparison. Column 2 shows the results for the 0.25 quantile, Column 3 or the median and Column 4 for the 0.75 quantile. All standard errors are robust. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

References

Abel, Wilhelm. 1978. *Handwerksgeschichte in neuer Sicht (Göttinger Beiträge zur Wirtschafts- und Sozialgeschichte) (German Edition)*. Schwartz. URL: https://www.amazon.com/Handwerksgeschichte-ttinger-Beitra-Wirtschafts-Sozialgeschichte/dp/350901068X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=350901068X.

Acemoglu, Daron. 1, 2012. *Why Nations Fail*. Random House LCC US. 544 pages. URL: https://www.ebook.de/de/product/15562217/daron_acemoglu_why_nations_fail.html. [245]

Acemoglu, Daron, Davide Cantoni, Simon Johnson, and James A Robinson. 2011. "The Consequences of Radical Reform: The French Revolution." *American Economic Review* 101 (7): 3286–307. DOI: 10.1257/aer.101.7.3286. [246, 248, 260, 261, 265, 285, 288, 290]

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369–401. DOI: 10.1257/aer.91.5.1369. [246]

Acemoglu, Daron, Simon Johnson, and James Robinson. 2002a. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117 (4): 1231–94. DOI: 10.1162/003355302320935025. [261, 262]

Acemoglu, Daron, Simon Johnson, and James Robinson. 2002b. "The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth." *NBER Working Paper No. 9378*, (December): DOI: 10.3386/w9378. [285, 290]

Acemoglu, Daron, Simon Johnson, and James Robinson. 2005a. "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth." *American Economic Review* 95 (3): 546–79. DOI: 10.1257/0002828054201305. [245, 246, 248, 261, 288]

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005b. "Chapter 6 Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth*. Elsevier, 385–472. DOI: 10.1016/s1574-0684(05)01006-3. [248]

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review* 101 (5): 2003–41. DOI: 10.1257/aer.101.5.2003. [262]

Becker, Sascha O., Francesco Cinnirella, Erik Hornung, and Ludger Woessmann. 2014. "iPEHD - The ifo Prussian Economic History Database." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 47 (2): 57–66. DOI: 10.1080/01615440.2013.852370. [293]

Becker, Sascha O., and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *Quarterly Journal of Economics* 124 (2): 531–96. DOI: 10.1162/qjec.2009.124.2.531. [263, 267, 268, 271, 288, 290, 294, 295]

Bovensiepen, Rudolf. 1909. *Die Kurhessische Gewerbepolitik und die wirtschaftliche Lage des zünftigen Handwerks in Kurhessen von 1816-1867*. Vereinigte Friedrichs-Universität Halle-Wittenberg. [253, 254, 286, 287]

Cantoni, Davide. 2015. "The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands." *Journal of the European Economic Association* 13 (4): 561–98. DOI: 10.1111/jeea.12117. [261]

Clark, Gregory. 2014. "The Industrial Revolution." In *Handbook of Economic Growth*. Elsevier, 217–62. DOI: 10.1016/b978-0-444-53538-2.00005-8. [261]

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. 2019. "Inference with Arbitrary Clustering." *SSRN Electronic Journal*, DOI: 10.2139/ssrn.3449578. [264]

Conley, T. G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45. DOI: 10.1016/s0304-4076(98)00084-0. [264, 272, 276, 277, 281, 282]

De la Croix, David, Matthias Doepke, and Joel Mokyr. 2017. "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Preindustrial Economy." *Quarterly Journal of Economics* 133 (1): 1–70. DOI: 10.1093/qje/qjx026. [250, 254]

Dipper, Christof. 1996. *Hessen in der Geschichte: Festschrift für Eckhart G. Franz zum 65. Geburtstag (Arbeiten der Hessischen Historischen Kommission) (German Edition)*. Hessische Historische Kommission Darmstadt. URL: https://www.amazon.com/Hessen-Geschichte-Festschrift-Geburtstag-Historischen/dp/3884430327?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3884430327. [291, 292]

Epstein, Stephan R. 1, 2010. *Guilds, Innovation and the European Economy, 1400 1800*. Cambridge University Press. 364 pages. URL: https://www.ebook.de/de/product/7280664/stephan_r_epstein_guilds_innovation_and_the_european_economy_1400_1800.html. [250]

Fernihough, Alan, and Kevin Hjortshøj O'Rourke. 2014. "Coal and the European Industrial Revolution." Working paper. National Bureau of Economic Research. DOI: 10.3386/w19802. [261, 265]

Fuchs-Schündeln, N., and T.A. Hassan. 2016. "Natural Experiments in Macroeconomics." In *Handbook of Macroeconomics*. Elsevier, 923–1012. DOI: 10.1016/bs.hesmac.2016.03.008. [248]

Gerd Dethlefs Armin Owzar, Gisela Weiß. 2008. *Modell und Wirklichkeit: Politik, Kultur und Gesellschaft im Großherzogtum Berg und im Königreich Westphalen 1806-1813 (German Edition)*. Schoeningh Ferdinand GmbH. URL: https://www.amazon.com/Modell-und-Wirklichkeit/dp/3506757474?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3506757474. [286, 289]

Hahn, Hans-Werner. 1984. *Geschichte des Deutschen Zollvereins (Kleine Vandenhoeck-Reihe) (German Edition)*. Vandenhoeck & Ruprecht. URL: https://www.amazon.com/Geschichte-Deutschen-Zollvereins-Kleine-Vandenhoeck-Reihe/dp/3525335008?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3525335008. [255, 256]

Heinemeyer, Walter (Hrsg.): 1986. *Das Werden Hessens (Veröffentlichungen der Historischen Kommission für Hessen) (German Edition)*. N.G. Elwert. URL: https://www.amazon.com/Werden-Hessens-ffentlichungen-Historischen-Kommission/dp/3770808495?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3770808495. [256, 257, 280, 289–292]

Hessisches Statistisches Landesamt. 1968a. *Historisches Gemeindeverzeichnis für Hessen. H. 1. Die Bevölkerung der Gemeinden 1834 bis 1967*. Hessisches Statistisches Landesamt.

Hessisches Statistisches Landesamt. 1968b. *Historisches Gemeindeverzeichnis für Hessen. H. 2. Gebietsänderungen der hessischen Gemeinden und Kreise 1834 bis 1967*. Hessisches Statistisches Landesamt.

Hundt, Michael. 1996. *Die mindermächtigen deutschen Staaten auf dem Wiener Kongress (Veröffentlichungen des Instituts für Europäische Geschichte Mainz) (German Edition)*. P. Zabern. [258, 279]

Hundt, Michael, and Joachim Robert Moeschl. 2004a. "Der Deutsche Bund nach dem Frankfurter Territorialrezess um 1820." de. DOI: 10.25359/issn.1614-6352.map004.

Hundt, Michael, and Joachim Robert Moeschl. 2004b. "Die deutschen Staaten unter französischer Vorherrschaft Ende 1812." de. DOI: 10.25359/issn.1614-6352.map001.

Keller, Wolfgang, and Carol H. Shiue. 2014. "Endogenous Formation of Free Trade Agreements: Evidence from the Zollverein's Impact on Market Integration." *Journal of Economic History* 74 (04): 1168–204. DOI: 10.1017/s0022050714000898. [249, 256]

Keller, Wolfgang, and Carol H. Shiue. 2016. "Market Integration as a Mechanism of Growth." *CESifo Working Paper*, (6070): [250, 261]

Kiesewetter, Hubert. 2004. *Industrielle Revolution in Deutschland: Regionen als Wachstumsmotoren (German Edition).* Franz Steiner Verlag. URL: https://www.amazon.com/Industrielle-Revolution-Deutschland-Regionen-Wachstumsmotoren/dp/3515086137?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3515086137. [286]

Köllmann, Wolfgang. 1980. *Quellen zur Bevölkerungs-, Sozial- und Wirtschaftsstatistik Deutschlands, 1815-1875 (Forschungen zur deutschen Sozialgeschichte) (German Edition).* Boldt. URL: https://www.amazon.com/lkerungs-Wirtschaftsstatistik-Deutschlands-Forschungen-Sozialgeschichte/dp/3764617365?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3764617365. [262]

Kopsidis, Michael, and Daniel W. Bromley. 2015. "The French revolution and German industrialization: dubious models and doubtful causality." *Journal of Institutional Economics* 12 (01): 161–90. DOI: 10.1017/s1744137415000223. [260, 288–290, 292]

Kunz, Andreas, and Leonard Dietze. 2007. "German Historical GIS." HGIS Germany.

Lentz, Thierry. 2014. *1815*. Siedler. URL: https://www.amazon.com/1815/dp/3827500478?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3827500478. [258]

Mascher, Heinrich Anton. 1866. *Das deutsche Gewerbewesen von der frühesten Zeit bis auf die Gegenwart: nach Geschichte, Recht, Nationalökonomie und Statistik*. Döring. [291, 292]

Michalopoulos, Stelios, and Elias Papaioannou. 2013. "National Institutions and Subnational Development in Africa." *Quarterly Journal of Economics* 129 (1): 151–213. DOI: 10.1093/qje/qjt029. [247]

Murphy, Kevin M., Andrei Shleifer, and Robert Vishny. 1989. "Income Distribution, Market Size, and Industrialization." *Quarterly Journal of Economics* 104 (3): 537. DOI: 10.2307/2937810. [245, 252, 255]

Newey, Whitney K, and Kenneth D West. 1983. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–8. [264]

Nipperdey, Thomas. 1983. *Deutsche Geschichte 1800-1866: Burgerwelt und starker Staat (German Edition).* C.H. Beck. URL: https://www.amazon.com/Deutsche-Geschichte-1800-1866-Burgerwelt-starker/dp/340609354X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=340609354X. [256, 262, 269, 285, 288–290, 292, 293]

Nordrhein-Westfalen, Statistisches Landesamt. 1966. *Gemeindestatistik des Landes Nordrhein-Westfalen: Bevölkerungsentwicklung 1816-1871.* Beiträge zur Statistik des Landes Nordrhein-Westfalen: Sonderreihe Volkszählung 1961. URL: https://books.google.de/books?id=rWAoAQAAIAAJ.

North, Douglass C., and Robert P. Thomas. 1973. *The rise of the western world: A new economic history*. Cambridge University Press. [245]

North, Douglass C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History* 49 (4): 803–32. DOI: 10.1017/s0022050700009451. [245]

Ogilvie, Sheilagh. 2004. "Guilds, efficiency, and social capital: evidence from German proto-industry." *Economic History Review* 57 (2): 286–333. DOI: 10.1111/j.1468-0289.2004.00279.x. [250, 254]

Ogilvie, Sheilagh. 2014. “The Economics of Guilds.” *Journal of Economic Perspectives* 28 (4): 169–92. DOI: 10.1257/jep.28.4.169. [250, 253–255]

Ogilvie, Sheilagh, and A.W. Carus. 2014. “Institutions and Economic Growth in Historical Perspective.” In *Handbook of Economic Growth*. Elsevier, 403–513. DOI: 10.1016/b978-0-444-53538-2.00008-3. [248]

Ohnishi, Takeo. 1973. *Zollpolitik Preußens bis zur Gründung des deutschen Zollvereins*. Otto Schwartz & Co. [255, 256]

Pascali, Luigi. 2017. “The Wind of Change: Maritime Technology, Trade, and Economic Development.” *American Economic Review* 107 (9): 2821–54. DOI: 10.1257/aer.20140832. [245, 249]

Ploeckl, Florian. 2013. “The internal impact of a customs union; Baden and the Zollverein.” *Explorations in Economic History* 50 (3): 387–404. DOI: 10.1016/j.eeh.2013.05.001. [249]

Ploeckl, Florian. 2015. “The Zollverein and the Sequence of a Customs Union.” *Australian Economic History Review* 55 (3): 277–300. DOI: 10.1111/aehr.12083. [249]

Richter, Klaus W. 2007. *Die Wirkungsgeschichte Des Deutschen Kartellrechts VOR 1914: Eine Rechtshistorische-Analytische Untersuchung (Die Einheit Der Gesellschaftswissenschaften) (German Edition)*. Mohr Siebeck. URL: https://www.amazon.com/Wirkungsgeschichte-Deutschen-Kartellrechts-Rechtshistorische-Analytische-Gesellschaftswissenschaften/dp/3161492323?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3161492323. [287]

Siemann, Wolfram. 1995. *Vom Staatenbund zum Nationalstaat: Deutschland 1806-1871 (Neue deutsche Geschichte) (German Edition)*. Beck. URL: https://www.amazon.com/Vom-Staatenbund-zum-Nationalstaat-Deutschland/dp/3406308198?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3406308198. [262]

Smalian. 1809. “Charte von dem Departemente der Fulda des Königreichs Westphalen.” Weimar, Verlag des Geographischen Instituts. [275]

Spielmann, Christian. 1926. *Geschichte von Nassau: 2. Kultur- und Wirtschaftsgeschichte*. [292]

Stauber, Reinhard. 2014. *Der Wiener Kongress*. UTB GmbH. URL: https://www.amazon.com/Der-Wiener-Kongress/dp/3825240959?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3825240959. [258]

Steindl, Harald. 1986. “Die Einführung der Gewerbefreiheit.” In. Vol. 3, Helmut Coing, 3527–633. [287]

Tabellini, Guido. 2010. “Culture and Institutions: Economic Development in the Regions of Europe.” *Journal of the European Economic Association* 8 (4): 677–716. DOI: 10.1111/j.1542-4774.2010.tb00537.x. [246, 285, 288, 290]

Viner, Jacob. 1950. “The customs union issue.” *Carnegie Endowment for International Peace*, [249]

Vogel, Barbara. 1983. *Allgemeine Gewerbefreiheit: Die Reformpolitik des preussischen Staatskanzlers Hardenberg (1810-1820) (Kritische Studien zur Geschichtswissenschaft) (German Edition)*. Vandenhoeck & Ruprecht. URL: https://www.amazon.com/Allgemeine-Gewerbefreiheit-Reformpolitik-Staatskanzlers-Geschichtswissenschaft/dp/3525357168?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=3525357168. [287]

Voth, Hans Joachim. 2001. “Modern Prussian History: 1830-1947.” In. Philip Dwyer. Chapter The Prussian Zollverein and the bid for economic superiority, 109–25. [249]

Wolf, Nikolaus, and Thilo Huning. 2019. “How Britain Unified Germany: Endogenous Trade Costs and the Formation of a Customs Union.” *CEPR Discussion Paper No. DP13634*, [249]

Publikationsliste

Dohmen, Martin. 2022. „Freedom of enterprise and economic development in the German industrial take-off". *European Economic Review* 148 (September): DOI: 10.1016/j.euroecorev.2022.104163.