STAFF REPORTS

NO. 999
DECEMBER 2021

Superstar Returns

Francisco Amaral | Martin Dohmen | Sebastian Kohl | Moritz Schularick

FEDERAL RESERVE BANK of NEW YORK

Superstar Returns
Francisco Amaral, Martin Dohmen, Sebastian Kohl, and Moritz Schularick
Federal Reserve Bank of New York Staff Reports, no. 999
December 2021
JEL classification: G10, G12, N90, R21, R31

Abstract

We study long-term returns on residential real estate in twenty-seven "superstar" cities in fifteen countries over 150 years. We find that total returns in superstar cities are close to 100 basis points lower per year than in the rest of the country. House prices tend to grow faster in the superstars, but rent returns are substantially greater outside the big agglomerations, resulting in higher long-run total returns. The excess returns outside the superstars can be rationalized as a compensation for risk, especially for higher co-variance with income growth and lower liquidity. Superstar real estate is comparatively safe.

Key words: housing returns, housing risk, superstar cities, regional housing markets

---

Schularick: MacroFinance Lab, University of Bonn, Sciences Po Paris, and Federal Reserve Bank of New York (email: schularick@uni-bonn.de). Amaral: Macro Finance Lab, University of Bonn (email: francisco.amaral@uni-bonn.de). Dohmen: Macro Finance Lab, University of Bonn (email: mdohmen@uni-bonn.de). Kohl: MPIfG Cologne (email: kohl@mpifg.de). This work is part of a larger project supported by the European Research Council Grant (ERC-2017-COG 772332). The authors thank a large number of researchers who helped with the data on individual cities, especially Jacques Friggit, Matthijs Korevaar, Avner Offer and Chihiro Shimizu. They also thank Andrea Balestracci, Melina Cosentino, Gregor Jost, Dimitrios Kanelis, Gerwin Kiessling, Liudmila Kiseleva, Carla Revilla, Markus Schick, and Masashi Tanigaki for their excellent research assistance. In addition, they thank Benjamin Born, Esther Eiling, Marco Giacoletti, Christian Hilber, Nir Jaimovich, Christine Laudenbach, Lara Loewenstein, Ronan Lyons, Jonathan Moreno Medina, Marco Pagano, Joseph E. Stiglitz, Daniel Sturm, and Sofie Waltl for helpful feedback. The authors acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy – EXC 2126/1 – 39083886. They 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.

To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr999.html.

1 Introduction

This paper introduces and analyzes a new long-run city-level data set covering annual house prices and rents in 27 "superstar" cities in 15 OECD countries over the past 150 years. We borrow the "superstar city" terminology from the well-known paper by Gyourko, Mayer, and Sinai (2013) for the U.S., but take it global in the sense that we study the main economic agglomerations in these 15 countries. For each national superstar 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 the rest 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 rental series from city yearbooks or primary sources such as newspapers, tax records, and notary archives.[1]

Our central finding is that, over the long-run, superstar cities have witnessed lower total returns on residential real estate than other parts of the same country. With 5.75 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 superstar cities comes with a negative return premium of approximately 90 basis points annually relative to national returns (including the superstars) and about 100 basis points lower than the rest of the country (excluding the superstars). 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 superstar 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.

Why are housing returns persistently lower in large cities when compared to the rest of the economy? Our key finding can be rationalized in a standard asset pricing framework where excess returns are a compensation for risk. Observable long-run return

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

differences between different assets must be attributable to differences in risk, or to violations of standard assumptions (such as persistent behavioral biases in expectations). Suppose that everything that makes a national superstar city – its diversified economy, its large market, its amenities, the international demand – also makes it a safer place as an investment.[2] 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 superstars 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 empirical support for this interpretation of the (negative) superstar premium. On the one hand, we present evidence that the co-variance between housing returns and income growth is lower in large cities. Between 1950 and 2018 within the US the co-variance between MSA-level income growth and MSA-level excess 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 show 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 superstar cities witnessed lower housing returns might seem counter-intuitive 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 superstar 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 superstars and other parts of the country (albeit with mixed statistical significance). At the same time, rent returns are considerably higher outside the large cities. Our point estimate puts the mean rent return differential at 160 basis points per annum. The differences in rent-price ratios exceed the differences in capital gains in the long run and lead to lower total housing returns.

[2]The urban economics literature has documented a series of additional facts about large cities that set them apart from the rest of the country (Black and Henderson (1999), Desmet and Henderson (2015)).

Differences in housing risk can also rationalize these divergent patterns in capital gains and rent returns. Assuming housing risk is lower in superstars, 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, 2015), explaining higher price gains in the superstars *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 superstar 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 national superstar cities, and experimented with different size cut-offs in different eras. Once more, none of this altered the core findings in a systematic way.

The lack of available data remains a central challenge to research on housing markets (Piazzesi, 2018). Our work is the first to put together international long-run housing 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 et al., 2019) and an international dimension to individual papers on long-run housing returns in individual regions (Demers and Eisfeldt, 2021; Eichholtz et al., 2020; Keely and Lyons, 2020). The 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 been focused on the spatial distribution of house price appreciation (Gyourko, Mayer, and Sinai, 2013; Hilber and Vermeulen, 2015; Saiz, 2010), we demonstrate that the spatial distribution of housing returns is different – a fact that we explain with differences in local housing market risks. Finally, our paper also contributes to the nascent literature on the risk-return relation in housing markets (Demers and Eisfeldt, 2021; Jordà et al., 2019; Giacoletti, 2021; Sagi, 2021)

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.

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 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.

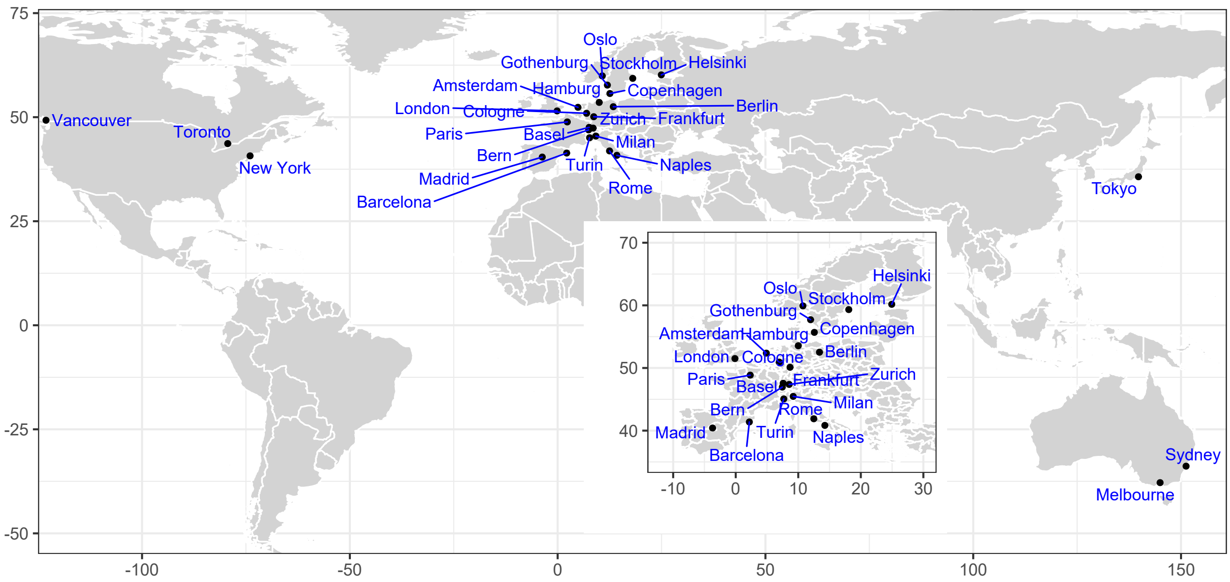
2.1 City sample

We focused our data collection on the largest cities within 15 developed countries that are covered by the Jorda-Schularick-Taylor Database and Jordà 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

[3] We included fewer cities in countries such as France where a dominant urban center was well established by the 19$^{\text{th}}$ 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.

Figure 1: *Geographical distribution of our city sample*



Note: 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).

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. 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.

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: *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

Note: 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.

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.

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 Eichholtz et al. (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 2 panel (a) shows an example of the property ad section of the *Kensington Post* in 1965. The advertisements used in our final index are marked.

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.

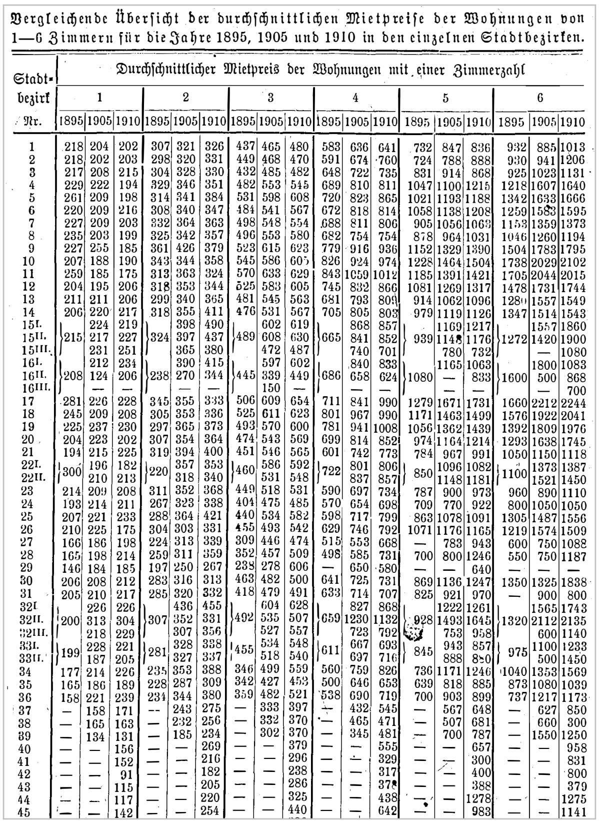
Regarding the construction of rent indices, one concern is that the rise of urban home ownership might have made rents less important than they have been historically. In cities like Oslo or Rome, private rentals occupy a small segment of the market, and the majority of homes are in private ownership. However, Figure 3 shows that the majority of urban households are tenants, as in all German-speaking cities but also in New York and Paris.

That said, rent data was often harder to find than house price data. We rely on rent indices from statistical agencies as city-level rent indices were constructed by city

Figure 2: *Examples of primary and secondary sources*

(a) *London, 1965*

(b) *Frankfurt, 1895-1910*



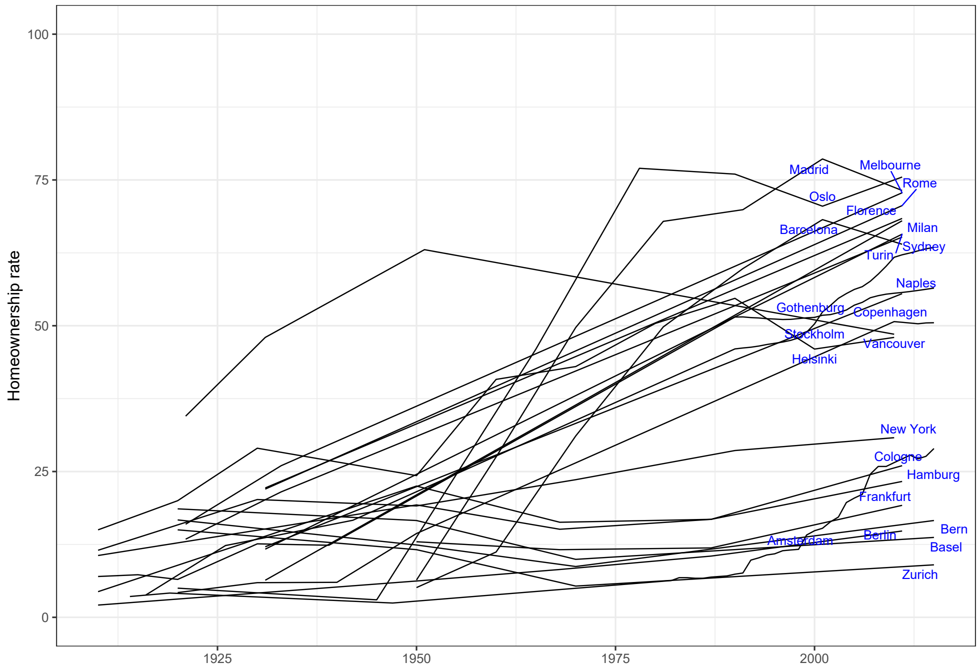
Note: Panel (a): Extract of the real estate part of the ad section for the 14th of May 1965 from the newspaper Kensington Post. Panel (b): From Beiträge zur Statistik der Stadt Frankfurt am Main 11. NF (1919) published in Busch (1919).

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 ten years and typically covered all rental units, providing a precise picture of the universal level of rents in a specific city. Figure 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

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

Figure 3: *Urban home ownership rates, long sample cities*



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

so-called *Stadtbezirke*. We use this data to benchmark our rent series between 1895 and 1910. Table 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.

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]

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

[5] In appendix 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.

Table 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

Note: 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.

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:

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

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: $Rent\ return = \frac{R_t}{P_t} * \frac{HPI_t^{nom}}{HPI_{t-1}^{nom}}$.

Rent-price ratio estimates are constructed following the rent-price approach used in Jordà et al. (2019) and Brounen et al. (2013). To do so, we first use benchmark rent-price ratios for the end of our sample period in 2018. We again follow Jordà 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:

$$\frac{RI_{t+1}}{HPI_{t+1}} = \left(\frac{RI_{t+1}/RI_t}{HPI_{t+1}/HPI_t}\right)\frac{RI_t}{HPI_t} \tag{2}$$

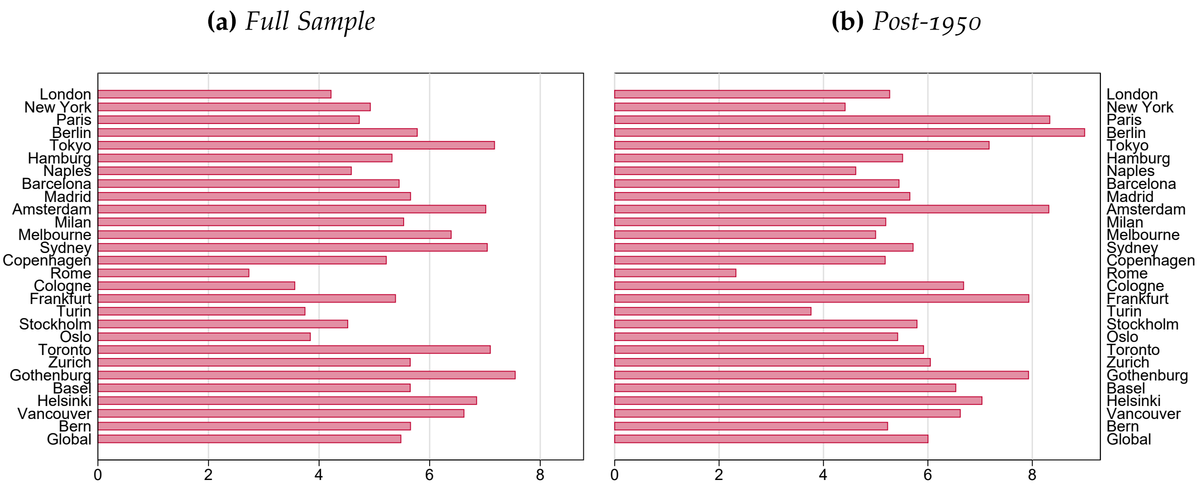
The disadvantage of this methodology is that possible measurement errors accumulate over time due to extrapolation. To account for this, we collected historical rent-price ratios to verify our rental yield series. Whenever the rent-price approach estimates diverge from these historical sources, we adjust 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 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 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 A.2.[6]

[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.

Figure 4: *City-level real average total housing returns (log points)*



Note: 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. Panel (b) shows average housing return data by city starting in 1950.

3 Long-term superstar returns

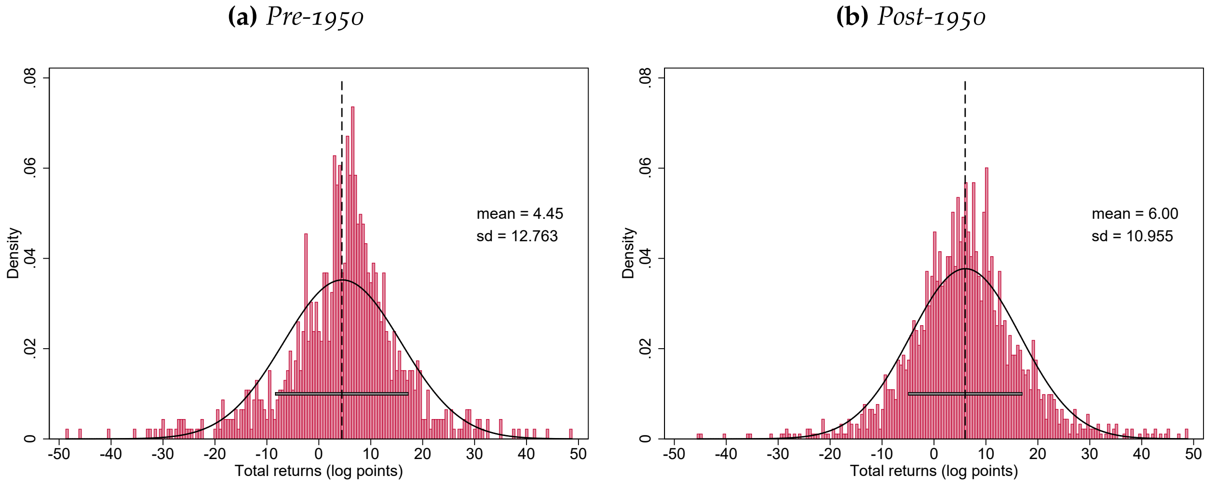
In this section, we first establish the main stylized facts on long-run superstar housing returns. We then proceed to analyze trends in capital gains and rent returns, as well as their contributions to total returns, and compare superstars to the rest of the country.

We start with summary statistics on real log housing returns and its components for our new data set. The left-hand panel of Figure 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 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, 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 superstar returns were close to 2 percentage points higher

[7] Appendix Table 12 shows summary statistics by city in numbers including standard deviations. Appendix Table 13 adds summary statistics with average percentage point (simple) returns for comparison to other literature.

Figure 5: *Distribution of annual real housing returns (log points)*



Note: 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. Panel (b) covers the entire sample of cities after 1950.

with a lower standard deviation.

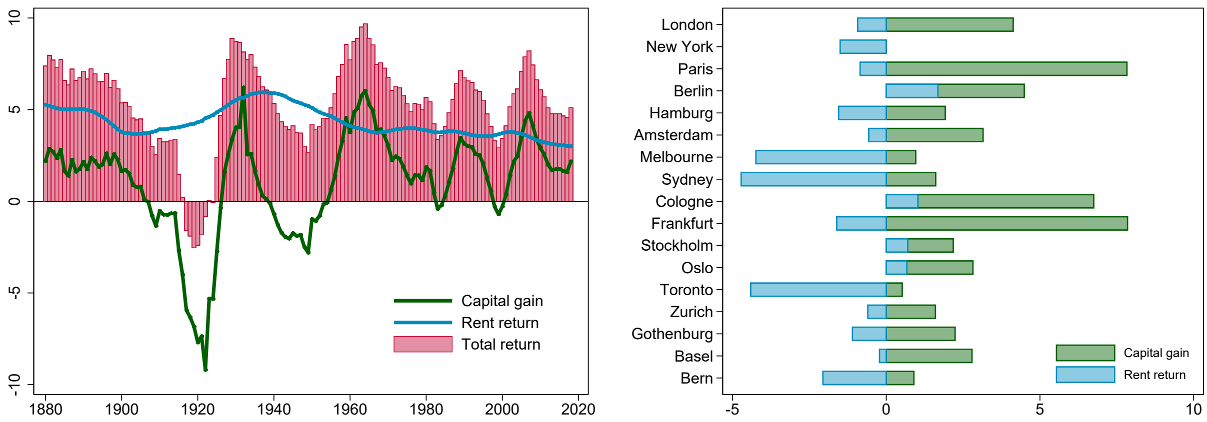
Panel (a) of Figure 6 plots 10-year lagged moving averages of log real housing returns averaged over the 27 national superstars over time. Housing returns dropped sharply around the World Wars as house prices dropped. The effect is particularly pronounced after World War I, as many governments introduced 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 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 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 returns mirrors the secular decline in the yield component of stock returns over the same period.

Figure 6: *City-level average returns (log points)*

(a) *All cities average* **(b)** *Absolute change by city, pre- vs post-1950*



Note: 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.

Rent returns represent approximately 67% of total housing returns over the last 150 years. Panel (a) of Figure 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 et al. (2019) and Demers and Eisfeldt (2021).

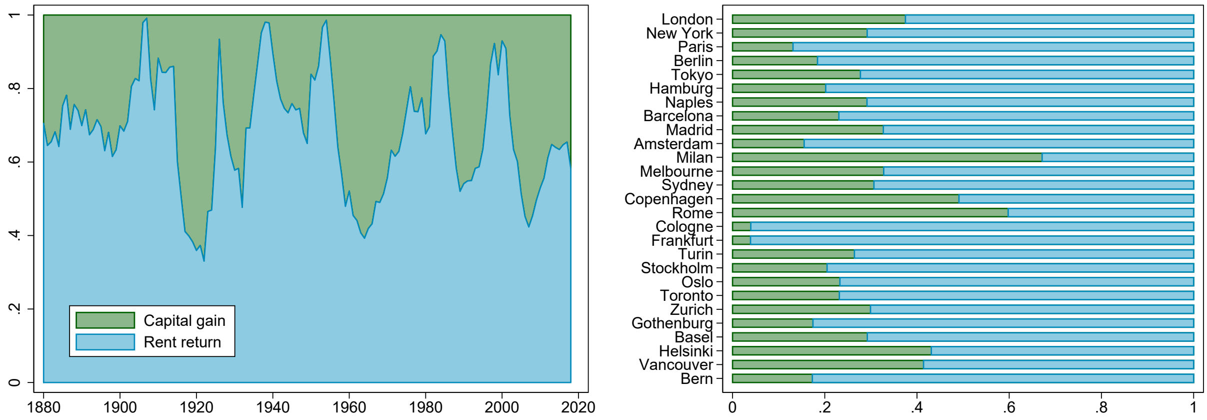
3.1 Superstars vs. national housing markets

In the next step, we merge our city-level data set with national housing returns from Jordà et al. (2019) in order to compare returns in the superstars to those in the rest of the country. Jordà 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 B.1.

The national housing portfolios in Jordà et al. (2019) include the national superstar cities. For transparency and comparability reasons, we will still compare the superstars

Figure 7: *Share of log total returns, 1870-2018*

(a) *Full sample* **(b)** *By city*



Note: 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 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 non-superstar locations in the country. National returns can be expressed as:

$$National\ return_t \quad = \quad w_{t-1} * National\ superstars\ return_t \ + \ (1 - w_{t-1}) * RoC\ return_t, \quad (3)$$

where w is the relative weight of the national superstar in the respective national housing series. Using equation 3 and our national superstar cities return series, we can approximate the housing returns in the rest of the country (*RoC return*) by subtracting the national superstars 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 superstars). 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 national superstars than a market capitalization weighted index. As a consequence, the rest of the country returns that we back-out from national series in Jordà et al. (2019) likely mark a lower bound.

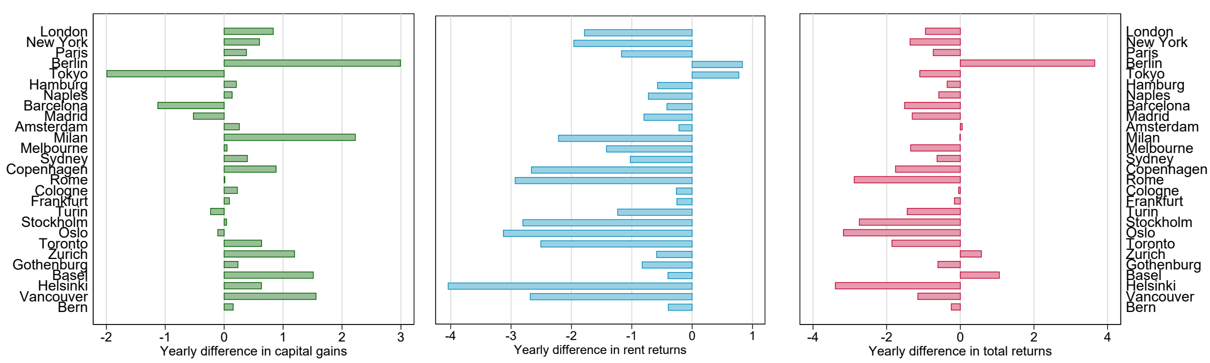
In some cases, the geographical coverage of the national housing series is too narrow in the pre-World War II era to allow a meaningful comparison between the superstar cities and the rest of the country. Appendix Table 14 shows the geographical coverage of the national house price series by country. For the comparison between superstar 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 B.3).

To guide the reader through the results, we start with an example of an undisputed national 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 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, the superstar city Paris underperformed the rest of France with respect to total returns on housing investment.

Figure 8: *Average differences in city-level and national returns (log points), 1950-2018*

(a) *Capital gain* **(b)** *Rent return* **(c)** *Total return*



Note: 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 8, we broaden the perspective to all 27 superstars in the sample and compare them to their national real estate markets. Figure 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 et al. (2019).[8] A general pattern can be easily discerned. Just like in the French case, capital gains are higher in nearly all superstars. 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]

Table 3 formalizes the analysis of different superstar/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 superstar city-level (Column 1) and for the national housing portfolios as defined in Jordà et al. (2019) (Column 2). Column (4) shows the population-weighted return for the rest of country (excluding the superstars), as defined above. The lower panel narrows the superstar definition to the single largest city in each country (New York, London, Paris, etc.), providing an even stronger superstar vs. rest of country comparison.[10]

The results essentially mirror those in the Parisian example before. At 2.24 log points capital gains have been about 41 basis points higher in the 27 national superstars than in the national portfolio, and 60 basis points higher than in the rest of the country. Rent returns, in contrast, have been lower in the superstars with a difference of 1.35 or 1.61 log points, depending on the comparison portfolio. The higher rent returns outside the superstars more than compensate for the lower rate of capital appreciation. Our overall benchmark estimate is that in the long-run total returns in the superstars were 90-100 basis points lower per year than in the national portfolio and rest of the country.

[8] Appendix Table 17 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 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.

Table 3: *City-level and national yearly housing returns (log points), 1950-2018*

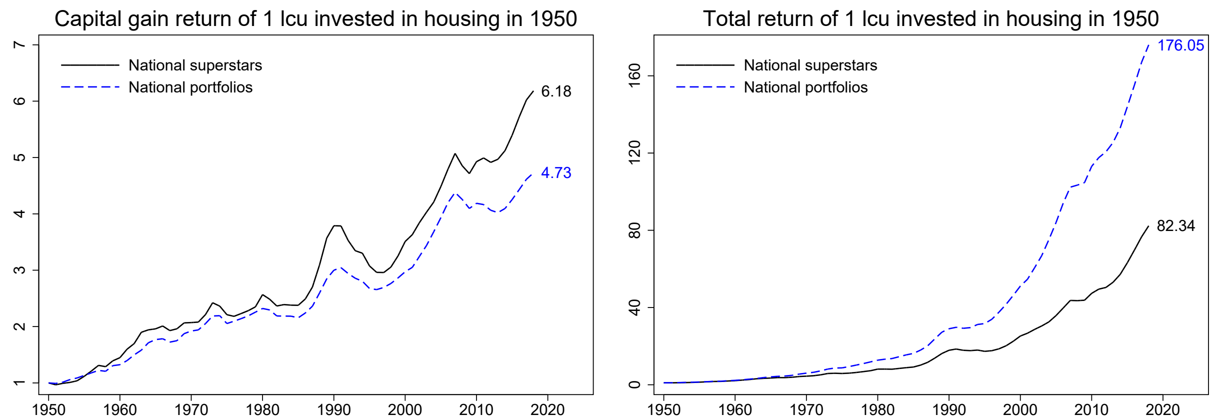
	27 national superstars				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	2.24	1.82	0.41* (0.24)	1.64	0.60** (0.26)
Rent return	3.59	4.94	-1.35*** (0.04)	5.21	-1.61*** (0.05)
Total return	5.75	6.68	-0.93*** (0.23)	6.76	-1.01*** (0.26)
N	1767				
	Only largest city/country				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	2.42	2.12	0.31 (0.30)	1.99	0.44 (0.34)
Rent return	3.60	5.17	-1.57*** (0.06)	5.41	-1.82*** (0.07)
Total return	5.93	7.18	-1.25*** (0.30)	7.30	-1.37*** (0.34)
N	1061				

Note: 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 superstars. 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 national superstars 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. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

The lower panel of Table 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 1.37 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 superstar 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.

At first sight the return differences of 90-100 basis points between the superstars and the rest might seem small. But differences in yearly returns can accumulate quickly and generate substantial return differences over time. Figure 9 shows that the city-level portfolio increased much faster in value than the national one, because house prices in the superstar cities appreciated more, namely by a factor of 6.2 compared to 4.7 for

Figure 9: *Cumulative portfolio returns of city-level vs national portfolios*



Note: 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.

the national portfolio.[11] Once again, the picture changes when rent returns are added in the right panel of Figure 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 twice as high as the returns in the superstars.

3.2 Further tests

Capital gains are higher in superstars, 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 D where we show that differences in real estate taxation do not affect our results.

[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 3.

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 superstars 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 4: *Yearly housing returns (log points) using alternative rental yield benchmarks, 1950-2018*

	27 national superstars			Only largest city/country		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.24	1.82	0.41* (0.24)	2.58	2.13	0.46 (0.29)
Rent return	3.37	4.94	-1.57*** (0.04)	3.49	5.20	-1.72*** (0.06)
Total return	5.53	6.68	-1.15*** (0.24)	5.97	7.22	-1.25*** (0.29)
N	1767			1004		

Note: 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. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 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 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.

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 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 5 shows the results for the 27 national superstars relative to the national index. Our key results also hold for the most recent period: superstar 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 B.6 presents results for other sub-periods as well as moving averages over the entire time period.

Table 5: *Yearly housing returns (log points) for 27 national superstars until and post 1990*

	Until 1990			Post 1990		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.64	2.21	0.43 (0.38)	1.69	1.31	0.39* (0.22)
Rent return	3.77	5.37	-1.60*** (0.07)	3.35	4.36	-1.01*** (0.04)
Total return	6.32	7.47	-1.15*** (0.38)	4.98	5.62	-0.63*** (0.22)
N	1011			756		

Note: 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. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

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 6 confirm that, independently of rent control regimes, capital gains are higher and rent returns lower in the superstar cities compared to the national average. The difference between superstars and the national returns is even slightly higher in weaker rent control regimes.

Table 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.	0.43 (0.45)	-1.67*** (0.08)	-1.22*** (0.45)	687

Note: 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. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

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. Otherwise the approach and the methodology are the same within data sets. The central question is whether the findings from the long-run comparison of national superstar 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?

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 Commu-*

nity 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 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 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 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 definition of superstar cities, 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

[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 C.1 and Gyourko, Mayer, and Sinai (2013).

[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.

Table 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

Note: 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 95^{th} 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 5^{th} 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. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$

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.

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.

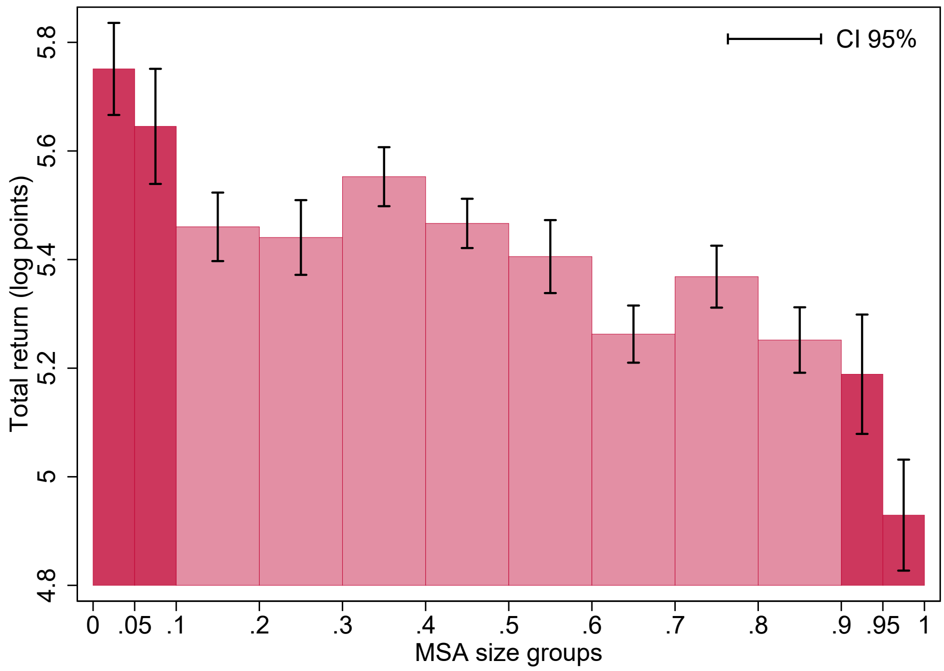
Average log total returns within each bin are plotted in Figure 10, which shows that overall housing returns decrease with city size. The relation is not perfectly monotonic across all size bins, but clearly visible overall.[15] Appendix Table 22 shows the decomposition of housing returns within size 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.

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

[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).

Figure 10: *Total returns for 316 MSAs (log points) by population size, 1950-2018*



Note: 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.

comparably large cities that correspond to urban municipalities excluding rural 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 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

[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.*

cities as being at or above the 5% largest of the size distribution.

Table 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

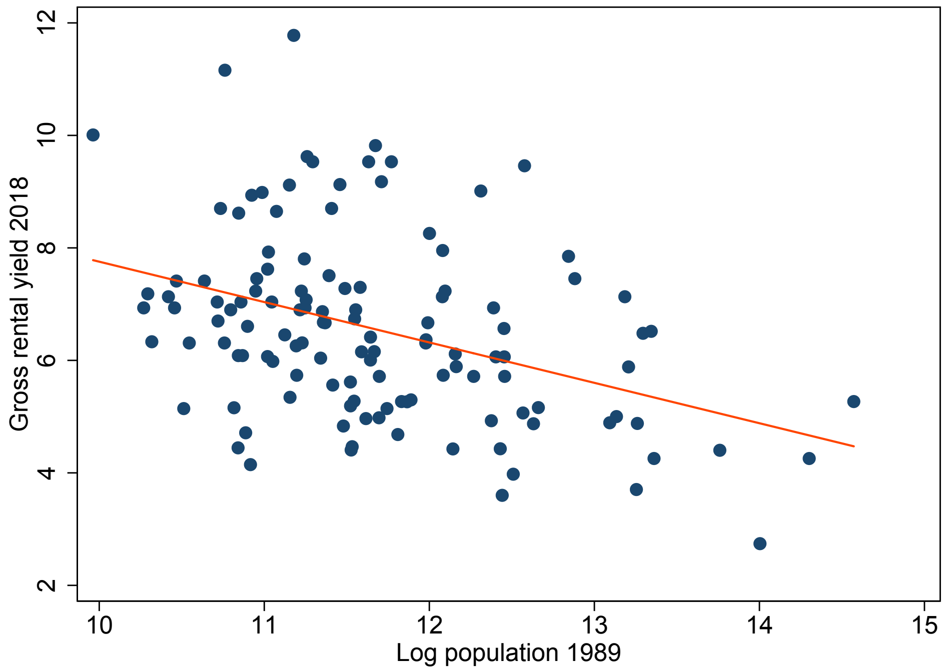
Note: The table shows differences in annual housing returns between large cities and the rest of the sample or small cities. It covers 42 major German cities between 1975 and 2018. Differences are measured as coefficients in a random effects panel regression of the dependent variable (log capital gain, 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 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 23: once more, rent returns are monotonically decreasing with city size, while capital gains are higher in large cities. Figure 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 considerably within size bins, a clear negative relation between city size and gross rental yields is visible in the raw data.[19]

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 superstar 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.

[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.

Figure 11: *Correlation of gross rental yields (log points) in 2018 and log population size*



Note: 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.

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 superstar/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 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 et al., 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 superstars and neglect the rent return component in total housing returns. In appendix 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:

$$\left(\frac{R_{t+1}^A}{P_t^A} + cg_{t+1}^A\right) * \frac{1}{\delta^A} = \left(\frac{R_{t+1}^B}{P_t^B} + cg_{t+1}^B\right) * \frac{1}{\delta^B}, \tag{4}$$

with P_t^l being the house price at time t in location l, R the rent payment, and $cg^l = \frac{P_{t+1}^l - P_t^l}{P_t^l}$ 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:

$$\frac{R_{t+1}^A}{P_t^A} + cg_{t+1}^A < \frac{R_{t+1}^B}{P_t^B} + cg_{t+1}^B \tag{5}$$

For simplicity of exposition, we assume that houses in both cities feature the same expected future rental cash-flow: $R_{t+1}^A = R_{t+1}^B$. 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 (4) to hold, current 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:

$$\frac{R_{t+1}^A}{P_t^A} < \frac{R_{t+1}^B}{P_t^B}, \tag{6}$$

This helps rationalize the empirical finding (ii) that rent returns are lower in large cities. In a next step, we can rewrite inequality (5) as:

$$cg_{t+1}^A - cg_{t+1}^B < \frac{R_{t+1}^B}{P_t^B} - \frac{R_{t+1}^A}{P_t^A}, \tag{7}$$

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,

[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.

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 (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 (2015) 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 national superstar 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 the superstars despite higher capital gains driven by lower rent returns. We will now explore the empirical evidence that housing risk is indeed lower in superstar cities and, if so, what risks investors outside the large cities are compensated for with higher returns.

5.1 Two sources of housing risk

When examining the difference in housing investment risk between cities, two separate sources of risk need to be considered: 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 show some 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.

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.[21]

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

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:

$$ln\mathbb{E}[R_{t+1}] - lnR_f = \gamma \; Cov \left[ln\left(\frac{C_{t+1}}{C_t}\right), lnR_{t+1} - lnR_f \right] \tag{8}$$

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.

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

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 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.[22]

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.[23] We then test whether these co-variances are smaller in large MSAs. The results are depicted in Table 9 column 3. It shows that the co-variances of income and total housing excess returns are indeed significantly smaller in large MSAs compared to the rest. Like for the difference in housing returns, the difference in co-variances becomes larger when we compare the largest MSAs to only the smallest ones. Appendix H.2 shows results for the entire distribution of MSAs as well as betas in the sense of the consumption capital asset pricing model (CCAPM). The results for betas are very similar.

Table 9: *Differences in co-variances between income and housing returns by city size, U.S, 1950-2018*

Sample	Capital gain	Rental yield	Total return	N
Large vs rest	-0.68** (0.317)	-0.65*** (0.126)	-0.75** (0.298)	316
Large vs small	-2.00*** (0.606)	-1.23*** (0.250)	-2.25*** (0.584)	31

Note: The table shows differences in the co-variance between income growth and 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 (co-variance) 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. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

[22] 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.

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

We do the same analysis for the two components of log total returns: log capital gains and log rental yields. We calculate the co-variances for each one of the components separately. The results can be found in Table 9 columns 1 and 2, which also show that co-variances for both components are smaller in the largest cities. In brief, the available evidence points to larger risks in small cities as income co-varies more with local housing returns.

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 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:[24]

$$\Delta p_{i,l,t} = \Delta v_{l,t} + BX_i + \sigma_{l,idiosyncratic}\,\varepsilon_{i,t}, \tag{9}$$

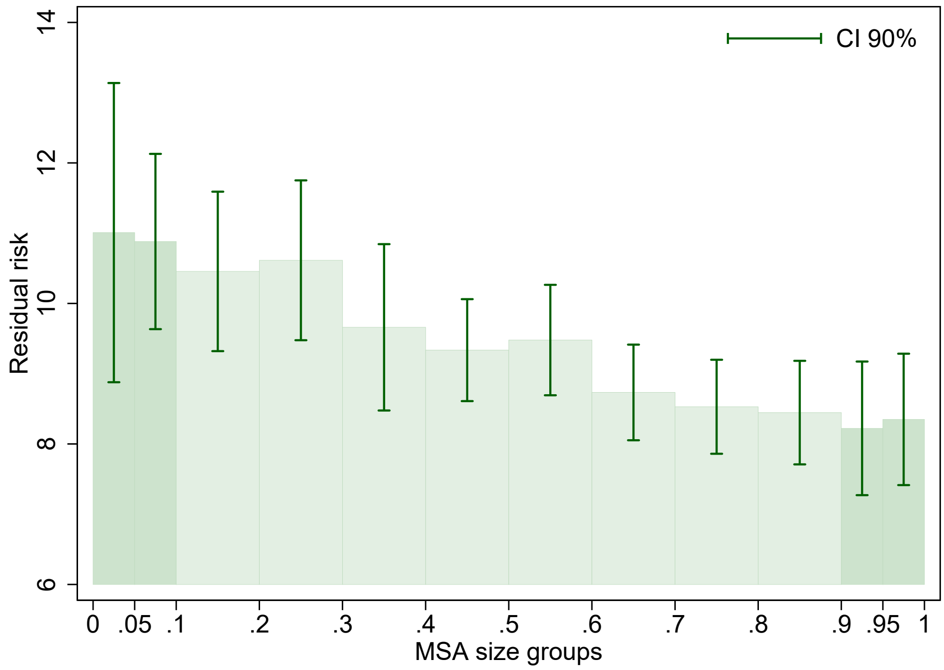
where $\Delta v_{l,t}$ is the growth in local county house prices, BX_i is a vector of house and transaction characteristics, which includes zip-code and time fixed effects, and $\sigma_{l,idiosyncratic}\ \varepsilon_{i,t}$ is a sales-specific shock. We then measure idiosyncratic risk as the standard 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 F and G.

[24]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.

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 house price reactions to positive demand shocks (Hilber and Vermeulen, 2015). However, as shown in the last section, overall house price growth co-varies less with income in the largest MSAs.[25] Idiosyncratic risk represents the largest share of total house price growth variation as can be seen in Appendix G.

Figure 12: *Annual idiosyncratic house price risk by MSA size, 1990-2020*



Note: 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 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 11.01% of the house

[25]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.

sales price, but about 25% lower in the largest MSAs at 8.35%.

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 a fixed parameter, but will be a function of local housing market and other factors: $\sigma_{l,idiosyncratic}(market\ characteristics_l, ...)$. The next 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 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.[26]

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).

[26]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.

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

Table 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

Note: 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 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 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

Note: 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.[27] Moving beyond U.S.

[27] 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.

data, in Appendix 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.

6 Conclusion

This paper constructed a novel data set covering long-run house prices, rent and housing return series for 27 national superstar cities. The historical data shows that the superstars tend to under-perform the rest of the country in terms of total returns.

This core finding can be rationalized by differences in risk. The data suggest 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 makes first steps towards a better understanding of spatial risk and return patterns in housing markets over the long run. We expect that the detailed data put together for this paper will allow future researchers to analyse and understand risk and return patterns in housing markets in greater detail.

References

Adam, Klaus, Oliver Pfäuti, and Timo Reinelt (2020). *Falling Natural Rates, Rising Housing Volatility and the Optimal Inflation Target*. Tech. rep. University of Bonn and University of Mannheim, Germany.

Ansell, Ben W. (2019). "The Politics of Housing". In: *Annual Review of Political Science* 22(1), pp. 165–185.

Armona, Luis, Andreas Fuster, and Basit Zafar (2018). "Home Price Expectations and Behaviour: Evidence from a Randomized Information Experiment". In: *The Review of Economic Studies* 86(4), pp. 1371–1410.

Arundel, Rowan and Cody Hochstenbach (2019). "Divided Access and the Spatial Polarization of Housing Wealth". In: *Urban Geography*, pp. 1–27.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel (2018). "The Economic Effects of Social Networks: Evidence from the Housing Market". In: *Journal of Political Economy* 126(6), pp. 2224–2276.

Becker, Richard A., Allan R. Wilks, Ray Brownrigg, Thomas P Minka, and Alex Deckmyn (2018). ""maps" - Draw Geographical Maps". In: *R-package*.

Black, Duncan and Vernon Henderson (1999). "Spatial Evolution of Population and Industry in the United States". In: *American Economic Review* 89(2), pp. 321–327.

Blanchard, Olivier J. and Lawrence F Katz (1992). "Regional evolutions". In: *Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution* 23(1).

Bogin, Alexander, William Doerner, and William Larson (2018). "Local House Price Dynamics: New Indices and Stylized Facts". In: *Real Estate Economics* 47(2), pp. 365–398.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020). "Overreaction in Macroeconomic Expectations". In: *American Economic Review* 110(9), pp. 2748–82.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2019). "Diagnostic Expectations and Stock Returns". In: *The Journal of Finance* 74(6), pp. 2839–2874.

Bracke, Philippe (2013). "How Long Do Housing Cycles Last? A Duration Analysis for 19 OECD Countries". In: *Journal of Housing Economics* 22(3), pp. 213–230.

Brounen, Dirk, Piet Eichholtz, Stefan Staetmans, and Marcel Theebe (2013). "Inflation Protection from Homeownership: Long-Run Evidence, 1814-2008". In: *Real Estate Economics* 42(3), pp. 662–689.

Bureau, Census (2018). "Metropolitan and Micropolitan Statistical Areas and Related Statistical Areas - States". In: *Cartographic Boundary Files - Shapefile,*

Busch, Dr. August (1919). "Beiträge zur Statistik der Stadt Frankfurt am Main". In: *Erstes Heft: Die Besiedlung des Frankfurter Stadtgebiets und die Befriedigung des Wohnungsbedürfnisses der Bevölkerung*.

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". In:

Cochrane, John H (2009). *Asset pricing: Revised edition*. Princeton university press.

De Stefani, Alessia (2020). "House Price History, Biased Expectations, and Credit Cycles: The Role of Housing Investors". In: *Real Estate Economics*.

Demers, Andrew and Andrea L. Eisfeldt (2021). "Total Returns to Single-Family Rentals". In: *Real Estate Economics*.

Desmet, Klaus and Vernon Henderson (2015). "The Geography of Development Within Countries". In: *Handbook of Regional and Urban Economics*. Elsevier, pp. 1457–1517.

Duranton, Gilles (2007). "Urban Evolutions: The Fast, the Slow, and the Still". In: *American Economic Review* 97(1), pp. 197–221.

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.

Eeckhout, Jan (2004). "Gibrat's Law for (All) Cities". In: *American Economic Review* 94(5), pp. 1429–1451.

Eichholtz, Piet, Matthijs Korevaar, Thies Lindenthal, and Ronan Tallec (2020). "The Total Return and Risk to Residential Real Estate". In: *SSRN Electronic Journal*.

Eurostat (2013). *Handbook on Residential Property Prices (RPPIs)*. International Monetary Fund.

Ewens, Michael, Charles M. Jones, and Matthew Rhodes-Kropf (2013). "The Price of Diversifiable Risk in Venture Capital and Private Equity". In: *Review of Financial Studies* 26(8), pp. 1854–1889.

Fama, Eugene F and Kenneth R French (1993). "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics*.

Gennaioli, Nicola and Andrei Shleifer (2018). *A Crisis of Beliefs*. Princeton University Press.

Giacoletti, Marco (2021). "Idiosyncratic Risk in Housing Markets". In: *The Review of Financial Studies* 34(8). Ed. by Stijn Van Nieuwerburgh, pp. 3695–3741.

Goetzmann, William N. and Alok Kumar (2008). "Equity Portfolio Diversification*". In: *Review of Finance* 12(3), pp. 433–463.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai (2013). "Superstar Cities". In: *American Economic Journal: Economic Policy* 5(4), pp. 167–199.

Han, Lu and William C. Strange (2015). "The Microstructure of Housing Markets". In: *Handbook of Regional and Urban Economics*. Elsevier, pp. 813–886.

Hilber, Christian and Andreas Mense (2021). "Why Have House Prices Risen so Much More than Rents in Superstar Cities?" In:

Hilber, Christian and Wouter Vermeulen (2015). "The Impact of Supply Constraints on House Prices in England". In: *The Economic Journal* 126(591), pp. 358–405.

Hill, Robert J. (2012). "Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy*". In: *Journal of Economic Surveys*.

Hub, ArcGIS (2019). "Bundesländer 2019 mit Einwohnerzahl". In: *Esri Deutschland Content*.

Hudson, Robert S. and Andros Gregoriou (2015). "Calculating and Comparing Security Returns is Harder than You Think: A Comparison between Logarithmic and Simple Returns". In: *International Review of Financial Analysis* 38, pp. 151–162.

Hull, John C (2019). *Options, Futures and Other Derivatives*.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M Taylor (2019). "The Rate of Return on Everything, 1870-2015*". In: *The Quarterly Journal of Economics*.

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

Keely, Richard and Ronan C. Lyons (2020). "Housing Prices, Yields and Credit Conditions in Dublin since 1945". In:

Kemeny, Jim (1995). *From Public Housing to the Social Market: Rental Policy Strategies In Comparative Perspective*. Routledge: London et al.

Kholodilin, Konstantin (2020). "Long-Term, Multicountry Perspective on Rental Market Regulations". In: *Housing Policy Debate* 30(6), pp. 994–1015.

Kholodilin, Konstantin and Sebastian Kohl (2021). "Social Policy or Crowding-out? Tenant Protection in Comparative Long-run Perspective". In: *Housing Studies*, pp. 1–24.

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". In: *DIW Discussion Papers*.

Klick, Larissa and Sandra Schaffner (2020). "Regional Real Estate Price Indices for Germany, 2008 – 2019: RWI–GEO–REDX". In: *Jahrbücher für Nationalökonomie und Statistik* 241(1), pp. 119–129.

Knoll, Katharina, Moritz Schularick, and Thomas Steger (2017). "No Price Like Home: Global House Prices, 1870-2012". In: *American Economic Review* 107(2), pp. 331–353.

Kohl, Sebastian (2017). *Homeownership, Renting and Society: Historical and Comparative Perspectives*. Routledge: London.

Kohl, Sebastian and Jardar Sørvoll (2021). "Varieties of Social Democracy and Cooperativism: Explaining the Divergence between Housing Regimes in Nordic and German-Speaking Countries." In: *Social Science History*.

Kotova, Nadia and Anthony Lee Zhang (2019). "Search Frictions and Idiosyncratic Price Dispersion in the US Housing Market". In: *SSRN Electronic Journal*.

Kuchler, Theresa and Basit Zafar (2019). “Personal Experiences and Expectations about Aggregate Outcomes”. In: *The Journal of Finance* 74(5), pp. 2491–2542.
Kuvshinov, Dmitry and Kaspar Zimmermann (2020). “The Expected Return on Risky Assets: International Long-run Evidence”. In: *SSRN 3546005*.
Merton (1987). “A Simple Model of Capital Market Equilibrium with Incomplete Information”. In: *The Journal of Finance*.
Meucci, Attilio (2010). “Quant Nugget 2: Linear vs. Compounded Returns–Common Pitfalls in Portfolio Management”. In: *GARP risk professional*, pp. 49–51.
Moretti, Enrico (2013). “Real Wage Inequality”. In: *American Economic Journal: Applied Economics* 5(1), pp. 65–103.
Mueller, Elisabeth (2010). “Returns to Private Equity - Idiosyncratic Risk Does Matter!” In: *Review of Finance* 15(3), pp. 545–574.
Paciorek, Andrew (2013). “Supply Constraints and Housing Market Dynamics”. In: *Journal of Urban Economics* 77, pp. 11–26.
Parker, Jonathan A. and Christian Julliard (2005). “Consumption Risk and the Cross Section of Expected Returns”. In: *Journal of Political Economy* 113(1), pp. 185–222.
Piazzesi, Monika (2018). “Asset Pricing Program”. In: *NBER Reporter* 2018(2), p. 1.
Piazzesi, Monika and Martin Schneider (2016). “Housing and Macroeconomics”. In: *Handbook of Macroeconomics*. Elsevier, pp. 1547–1640.
Pooley, Colin G. (1992). *Housing Strategies in Europe, 1880-1930*. Leicester University Press.
Reba, Meredith, Femke Reitsma, and Karen C. Seto (2016). “Spatializing 6,000 Years of Global Urbanization from 3700 BC to AD 2000”. In: *Scientific Data* 3(1).
Robinson, David T. and Berk A. Sensoy (2016). “Cyclicality, Performance Measurement, and Cash Flow Liquidity in Private Equity”. In: *Journal of Financial Economics* 122(3), pp. 521–543.
Rozsypal, Filip and Kathrin Schlafmann (2020). “Overpersistence Bias in Individual Income Expectations and Its Aggregate Implications”. In: *Working Paper*.
Sagi, Jacob S (2021). “Asset-Level Risk and Return in Real Estate Investments”. In: *The Review of Financial Studies* 34(8). Ed. by Stijn Van Nieuwerburgh, pp. 3647–3694.
Saiz, Albert (2010). “The Geographic Determinants of Housing Supply”. In: *The Quarterly Journal of Economics* 125(3), pp. 1253–1296.
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”. In: *Comparative European Politics* 6, pp. 237–261.
Soo, Cindy K (2018). “Quantifying Sentiment with News Media across Local Housing Markets”. In: *The Review of Financial Studies* 31(10), pp. 3689–3719.
Sorensen, Morten, Neng Wang, and Jinqiang Yang (2014). “Valuing Private Equity”. In: *Review of Financial Studies* 27(7), pp. 1977–2021.

Stapledon, Nigel (2007). "Long Term Housing Prices in Australia and Some Economic Perspectives". In: *PhD diss., University of New South Wales, Sydney*.

Stapledon, Nigel (2012). "Trends and Cycles in Sydney and Melbourne House Prices from 1880 to 2011". In: *Australian Economic History Review* 52(3), pp. 293–317.

Tversky, Amos and Daniel Kahneman (1983). "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." In: *Psychological review* 90(4), p. 293.

United Nations (2018). "World Urbanization Prospects: The 2018 Revision, Online Edition". In: *Department of Economic and Social Affairs, Population Division*.

Van Dijk, Dorinth (2019). "Local Constant-Quality Housing Market Liquidity Indices". In: *SSRN Electronic Journal*.

White, Eugene N, Kenneth Snowden, and Price V Fishback (2014). *Housing and Mortgage Markets in Historical Perspective*. The University of Chicago Press.

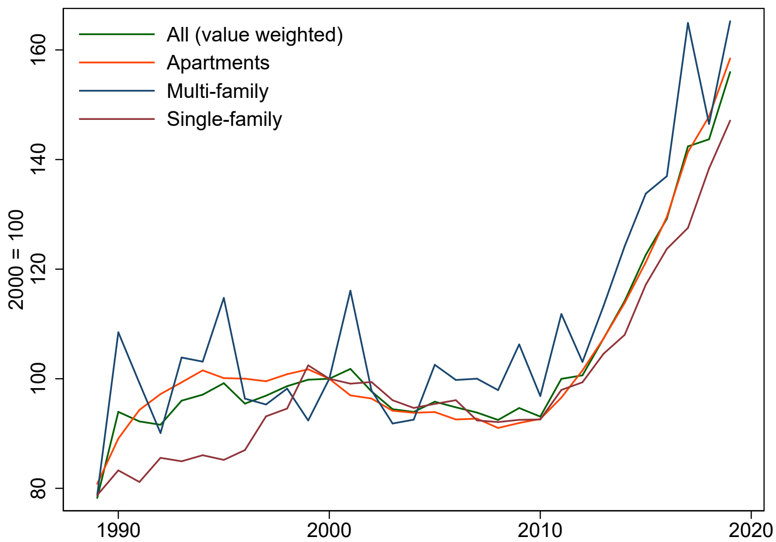
Appendix

A Additional data analyses

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 13: *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.

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

A.2 Log returns

Throughout the paper and in contrast to Jordà 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:

$$\text{Total return}_t \;=\; ln\left(\frac{P_t + R_t}{P_{t-1}}\right) \tag{10}$$

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.[28] 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,[29] 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

[28] See Hudson and Gregoriou (2015).

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

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 5 suggests that log total housing returns are indeed close to be normally distributed. Even if the assumption of normally distributed log returns 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.[30] But there are also disadvantages of using log returns instead of simple returns.[31] 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.

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

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

A.3 Summary statistics

Table 12: *Summary statistics on city-level housing returns (log points)*

City	Full sample			Post 1950		
	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	1.27 (15.73)	4.24 (1.43)	5.45 (15.52)	1.27 (15.73)	4.24 (1.43)	5.45 (15.52)
Madrid	1.87 (16.77)	3.85 (1.08)	5.65 (16.50)	1.87 (16.77)	3.85 (1.08)	5.65 (16.50)
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)	2.69 (1.02)	5.22 (8.93)	2.86 (8.80)	2.40 (0.73)	5.18 (8.88)
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)	3.61 (1.04)	4.52 (8.51)	1.93 (8.48)	3.93 (1.08)	5.79 (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	1.67 (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	1.44 (14.86)	4.10 (1.96)	5.48 (11.62)	2.43 (11.02)	3.66 (1.64)	6.00 (10.95)

Note: 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. 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 13: *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	2.51 (16.11)	4.34 (1.50)	6.84 (16.53)	2.51 (16.11)	4.34 (1.50)	6.84 (16.53)
Madrid	3.33 (17.82)	3.94 (1.12)	7.26 (18.24)	3.33 (17.82)	3.94 (1.12)	7.26 (18.24)
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)	2.73 (1.05)	5.77 (9.33)	3.29 (9.02)	2.43 (0.75)	5.72 (9.29)
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)	3.69 (1.08)	4.99 (8.75)	2.30 (8.53)	4.02 (1.13)	6.32 (8.73)
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	2.14 (12.04)	4.20 (2.08)	6.34 (12.41)	3.08 (11.40)	3.74 (1.71)	6.82 (11.77)

Note: 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. 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.

B Additional results for city vs national comparison

B.1 National housing data

Table 14 shows the geographical coverage of the national house price series used by Jordà 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.

National rent series from Jordà 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 et al. (2019).

We adapted the housing series of Jordà 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 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 et al. (2019).

Table 14: *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

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 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 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 et al. (2019).

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

In this section, we perform our main analysis from section 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 15. The Table shows that our results are both present in Europe as well as outside Europe.

Table 15: *City-level and national yearly housing returns (log points), 1950-2018*

	Europe				
	Cities	National	Difference	RoC	Cities - RoC
Capital gain	2.32	1.86	0.46* (0.27)	1.66	0.66** (0.29)
Rent return	3.59	4.90	-1.31*** (0.05)	5.14	-1.55*** (0.05)
Total return	5.82	6.67	-0.84*** (0.27)	6.72	-0.89*** (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				

Note: 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. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

B.3 Long-run comparison between national superstars and national housing portfolios

In this section, we repeat our main analysis from section 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 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 national superstars 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,[32] Norway starting 1891, the United Kingdom starting 1930[33] and the United States starting 1920.[34]

Table 16: *City-level and national yearly housing returns (log points), long-run*

	25 national superstars			Only largest city/country		
	Cities	National	Difference	Cities	National	Difference
Capital gain	2.14	1.72	0.42* (0.24)	2.36	1.98	0.37 (0.30)
Rent return	3.65	5.11	-1.46*** (0.04)	3.74	5.39	-1.65*** (0.06)
Total return	5.70	6.75	-1.06*** (0.24)	6.01	7.27	-1.26*** (0.30)
N	1920			1039		

Note: 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 national superstars 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 national superstars 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. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

The results adding the national superstars within these countries before 1950 are depicted in Table 16. For the sample of all 27 national superstars, the results become, if anything, even stronger than when only including the data post 1950 in section 3.1. For the sample of only the largest city per country, the results stay virtually 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

[32] 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).

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

[34] As can be seen from Table 14, 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 et al. (2019), and are moreover implausible compared to the housing series for Sydney and Melbourne from Stapledon (2012); Stapledon (2007), 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.

precisely measured.[35]

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.

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

[35]The difference in total returns is -0.98** for all national superstars 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.

Table 17: *Difference in yearly housing returns (log points) by cities, 1950-2018*

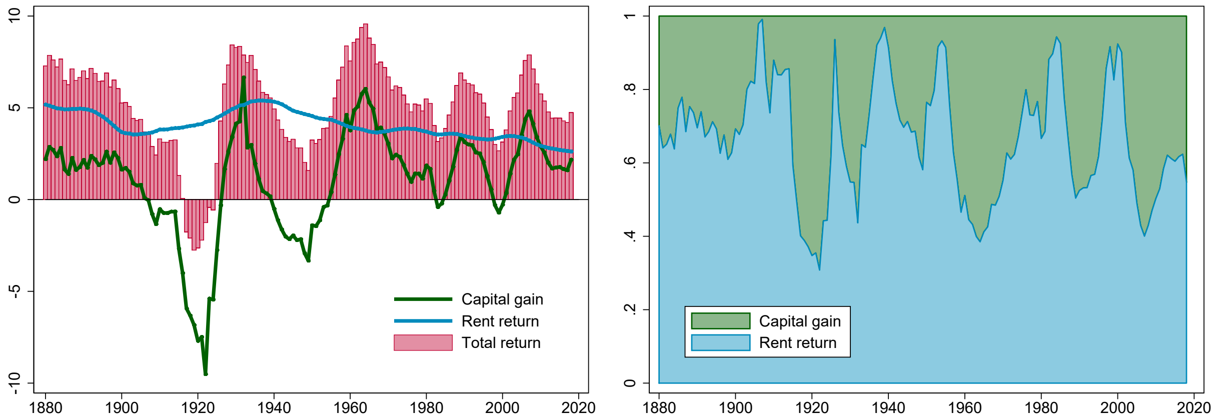
City	Capital gain	Rent return	Total return	N
London	0.83 (0.81)	-1.78*** (0.17)	-0.95 (0.83)	68
New York	0.60 (1.45)	-1.96*** (0.08)	-1.36 (1.43)	68
Paris	0.38 (0.79)	-1.17*** (0.11)	-0.74 (0.78)	68
Berlin	2.99** (1.15)	0.83*** (0.23)	3.65*** (1.16)	56
Tokyo	-1.99 (1.96)	0.77*** (0.24)	-1.10 (1.93)	59
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	-1.13 (2.09)	-0.42*** (0.15)	-1.51 (2.04)	68
Madrid	-0.52 (1.92)	-0.80*** (0.19)	-1.31 (1.89)	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)	-2.66*** (0.18)	-1.76*** (0.50)	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)	-2.80*** (0.20)	-2.74*** (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

Note: 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. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

B.5 Alternative rental yield benchmarks

Figure 14: *Log city-level returns and their decomposition, 1870-2018*

(a) *All cities average* **(b)** *Share of log total returns*



Note: 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 6.

Table 18: *Summary statistics on returns in log points*

City	Full sample			Post 1950		
	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	1.27 (15.73)	3.81 (1.45)	5.03 (15.50)	1.27 (15.73)	3.81 (1.45)	5.03 (15.50)
Madrid	1.87 (16.77)	3.77 (1.08)	5.57 (16.51)	1.87 (16.77)	3.77 (1.08)	5.57 (16.51)
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)	3.57 (0.98)	6.07 (8.99)	2.86 (8.80)	3.44 (0.94)	6.19 (8.94)
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)	2.70 (0.78)	3.61 (8.54)	1.93 (8.48)	2.94 (0.81)	4.81 (8.33)
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	1.44 (14.90)	3.81 (1.76)	5.19 (11.63)	2.43 (11.02)	3.42 (1.41)	5.77 (10.95)

Note: 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.

B.6 Results for different sub-periods

Table 19: *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 (0.441)	2.04	1.61	0.43 (0.324)
Rent return	3.80	5.75	-1.95*** (0.101)	3.23	4.44	-1.21*** (0.058)
Total return	6.66	8.11	-1.45*** (0.445)	5.21	5.98	-0.78** (0.324)
N	584			420		

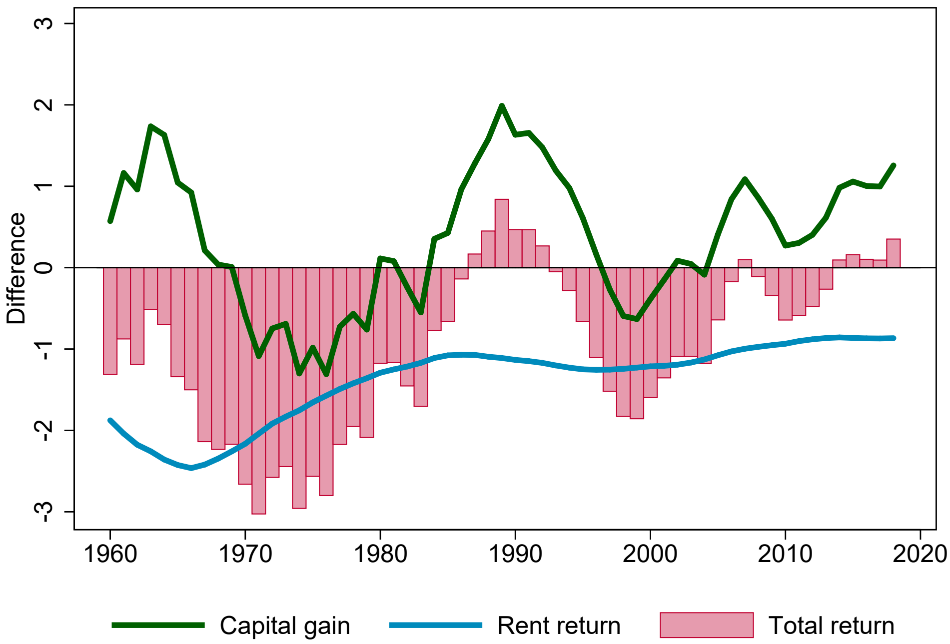
Note: The table shows averages of city-level and national log capital gains, log rent returns and log housing returns as well as the difference. National return averages are weighted by the number of cities in the respective country in the sample. Standard errors of differences (in parenthesis) and significance 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. As we want to minimize the effect of housing cycles, we compute 10 year lagged moving averages of this average difference.[36]

The outcomes are plotted in Figure 15. 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.

[36] 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.

Figure 15: *Average differences in city-level and national returns (log points) over time, 1950-2018*



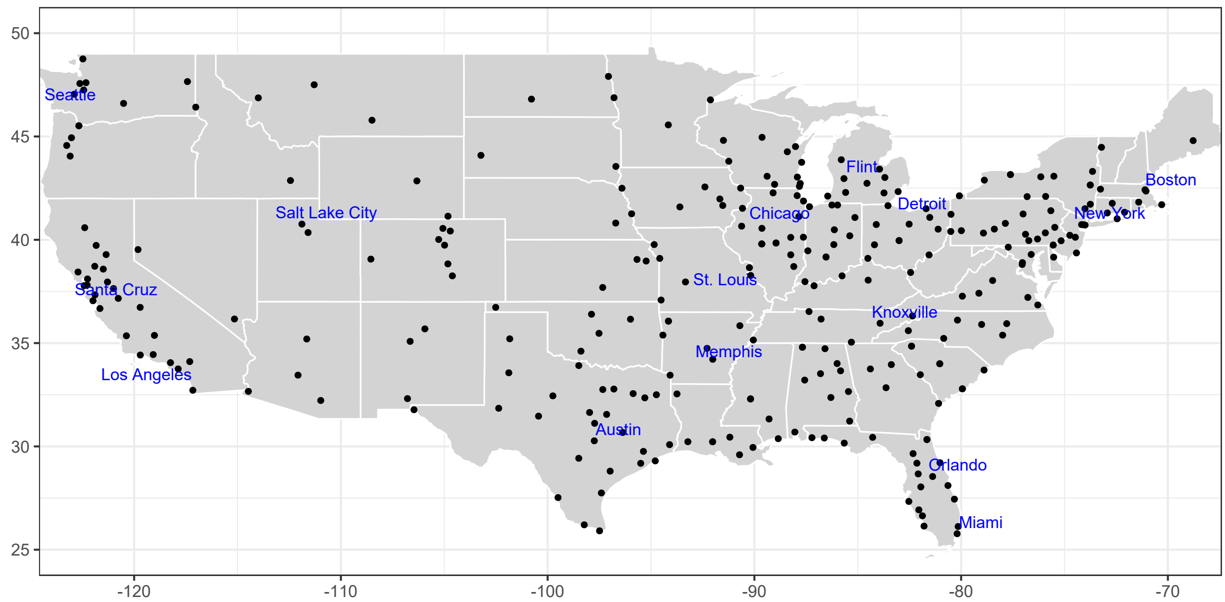
Note: This graph shows 10 year lagged moving averages of the mean difference in log capital gains, log rent returns and log total returns between the city-level and the respective national housing portfolios. The return period covered is 1951 to 2018, such that the moving averages start in 1960, except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

C Within country comparison - Data and further results

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 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 16 we show a map with the location of the MSAs in our sample.

Figure 16: *Geographical distribution of the American MSA sample*



Note: 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).[37] 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.[38]

[37] Unfortunately, the 2010 and 2020 Census did not include questions on housing anymore.

[38] 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

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 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 et al. (2019), we assume that one third of gross rents is spent on these costs.[39] 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.[40] 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 20.

induce a considerable bias.

[39]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.

[40]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 20: *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			

Note: 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).

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 17 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 Germany.[41] 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 construction 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.[42]

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 et al. (2019), we assume that one third of gross rents is used for maintenance and depreciation.[43] For the years prior to 1989 and the missing years in-between,[44] 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

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

[42] 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.

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

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

(*Gemeinden*) from the statistical office of Germany.[45] 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.[46] 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 21.

Table 21: *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			

Note: 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.

45Data is taken from the *Gemeindeverzeichnis* from the *Statistisches Bundesamt*.

46In contrast to counties.

C.3 Additional results

Table 22: *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

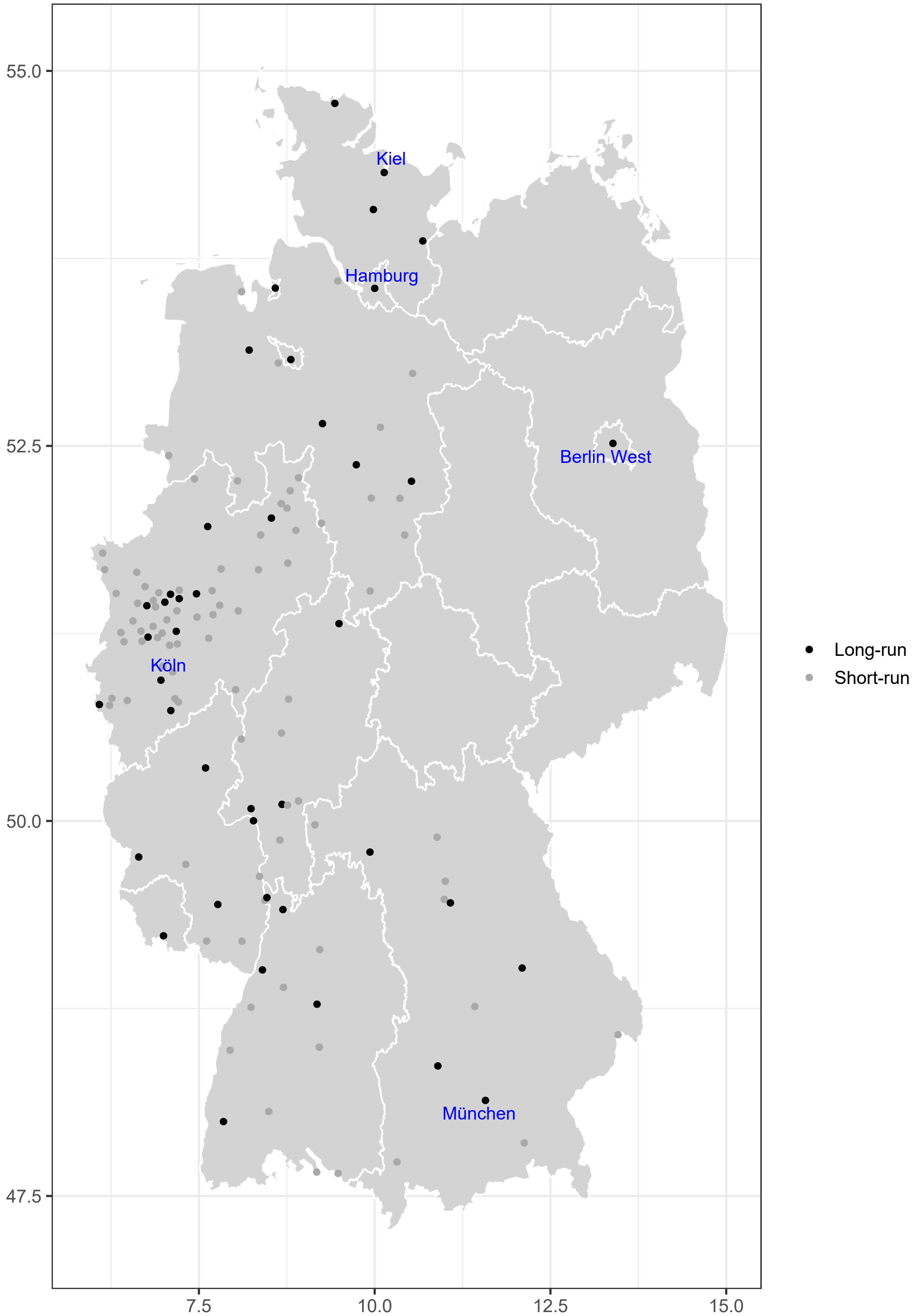
Note: All returns are log returns. Cities are divided into bins based on the size of MSA population in 1950. The middle 8 bins cover size deciles 2 to 9. The 4 extreme bins (1a, 1b, 10a, and 10b) split the smallest and largest deciles in half. As the data for American MSAs only exist in decadal steps,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

Table 23: *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

Note: 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.

Figure 17: *Geographical distribution of the German city sample*



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

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:

$$\text{Total return}_t = \frac{(P_t - P_{t-1})(1 - \tau_t^{capital})}{P_{t-1}} + \frac{R_t^{gross}(1 - \tau_t^{income} - \tau_t^{property})}{P_{t-1}}, \qquad (11)$$

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 superstar 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 2 we used data on net operating income yields from MSCI to benchmark our rent return series following the same procedure as in Jordà 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.

D.1 Rental income & capital gains taxes

From Sections 3.1 and 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.[47] Nevertheless, we can still explore the fact that in the post-World

[47] 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

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 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 et al. (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 24.

Table 24: *Difference in yearly housing returns (log points), 1950-2018*

Sample	Capital gain	Rent return	Total return	N
Not pro homeowner	0.00 (0.40)	-1.07*** (0.06)	-1.05*** (0.40)	859
Medium pro homeowner	0.90*** (0.31)	-1.64*** (0.06)	-0.75** (0.31)	683
Strong pro homeowner	0.84*** (0.26)	-1.73*** (0.06)	-0.89*** (0.26)	840

Note: 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. $*: p < 0.1; **: p < 0.05; ***: p < 0.01$.

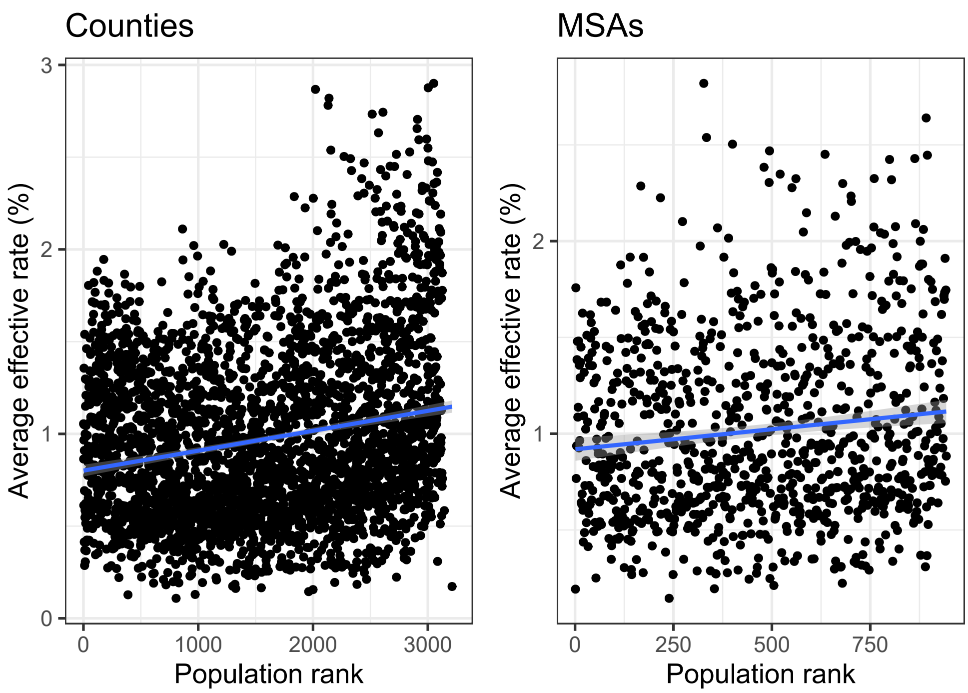
D.2 Property taxes in the US dataset

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.

from capital gains taxes if they have owned the property for more than 10 years.

Figure 18 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 18: *Effective property tax rates (percent) in counties and MSAs, 2010-2014*



Note: 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.

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 et al., 2019), but also in macroeconomic expectations, such as for consumption or investment (Bordalo et al., 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 et al. (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 3.1 have shown that 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 superstar 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 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.[48]

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

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

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 national superstar cities and the rest of the country. If this bias is persistent over time, this could, in turn, explain why house prices in superstar cities are elevated and, consequently, housing returns are smaller than in other cities as observed in the data.[49]

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 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:[50]

$$w^P * \overline{cap\ gain}^A + \overline{rent\ return}^A = w^P * \overline{cap\ gain}^B + \overline{rent\ return}^B \tag{12}$$

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.[51] 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 12 to hold in equilibrium over our long-run data. In the comparison between national superstars and national housing portfolios in section 3.1, the resulting weight on capital gains would approximately need to be 2.35.[52] This implies that home-buyers would need to attach more than double the weight (or

[49]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.

[50]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.

[51]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.

[52]To calculate the weight on capital gains we first transform the log returns from Table 3 into percentage returns, because log returns do not aggregate linearly across return components. By assuming that capital

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 big deals among institutional investors in Europe in the 2010s (see appendix 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.

F Corelogic Deed Dataset

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:

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.

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.

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 5, 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.[53]. 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:

$$LME_t = \frac{P^{loc}_{i,t_i} - P_{i,t_i}}{P_{i,t_i}} \tag{13}$$

$$P^{loc}_{i,t_i} = \frac{P_{i,T_i}}{R^{loc}_{t_i,T_i}} \tag{14}$$

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

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

into population size groups for the period between 1990 and 2020. The results can be seen in 25. 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 25: *Log LME across MSAs size bins, 1990-2020*

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Holding Period	42.33	40.28	39.14	41.53	34.45	35.52	33.92	32.33	31.96	30.15	30.46	29.77
Annual	15.23	14.40	14.28	14.93	12.75	12.84	12.66	12.04	11.90	11.28	11.44	11.63
N	13.00	12.00	25.00	25.00	25.00	24.00	25.00	25.00	25.00	25.00	12.00	12.00

Note: 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 ($l\tilde{m}e_i$), which is explained in the text.

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:

$$l\tilde{m}e_i = \alpha_{s,y} + \alpha_{e,y} + \alpha_{s,m} + \alpha_{e,m} \tag{15}$$

$$+\alpha_{zip} + \beta_P log(P_{i,t_i}) + BX_i + u_i \tag{16}$$

Where $l\tilde{m}e_i = \frac{lme_i}{\sqrt{hp_i}}$ 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 (between 2 and 5 years, between 6 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 26. 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 26: *Idiosyncratic risk across MSAs, 1990-2020*

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Holding Period	29.78	30.17	28.54	28.82	26.01	25.52	25.38	23.45	22.94	22.48	22.10	21.75
Annual	11.01	10.88	10.46	10.61	9.66	9.34	9.48	8.73	8.53	8.45	8.22	8.35
N	13.00	12.00	25.00	25.00	25.00	24.00	25.00	25.00	25.00	25.00	12.00	12.00

Note: 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.

H Housing risk distribution

H.1 Co-variance risk across the distribution

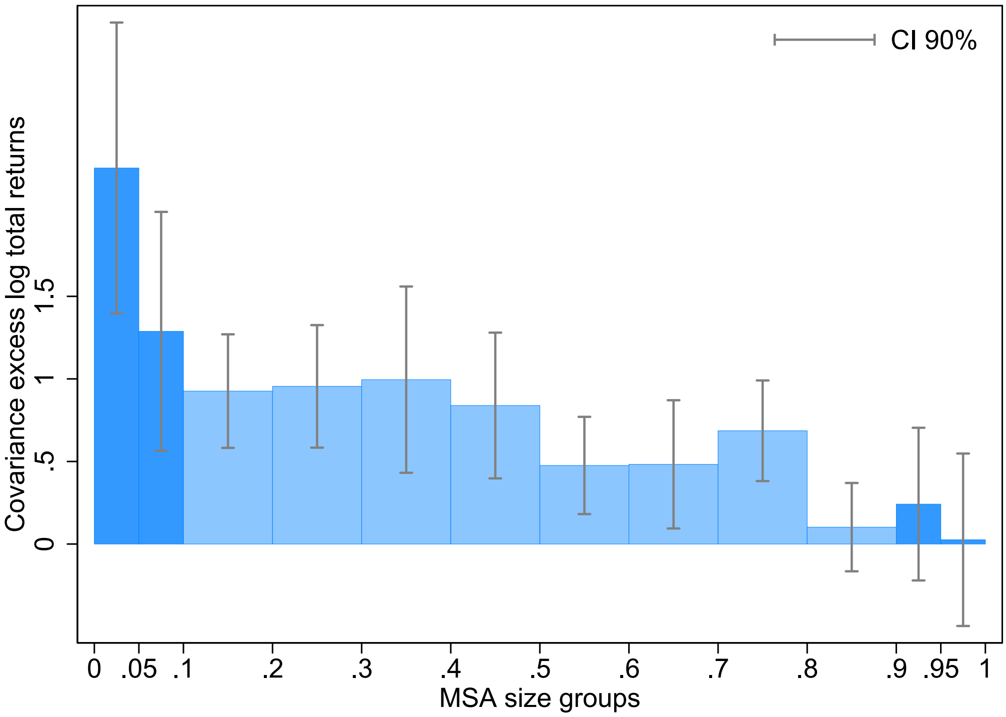
In this subsection of the appendix we show that the co-variance between excess housing returns and income growth decreases almost monotonically across the city-size distribution. In Figure 19, we plot the average co-variance between excess housing returns and income growth by MSA-size group for the period 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.

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

Figure 19: *Co-variance between log excess total housing returns and log income growth by MSA size, 1950-2018*



Note: 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.

market and because being a homeowner implies that a large share of their asset portfolio has a local component.[54] 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 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.

[54]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).

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.[55] We then test whether income betas are smaller in large MSAs. The results are depicted in Table 27 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 the largest MSAs to only the smallest ones. Appendix H.2 shows results for the entire distribution of MSAs.

Table 27: *Differences in income betas by city size, US, 1950-2018*

Sample	Capital gain	Rental yield	Total return	N
Large vs rest	-0.26** (0.106)	-0.25*** (0.040)	-0.32*** (0.096)	316
Large vs small	-0.70*** (0.252)	-0.43*** (0.089)	-0.76*** (0.233)	31

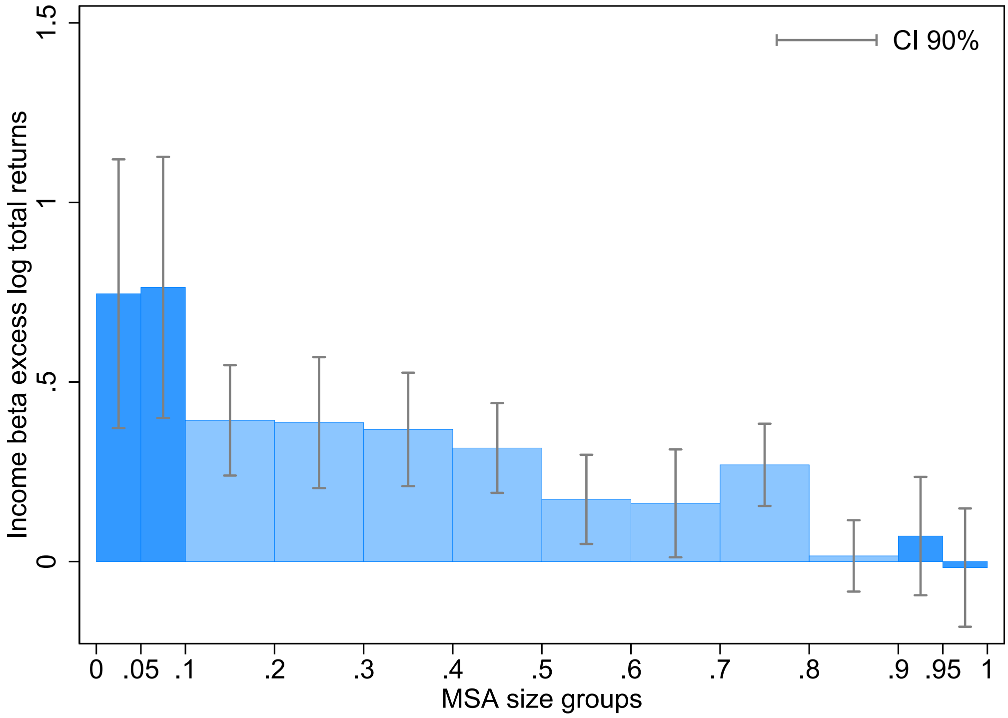
Note: The table shows differences in income betas for log excess total returns, log excess capital gains and log excess rental yields between large MSAs and the rest of the sample or small MSAs. Differences are measured as coefficients in a cross-sectional regression of the dependent variable (income 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 27 columns 1 and 2, which also show that betas for both components are smaller in the largest cities.

Figure 20 plots income betas for total housing returns by MSA-size bins. Section 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 28 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.

Figure 20: *Income betas on log excess total housing returns by MSA size, 1950-2018*



Note: 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.

Table 28: *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

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

Note: 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. $* : p < 0.1; ** : p < 0.05; *** : p < 0.01$.

Table 29: *Total housing risk and its decomposition by MSA size, 1990-2020*

	1A	1B	2	3	4	5	6	7	8	9	10A	10B
Local risk	4.91	5.03	5.70	5.15	5.01	5.25	5.46	5.68	5.69	5.09	6.58	6.63
Idiosyncratic risk	11.01	10.88	10.46	10.61	9.66	9.34	9.48	8.73	8.53	8.45	8.22	8.35
Share of idios. risk	0.79	0.79	0.73	0.78	0.75	0.74	0.74	0.71	0.68	0.72	0.60	0.61
Total risk	12.34	12.19	12.32	12.01	11.27	11.03	11.36	10.86	10.57	10.12	10.87	10.87
# Repeat sales	113587	126964	250780	312606	399287	502893	674340	969189	1416480	2718650	2867698	4281910
# MSAs	13	12	25	25	25	24	25	25	25	25	12	12

Note: 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.

H.3 House price risk distribution

Table 29 shows annual total house price risk and its decomposition across the MSA-size distribution for the period between 1990 and 2020. Following Giacoletti (2021), we define total house price risk 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,[56] but can be explained by the observation that large urban centers tend to have tighter housing supply constraints,[57] which amplify shocks to house prices leading to higher house price index volatility.[58] 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 29 reveals that total risk also decreases with MSA-size. While the smallest MSAs had on average an annual total house price risk of 12.34% of the sales price of a house between 1990 and 2020, the largest MSAs had a considerably lower total risk of 10.87% relative to the sales price.

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

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

[57]See, for example, Saiz (2010)

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

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 covariance between rental yields and capital gains at the property level attenuates capital gain volatility, but only to a small extent.[59]

The other source of rental yield volatility are changes in the rental income of a property. As done in section 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 remainder of this section we show empirical evidence that suggests that, if anything, both components of rental income risk are lower in large cities.

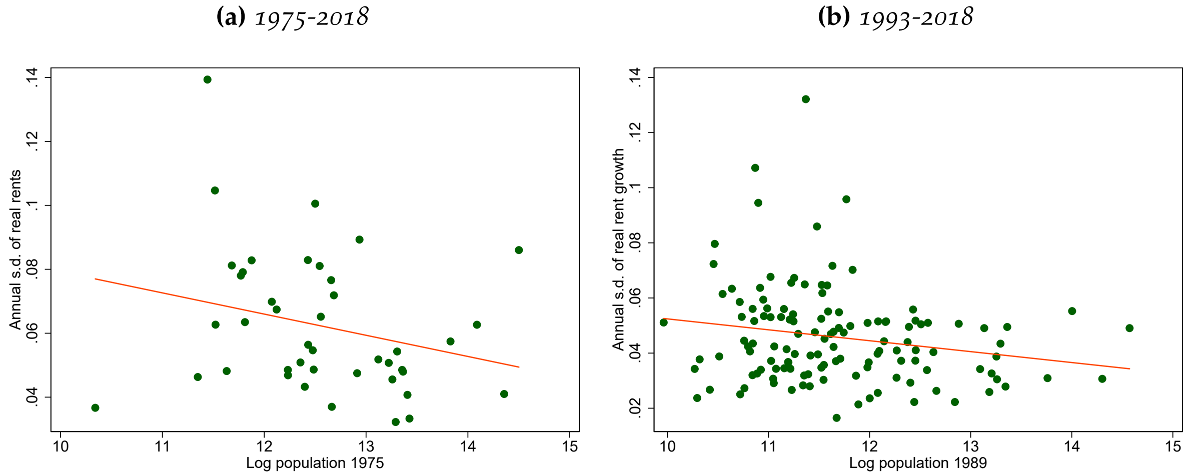
First, we analyze location-wide rent risk. Unfortunately, there does not exist a data set with long-run annual rent data on city- or MSA-level for the U.S. However, the German data set we constructed and use in section 4.2 does feature rent indices for a large cross-section of German cities.[60] We use these data to calculate location-wide rent volatility on city level. Figure 21 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.

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 30. It shows that the mean as well as the standard deviation of annual

[59]Eichholtz et al. (2020) also find a negative covariance of rental yields and capital gains empirically.

[60]For details on this data set please refer to Appendix C.2.

Figure 21: *Real rent growth volatility and population, Germany*



Note: 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 4.2 of the paper and Appendix C.2.

rental vacancies is lower in large cities.

Table 30: *Differences in mean and standard deviation of rental vacancies in p.p., US, 1985-2020*

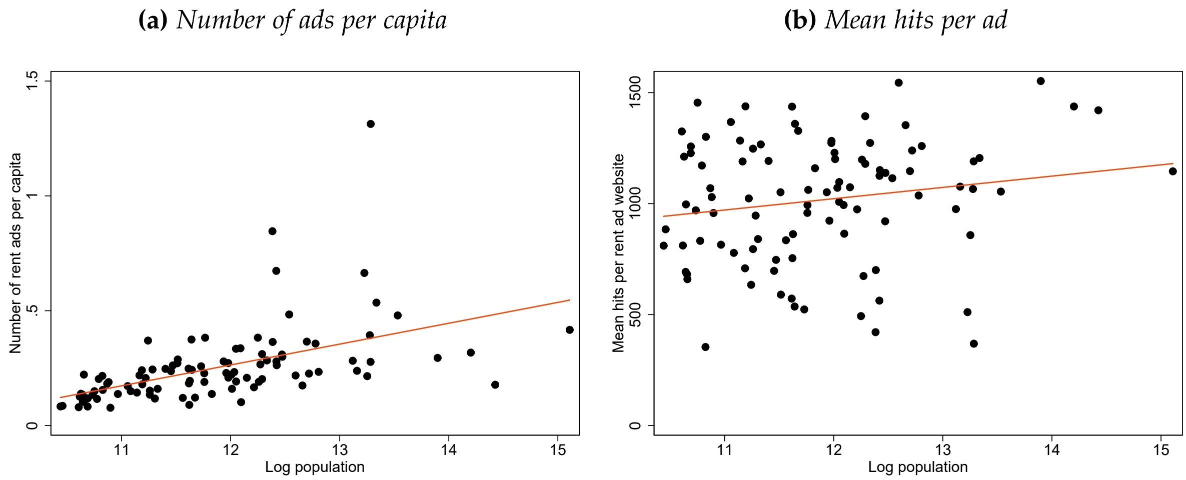
Sample	Mean	N	S.d.	N
Large vs rest	-2.06*(1.093)	1372	-0.73***(0.169)	1372
Large vs small	-1.25 (1.415)	168	-1.06***(0.274)	168

Note: 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.

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 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 22 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.

Figure 22: *Thickness of the rental market by city size, Germany*



Note: 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).

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.

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 5, 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.[61] This data is already used to extend the US data set used in section 4.1 and described in more detail in Appendix C.1. Additionally, we also use data on the real returns of US T-bills from Jordà et al. (2019) to construct excess total real housing returns. We calculate Sharpe ratios as:

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

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 31: *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	0.45	0.37	0.40	0.42	0.44	0.40	0.40	0.43	0.36	0.45	0.41	0.39
Total excess return	4.99	4.39	4.71	4.64	4.42	4.28	4.33	4.39	3.70	4.46	4.25	4.09
Total risk	12.34	12.19	12.35	12.02	11.26	11.05	11.35	10.85	10.57	10.13	10.79	10.62
Number of MSAs	13	12	25	25	25	24	25	25	25	25	12	12

Note: 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 31.[62] 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 32. This is in line with the theoretical prediction discussed above that higher housing returns in smaller cities are a compensation for higher housing risk.

[61]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.

[62]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 4.1. Still, the relation between housing returns and city size also shows up for the smaller sample and shorter time period.

Table 32: *Differences in Sharpe ratio, 1990-2020*

Sample	Sharpe Ratio	N
Large vs rest	-0.03(0.030)	248
Large vs small	-0.06(0.053)	25

Note: 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 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 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.

J Additional results on housing liquidity

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

Table 33: ***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

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 34: ***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

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.

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 *immobilien-scout24.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 23. 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 23 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.

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.[63]

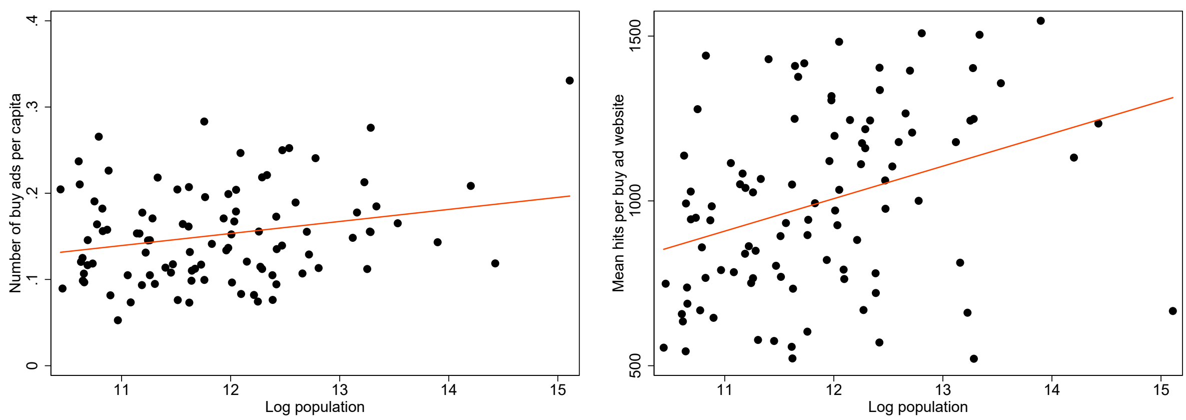
J.3 Real estate liquidity of institutional portfolios in European cities

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.

[63] 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.

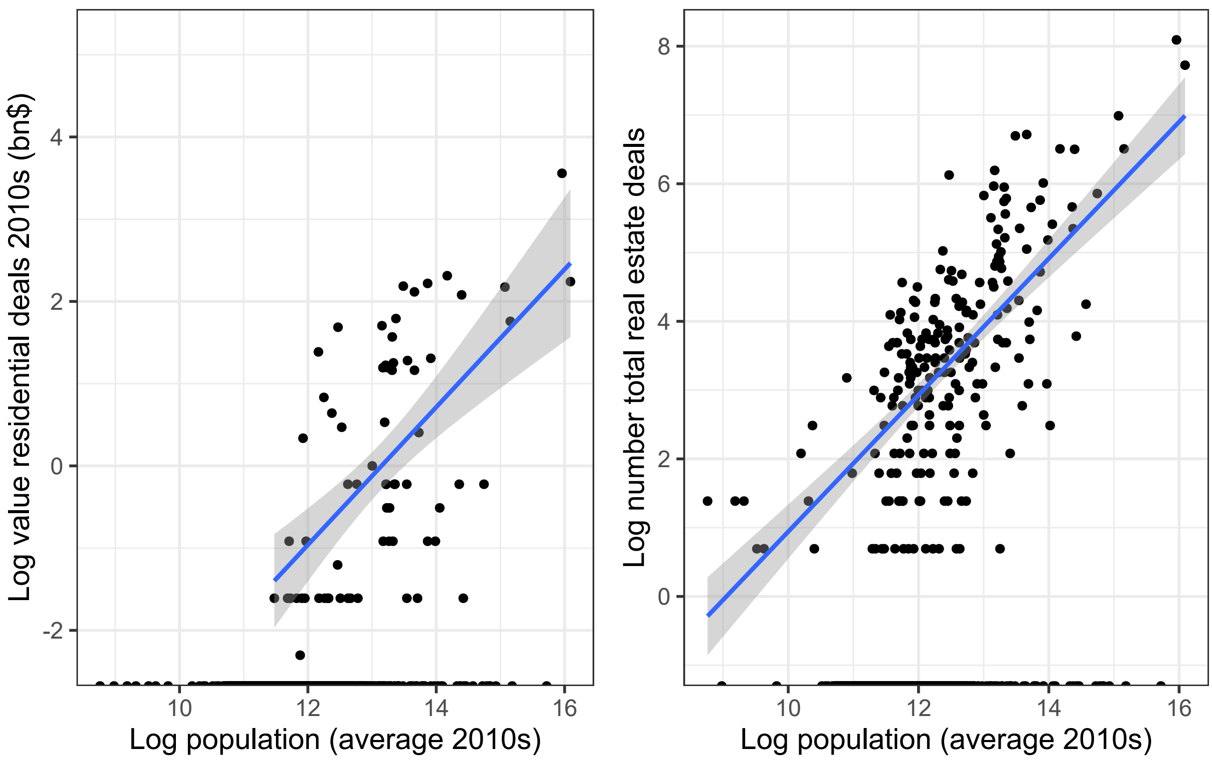
Figure 23: *Thickness of the housing market by city size, Germany*

(a) *Number of ads per capita* **(b)** *Mean hits per ad*



Note: 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).

Figure 24: *Liquidity of housing markets in European cities*



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