Housing Yields*

Stefano Colonnello[†] Roberto Marfè[‡] Qizhou Xiong[§]

This Draft: January 27, 2022

First Draft: February 16, 2021

Abstract

This paper investigates heterogeneity in residential property yields using rental and sale listings from a major German online real estate platform between 2007 and 2017. Equipped with property-level rent-to-price ratios obtained by matching properties for sale and for rent, we show that these yields strongly co-move with regional factors, such as population age structure, industry structure, housing supply rigidities, and the liquidity and size of the housing market. Differences are particularly pronounced between globally relevant cities and other areas. Despite the importance of regional factors, the degree of unexplained heterogeneity in yields is puzzlingly high relative to equity yields, whose variation can be largely understood through a few systematic factors. Specifically, a substantial fraction of housing yields heterogeneity is explained neither by local factors nor by an extensive array of property-specific observable features, possibly pointing to the crucial role of idiosyncratic factors, within-city aggregation effects, as well as of informational and regulatory frictions.

JEL Classification: G12, G51, R31

Keywords: Housing, Rent-to-Price Ratio, Geographic Heterogeneity

---

*We thank Ufuk Akcigit, Dmitry Kuvshinov (discussant), Francesco Lippi, Yiran Mao (discussant), Juan S. Morales, Loriana Pelizzon, Alberto Plazzi, Geert Rouwenhorst, Martin Simmler, Viktor Slavtchev, Gregor von Schweinitz, Michael Ungeheuer (discussant), Mungo Wilson, and seminar participants at Collegio Carlo Alberto, the 2021 European Meeting of the Urban Economics Association, the 2021 Annual Meeting of the German Finance Association, and the 2021 Paris December Finance Meeting for insightful comments. We are grateful to staff at FDZ Ruhr of RWI Essen for granting us access to the RWI-GEO-RED database.

[†]Ca' Foscari University of Venice and Halle Institute for Economic Research (IWH). E-mail: stefano.colonnello@unive.it.

[‡]University of Turin and Collegio Carlo Alberto. E-mail: roberto.marfe@carloalberto.org.

[§]Corresponding author. Saïd Business School, University of Oxford, Park End Street, OX1 1HP, United Kingdom. Phone: +44 7838606686. E-mail: qizhou.xiong@sbs.ox.ac.uk.

1 Introduction

The main residence is the largest component of wealth for many households (e.g., Flavin and Yamashita, 2002). Also in Germany, the country we focus on, despite a relatively low home ownership rate of 44% as of 2017 (vs. 64% in the US), real estate assets dominate the portfolio of the average household by a wide margin (Deutsche Bundesbank, 2019).[1] Nonetheless, the traditional view based on the permanent income hypothesis places little emphasis on the consequences of house price fluctuations for aggregate consumption and, in turn, the business cycle. Once the assumption of complete markets—i.e., of households' insurance against idiosyncratic shocks—is relaxed, changes in housing wealth become important (Berger, Guerrieri, Lorenzoni, and Vavra, 2018). In line with this conjecture, responses of consumption to local shocks to house prices are substantial in the US, with relevant consequences for the amplification of business cycles (Mian, Rao, and Sufi, 2013) and the effectiveness of monetary policy (Beraja, Fuster, Hurst, and Vavra, 2019).[2]

Understanding the drivers of house valuations is thus key to designing credible macroeconomic models. We focus on the rent-to-price ratios—the *housing yields*—, which incorporate market participants' expectations about properties' future discount and rent growth rates (Campbell, Davis, Gallin, and Martin, 2009; Plazzi, Torous, and Valkanov, 2010).[3] Our main contribution is to study the distribution and determinants of housing yields over a recent, highly granular, and comprehensive database on the market for residential properties of a large economy like Germany. We document a novel *heterogeneity puzzle*: a substantial degree of heterogeneity in housing yields can be explained neither by zipcode-level time-varying factors nor by a rich set of property-level characteristics. Such heterogeneity is economically sizable, with the the 90th−10th percentile range of unexplained yields in our dataset corresponding to a variation of EUR 60,778 in the value of the median flat, against a mean household net wealth estimated at EUR 202,541 according to Deutsche Bundesbank (2019). Hence, a substantial fraction of dis-

[1] Home ownership in Germany is especially low in (Eastern) urban areas; yet, private landlords own around two thirds of rental properties (Savills, 2019). Kaas, Kocharkov, Preugschat, and Siassi (2021) investigate the drivers of German low home ownership (which is coupled with a high house ownership for investment purposes), pointing to a high property transfer tax rate, tax deductions of mortgage interest payments available to landlords but not to owner-occupiers, and the accessibility of social housing.

[2] Guerrieri and Mendicino (2018) show that the effect of housing wealth changes on consumption is more modest for European countries, but is persistent and stronger in the long-run than in the short-run.

[3] The housing yield is a slow-moving variable, constituting a key metric to characterize the state of local housing markets, over and above the dividend-to-price ratio for stocks. Indeed, as pointed out by Plazzi et al. (2010), both the property price and the rent are observed market prices. By contrast, dividends also reflect to a large extent short-term managerial decisions (Vuolteenaho, 2002).

persion of the household wealth distribution should be imputed to unobservable property- or household-specific traits, within-zipcode agglomeration effects, and regulatory and/or informational frictions.

To examine rent-to-price ratios, we use sale and rental prices for flats listed on a major German online real estate platform between 2007 and 2017. One challenge is that we generally observe the market price of a property either as a sale price or as a rental price. To work around this problem, we build a synthetic measure of the property-level rent-to-price ratio, which relies on matching each rental property to a counterfactual property for sale based on a comprehensive set of observable property traits. To the best of our knowledge, we are the first to apply this matching approach on such a large-scale database.[4] The rent-to-price ratios so obtained vary greatly across geographic areas and their dispersion is remarkable not only across states or districts, but even across zipcodes within the same city.

Given the high degree of regional segregation of the housing market, we then investigate local economic and social conditions as plausible determinants of rent-to-price ratio variation. We find that district-level demographics, industry and economic fundamentals, rigidities in housing supply, and liquidity and size of the housing market explain a substantial fraction of variation. Differences in rent-to-price ratios across groups of districts split along selected dimensions (like their population age structure, income per capita, housing supply, and size) can be only marginally explained by disparities in observable traits of the housing stock across districts.

Cross-sectional dispersion in housing yields is remarkably stable with respect to the level of regional aggregation considered (federal state-, district-, or zipcode-level). Even after controlling for an extensive set of observable property-level characteristics and fine fixed effects at the zipcode-calendar quarter level, unexplained heterogeneity in yields remains large. We verify that potential matching errors inherent to our synthetic yields appear unlikely to drive their variation. Such heterogeneity is puzzling because residential properties offer a relatively homogeneous service to households, thus, once filtering out obvious differences in key dwelling traits (e.g., size and number of rooms, presence of a balcony, quality of facilities, etc.) and any time-varying zipcode trait (e.g., distance from schools or hospitals, quality of local services, number of nearby shops, etc.), one may expect that rent-to-price ratios exhibit little variation.[5] Besides pure idiosyncratic shocks

[4] A similar matching approach has been previously used over more limited datasets. Smith and Smith (2006) focus on ten US metropolitan statistical areas (MSAs) in 2005, whereas Bracke (2015) on the London area between 2006 and 2013.

[5] Moreover, we document that the total variation of housing yields is comparable to, if not smaller

hitting households, possible explanations for the "excess" heterogeneity in rent-to-price ratios, relate to informational frictions, regulatory restrictions affecting disproportionately certain properties (e.g., rent leveling), and local agglomeration effects operating within or across zipcodes.

Building on the US evidence of distinctive housing market dynamics in large metropolitan areas (e.g., Van Nieuwerburgh and Weill, 2010; Gyourko, Mayer, and Sinai, 2013), we explore whether and how house prices stand out in globally relevant cities relative to other areas of Germany. In line with the existence of pervasive agglomeration economies, median housing valuations are noticeably higher in such cities and this is largely unrelated to observable property characteristics. And the gap is rising: properties located in globally relevant cities substantially outperform those located elsewhere in terms of cumulative returns over 2007-2017.

The baseline analysis builds on a micro-level database that is akin to repeated cross-sectional data. Hence, to cross-validate our yield measure by means of a time series analysis, we finally resort to a pseudo-panel approach by aggregating properties based on their location, number of rooms, and size category. In this way, we are able to obtain quarterly housing returns and rent growth rates, together with rent-to-price ratios. Following a traditional present-value approach to housing valuation, we show that expectations about future discount and rent growth rates incorporated in the housing yields do predict future excess returns and rent growth, in line with the theory and existing evidence (Plazzi et al., 2010). This result further corroborates the reliability of our yields.

This paper contributes to the literature on the pricing of housing assets (for a recent survey on this, see Duca, Muellbauer, and Murphy, 2021). Real estate (especially if residential) has a dual nature: durable consumption good and investment. As a consequence, three different approaches to house pricing are common in the literature, each capturing this peculiarity to a different extent: the hedonic housing price model (e.g., Hill, 2013), the user cost of owning model (e.g., Himmelberg, Mayer, and Sinai, 2005), and the asset pricing analogy (e.g., Case and Shiller, 1989). The hedonic pricing model privileges the durable good nature of housing over its investment features. It assumes that its fundamental price relies on the inherent characteristics of the property, such as the size, the number of rooms, the floor of the building, the neighborhood, and so on. However, it remains silent on what should be its fundamental price and regards house prices as the mere result of the housing market dynamics. The user cost model accounts

than, that of yields for a widely researched asset class like US equities, pointing to the importance of the fine structure of our data and the documented unexplained heterogeneity.

for the opportunity cost of owning a property and posits that, when both the rental and sales market are functioning, the rent should equal the user cost. According to this modeling approach, as soon as the user cost is below rent, rational agents should seek house ownership. We follow the asset pricing analogy that is standard in the finance literature and mainly regard a property as a stock paying dividends periodically in the form of rents.

Despite the lack of a consensus on a specific pricing theory, a growing finance-oriented body of empirical work examines real estate assets. Jordà, Knoll, Kuvshinov, Schularick, and Taylor (2019), in a study of the aggregate rate of return on assets available in the economy, compare housing as an asset class against other forms of investment over a long time span across countries, and find that its country-level returns are akin to those on equities but exhibit lower volatility. By contrast, using data from UK portfolios of real estate investments between 1901 and 1983, Chambers, Spaenjers, and Steiner (2021) provide evidence of much lower long-run returns after adjusting for costs linked to owning properties. Eichholtz, Korevaar, Lindenthal, and Tallec (2021), relying on historical data from Paris and Amsterdam, highlight the primary role of property-level yields in explaining total housing returns.[6] A number of papers applies the present-value relationship approach of Campbell and Shiller (1988) to dissect the role of discount and rent growth rate expectations as captured by the rent-to-price ratio in predicting housing returns. Among others, Campbell et al. (2009) and Plazzi et al. (2010) focus on MSA-level data on residential and commercial properties from the US, respectively. Engsted and Pedersen (2015) extends the analysis to a cross-country setting. Evidence is overall supportive of some degree of predictability. We confirm this finding over a pseudo-panel constructed from property-level German data.

Several asset pricing studies look specifically at the cross-section of real estate properties. Sinai and Souleles (2005) provide insights into the determinants of rent-to-price ratios of residential properties using MSA-level data. They study the rent risk linked to renting a house vs. the asset price risk linked to owning it (which at the same time provides hedging against rent risk) and find that in the presence of volatile rents, rent-to-price ratios tend to be lower because of the higher hedging benefit linked to homeownership. Han (2013) studies how the risk-return relation for residential properties varies across local

[6] Eichholtz et al. (2021) also study housing yields dispersion. Differently from our extensive analysis of the entire housing market in Germany, they focus on time trends in two large cities and find that time-invariant neighborhood fixed effects capture most of the spatial heterogeneity in yields, whereas more detailed demarcation of submarkets or more granular neighborhood-level indexes do not improve the empirical fit.

markets, showing how hedging demand against housing consumption risk and housing supply rigidity can even turn such a relation negative in some US MSAs. Also Chang, Choi, Hong, and Kubik (2017) consider both rent price risk and rent hedging motives, and develop a model with search frictions in the matching process between households and dwellings, showing with MSA-level US data that such frictions on the housing market depress price-to-rent ratios. Using zipcode-level US data on housing returns, Eiling, Giambona, Lopez Aliouchkin, and Tuijp (2019) document that both MSA-level and statewide factors as well as idiosyncratic risk are priced only in about a fifth of MSAs, against the common wisdom of strong segmentation of real estate markets and of households' portfolio under-diversification. Giacoletti (2021) and Sagi (2021) use property-level data to investigate the idiosyncratic component of prices of residential and commercial real estate assets, finding that in neither case it follows a random walk (contrary to what asset pricing models typically assume). Tang, Zeng, and Zhu (2020) ascribe the secular increase in house pricing dispersion in the US to the existence of rational bubbles generated by the non-stationarity of the price-to-rent-ratio.[7] Closer to this paper is Kantak (2019), who examines the relation between an industry-based measure of local expected economic growth and price-to-rent ratios for the US at the MSA-level, documenting a positive link between them in the cross-section. We add to this strand of the literature by studying the distribution and drivers of property-level rent-to-price ratios over a granular and large dataset for Germany, uncovering that a remarkable fraction of their variation—which is economically sizeable—can be explained neither by local factors nor by property-specific observable traits.

Fluctuations and cross-sectional dispersion in house prices attracted significant attention also outside of the asset pricing literature. A line of research points to credit booms as a key driver of price fluctuations in the US (e.g., Mian and Sufi, 2009; Duca, Muellbauer, and Murphy, 2011; Saadi, 2020). This mechanism, however, is to some extent muted for Germany, where the real estate boom over our sample period was not coupled with a credit boom (Bednarek, Te Kaat, Ma, and Rebucci, 2020). More relevant for the German case are probably agglomeration economies (e.g., Combes and Gobillon, 2015), which indeed have been shown to be at work also in Germany, both across cities (Ahlfeldt and Feddersen, 2018) and within cities (Ahlfeldt, Redding, Sturm, and Wolf, 2015). Such agglomeration effects impact the cross-sectional dispersion of house prices.

[7]Focusing on time-series variation, Granziera and Kozicki (2015) incorporate rational bubbles in an asset pricing model to match observed fluctuations in the price-to-rent ratio. Ling, Ooi, and Le (2015) point to the sentiment of market participants as a key determinant of the volatility of housing valuations.

Gyourko et al. (2013) document how "superstar" US cities attract high income individuals because of location preferences, crowding out poorer households and triggering housing booms. Van Nieuwerburgh and Weill (2010) theorize and show empirically that price dispersion comes together with increased wage and productivity dispersion. Howard and Liebersohn (2020) encompass these cross-sectional effects in an asset pricing model able to match also time-series variation in valuations. Ruf (2016), using property listings from online platforms covering the Swiss real estate market, shows that the liquidity of the housing market is increasing in the pervasiveness of agglomeration effects. By exploiting the size and granularity of our data, we complement this body of work by illustrating that a remarkable fraction of the variation in housing yields in Germany could be linked, among other factors, to within-city aggregation economies.

2 Data and Housing Yields Construction

The empirical analysis relies on two main data sources: 1) prices and characteristics of residential properties for sale and for rent, and 2) regional and nationwide economic and social statistics.

2.1 Housing data

Through the RWI-GEO-RED database maintained by the Research Data Center Ruhr (FDZ Ruhr) at RWI Essen (Breidenbach and Schaffner, 2020), we obtain information on prices and characteristics of residential properties for the period January 2007-October 2017 from ImmobilienScout24, a major German real estate listings website. The platform covers about 50% of all real estate properties listed for sale or rent in Germany (an de Meulen, Micheli, and Schaffner, 2014), which guarantees the representativeness of the data provided by the platform. We restrict the analysis to flats excluding detached houses, because of the more standardized nature of the former properties. The raw data contain 16,429,909 listings of flats for rent, and 7,122,908 for sale. The standardization of these properties translates into liquid rental and sale markets, which eases the matching exercise we conduct below to recover synthetic rent-to-price ratios. By contrast, the German rental market for detached houses is thin, which would adversely impact the reliability of the matching procedure below. By focusing on flats, we arguably over-represent urban relative to rural areas in our sample.

Whereas RWI-GEO-RED data come in monthly vintages of listings, instances of flats reappearing on the platform for multiple months are relatively infrequent. We thus narrow down the analysis to listings appearing only once or at their first appearance in the dataset. We then remove observations for which information on any of the following traits is missing: price (for rental or for sale), surface, rooms, bathrooms, bedrooms, floor number, postal code, and district code. We also remove observations with sale (monthly rental) price below EUR 10,000 (EUR 50) and surface below 10 square meters (sqm). And we exclude observations in the top 0.5% of price, surface, number of rooms, bathrooms, bedrooms, and floor number.

It is worth noting that we observe *listed* and not *transaction* prices (an de Meulen et al., 2014). As such, they reflect the supply-side assessment of property value rather than the actual market price, meaning that they are likely to be upward biased. This is an objective limitation of our empirical setting, but two reasons alleviate concerns on the soundness of the analysis. First, the analysis below mostly focuses on rent-to-price ratios, so that the biases in rental and sale prices should to a large extent cancel out. Second, contributors to the platform are generally professional estate agents, which should ensure a degree of rationality in reported prices, being possibly based on the opinion of qualified real estate appraisers.

The rent-to-price ratio of a given property is an inherently unobservable quantity, if we abstract from the relatively few instances of rental properties in those periods in which they are sold from one owner to another. In the remaining cases, we do not simultaneously observe the rental and sale price of the same property. We use two different approaches to compute the ratio.

2.1.1 Matching approach

Transactional data combined with the hedonic housing price model are traditionally used in the literature to provide a regional-level rent-to-price ratio index, but not at the property level (e.g., Campbell, Giglio, and Pathak, 2011).[8]

To overcome this shortcoming, we follow a matching approach (our baseline) to obtain a counterfactual sale price for each flat for rent. Specifically, we adopt a parsimonious set of covariates to this end, namely: flat surface (distance minimization), number of rooms, number of bedrooms, number of bathrooms, floor category, five-digit zipcode, and calendar quarter (exact matching). The floor category indicates whether the flat is in the basement, at ground floor, at floors 1 to 3, or at higher floors. We remove any

[8]One exception is the study by Hill and Syed (2016), who use hedonic imputation to obtain property-level rent-to-price ratios. Though flexible and computationally efficient, such an approach tends to smooth out rent-to-price ratios and leave out of the picture meaningful property-specific variation.

match for which we do not obtain exact matching on each discrete variable or for which the absolute distance in terms of surface is larger than 10 sqm. We take one match, among flats for sale, for each flat for rent. However, because of "ties", around one fifth of rental properties have multiple matches: the average (respectively, maximum) number of matches is 1.63 (resp., 50).[9]

Given each match between the flat for rent and the counterfactual flat for sale, we compute the natural logarithm of the annual rent-to-price ratio (in %)—the quantity whose variation we seek to explain in our main regression analysis—for the "synthetic flat" f as

$$\ln\left(H/P_{f,t}\right) = \ln\left(100 \cdot \frac{12 \cdot H_{r,t}}{P_{\bar{s},t}}\right), \tag{1}$$

where r, s, and t denote the flat for rent, the flat for sale, and the calendar quarter, respectively. Follow the notation of Plazzi et al. (2010), $H_{r,t}$ is the monthly rent exclusive of heating expenses (*Kaltmiete* in German). Such a measure of the rental price is arguably a purer house pricing measure and is more comparable to sale prices, $P_{\bar{s},t}$, than the rent inclusive of expenses, which may reflect the pricing of additional services. In other words, $H_{r,t}$ better approximates the period income the property owner gets from his/her investment. The notation $\bar{s}$ indicates that—in case more than one flat for sale is matched to the flat for rent—we average out their prices.[10]

2.1.2 Pseudo-panel approach

As noted above, the RWI-GEO-RED database is de facto (repeated) cross-sectional in nature. Hence, a non-negligible drawback of the baseline matching approach to recovering rent-to-price ratios is the impossibility to conduct time-series tests. To work around this problem, rather than exploiting the panel nature of those (relatively few) listings that appear more than once in the dataset, for complementary analyses in Section 4 we resort to a pseudo-panel in the spirit of Deaton (1985) by creating cohorts of properties.

To obtain unbiased estimates from a pseudo-panel analysis, the cohorts must be defined using time-invariant attributes that are observed in all periods with all properties. Therefore, we define cohorts with respect to three traits: location of the property (at the district level), its number of rooms, and its surface.[11] Despite having a large number

[9] We remove the few properties with more than 50 matches.

[10] We apply the same notation below to indicate the mean of other characteristics of matched flats.

[11] Whereas the location of a property is unambigously time-invariant, its number of rooms or surface

of observations, the presence of 402 districts in Germany requires us to be parsimonious in the granularity of rooms-surface combinations to ensure that we have sufficient observations in each cohort to achieve statistically robust asymptotics. We, therefore, form relatively coarse categories of flats. We split them in three groups in terms of number of rooms: studio flats with one or two rooms, middle sized flat with two and half or three rooms; and big flats with more than three rooms. Similarly, we discretize the surface of properties in four intervals: small (sqm $\in (0, 50)$), medium (sqm $\in [50, 70)$), large (sqm $\in [70, 90)$), and very large (sqm $\in [90, +\infty)$). Moreover, again to achieve well-sized cohorts, we construct the pseudo-panel at quarterly (rather than monthly) frequency.

In each district-quarter, we can thus have up to 12 ($= 3 \cdot 4$) groups of properties based on our rooms-surface categories. However, not all the combinations of the number of rooms and surface are well-populated enough. For instance, it is extremely rare for a small flat to have more than three rooms (0.02% of the sample). By the same token, a large flat in unlikely to have less then three rooms (1.06 % of the sample). Removing rooms-surface combinations with few observations (accounting for less than 2% of total observations) leaves with at most 8 groups per district-quarter.

However, not only the distribution of observations across rooms-surface combinations is uneven, also the regional distribution is. Sample sizes for large metropolitan areas are large, whereas many of the rural districts do not have a meaningful number of observations even for the most common rooms-surface pairs. Besides the sheer difference in population, another reason is that we focus on flats instead of houses, which naturally tilts sample towards metropolitan areas. To meet the conditions for Type 1 asymptotics (Verbeek, 2008), we therefore disregard any district-rooms-surface-quarter with fewer than 5 properties for rent or for sale. At the same time, we require each district-rooms-surface cohort to have an average of at least 30 properties for rent and 10 properties for sale over the period for which it is in the dataset.[12] After screening the pseudo-panel according to such criteria, we end up with 672 cohorts from 175 districts, each of which we observe for up to 44 quarters.

For each cohort-quarter, we then compute the natural logarithm of the quarterly

may admittedly change following a major renovation. In other words, the consistency of our approach rests on the assumption that major renovations are rare enough events.

[12]The lower threshold for flats for sale reflects their lower frequency in the sample relative to flats for rent.

rent-to-price ratio as

$$\ln\left(H^q/P_{c,t}\right) = \ln\left(\frac{3\cdot\bar{H}_{c,t}}{\bar{P}_{c,t}}\right), \tag{2}$$

where c indicates the cohort of properties. $\bar{H}_{c,t}$ and $\bar{P}_{c,t}$ are the average monthly rent and sale price per sqm in a given cohort-quarter respectively. Though the pseudo-panel only allows us to compute rent-to-price ratios at a less granular level than the matching approach, it makes it possible to investigate their evolution through time.

In the same way, for each property cohort we compute its logarithmic total return between quarter $t-1$ and quarter t:

$$r_{c,t} = \ln\left(\frac{\bar{P}_{c,t} + 3\cdot\bar{H}_{c,t}}{\bar{P}_{c,t-1}}\right), \tag{3}$$

which reflects both property price appreciation and rental income (e.g., Plazzi et al., 2010; Jordà et al., 2019).[13] We then denote the pure price growth component of returns as $r^*_{c,t}$. Analogously, using the same notation as Plazzi et al. (2010), we obtain the quarterly rent growth over the same horizon:

$$\Delta h_{c,t} = \ln\left(\frac{\bar{H}_{c,t}}{\bar{H}_{c,t-1}}\right). \tag{4}$$

Finally, we define the housing premium (i.e., the return in excess of the risk-free rate) as

$$r^e_{c,t} = r^e_{c,t} - r^f_t, \tag{5}$$

where r^f_t is the 3-month interbank rate for Germany.

Our main analysis in Section 3 focuses on the local economic and social determinants of the synthetic rent-to-price ratio obtained through matching. In Section 4, we then use the analogous ratio from the pseudo-panel, first, to validate the baseline findings, and, second, to assess its predictive ability with respect to housing returns and rent growth. Moreover, we evaluate the correlation of total return and rent growth with the growth of consumption per capita.

[13] Unlike Chambers et al. (2021), we do not observe costs incurred by landlords, so we are not able to adjust rental income accordingly.

2.2 Regional and nationwide data

We obtain national and regional economic and social statistics from the German Federal Statistical Office (*Statistisches Bundesamt* and *Statistische Ämter des Bundes und der Länder*) for the period 2007-2017.[14] We retrieve information on registrations of new passenger vehicles—our proxy for local consumption of durable goods—from the German Federal Motor Transport Authority (*Kraftfahrt Bundesamt*).

The housing market is often regarded as highly segregated across regions. We therefore reach the lowest administrative level for which a comprehensive set of economic and social indicators are publicly available in Germany, namely the district-level. German districts are aggregations of municipalities (*Gemeinde*). These districts are akin to US counties and are divided into rural districts (*Landkreis*) and urban districts (*Stadtkreis* or *Kreisfreie Stadt*).

To ensure consistency of regional variables, we account for those instances in which districts changed code over our sample period (e.g., because of statewide reforms such as those of Sachsen-Anhalt in 2007, Sachsen in 2008, and Mecklenburg-Vorpommern in 2011). To merge regional data with housing data, we use the 2015 vintage of district codes provided by the RWI-GEO-RED database for listed properties.

Nationwide data on inflation and interest rates are from from Federal Reserve Economic Data (FRED) of the St. Louis Federal Reserve Bank. Both in the housing and in the regional dataset, all monetary variables are expressed in 2007 euros (EUR). Similarly, all returns and growth rates are expressed in real terms. Moreover, to reduce the impact of outliers, all variables in levels are trimmed at the 99.5%, whereas ratios, returns, and growth rates are trimmed at the 0.5% and 99.5% level.

3 Heterogeneity in Housing Yields

We start by examining the degree of heterogeneity in rent-to-price ratios at the matched property-level. Table 1 reports summary statistics for flat characteristics (Panel A) and district-level variables (Panel B).[15] The final sample in Panel A contains 1,623,237 flats for rent matched to counterfactual flats for sale. By construction, differences between

[14] Regional data are available at annual frequency, at year end. As our main dataset is at quarterly frequency, we assume that regional variables stay constant between the fourth quarter of year y and the fourth quarter of year $y + 1$. More generally, if a regional variable is missing in between two dates, we assume it stays constant until a new non-missing observation is available. Note then that data on local property tax rates are available only up to 2015.

[15] Variable definitions are presented in Appendix Table A.1.

these two groups of flats are statistically indistinguishable from zero for matching covariates. Still, flats for rent are generally significantly different from flats for sale with respect to other covariates, although most of the differences are economically modest. Nonetheless, below we augment rent-to-price ratio specifications with such observable differences to absorb possible systematic patterns in ratios arising artificially from the matching exercise.

With this caveat in mind, Figure 1 visualizes the empirical distribution of the natural logarithm of the rent-to-price ratio obtained from equation (1) together with the national and local macroeconomic conditions over the sample period. In Panel A, we examine the distribution conditional on the size category of the property, uncovering a positive relation between valuations and flat surface. Variation across categories is limited with a median roughly ranging between 5.75% for small flats and 4.5% for very large flats. In Panel B, in which we condition on the federal state where properties are located, both between and within-group variation is more pronounced. States such as Bavaria and Hamburg exhibit substantially lower ratios than Eastern states like Saxony-Anhalt or areas that underwent massive de-industrialization like Saarland. Very intuitively, the economic success of states appears to correlate negatively with rent-to-price ratios. Panel C highlights that the median ratio is typically stable over the sample period, with only a slight increase around the 2008-2009 recession, and displays an increasing trend in within-period heterogeneity.[16] This is broadly consistent with the overall steady growth of economy experienced both nationally and locally in Germany between 2007-2017. Over the sample, the only recession was the 2008-2009 while around the European debt crisis only a slowdown took place (Panel D).

This first, aggregate evidence suggests that variation in rent-to-price ratios is to a large extent cross-sectional and tends to grow over time, which could be driven by a rise in productivity dispersion that emerges as most skilled workers move into large cities (Van Nieuwerburgh and Weill, 2010). Cross-sectional heterogeneity in the housing market relates both to property-specific and regional features. Below we focus on the role of the latter. Federal states display substantial median differences in rent-to-price ratios, but patterns become more and more nuanced as we consider finer geographical subdivisions. Figure 2 documents within-state median disparities in ratios that are more remarkable than those between states (e.g., the Munich area vs. the districts on the Czech border in Bavaria). If we zoom in on the seven "global" German cities—i.e., those with an

[16] Appendix Figure A.1 confirms the negative (resp., positive) trend in the level (resp., dispersion) of rent-to-price ratios.

advanced service sector and that serve as hubs of international transportation networks, where agglomeration economies are most likely to emerge—and consider variation at the zipcode level in Figure 3, we observe that median ratios greatly vary even within some of the most thriving metropolitan areas like Hamburg or Frankfurt.[17]

It is also instructive to compare the heterogeneity of rent-to-price ratios against that of valuation ratios for a well-known asset class like the US equities. To this end, we retrieve information on the quantiles of ratios over the cross-section of NYSE stocks from Kenneth French's website. Table 2 reports selected percentiles for (actual) rent-to-price ratios for German residential properties vs. dividend-, earnings-, and cash flow-to-price ratios for US stocks over 2007-2017.[18] Focusing on the interquartile range and on the spread between the 95th and 5th percentile, rent-to-price ratios exhibit a degree of cross-sectional dispersion in line with stock dividend-to-price ratios, but lower than earnings- and cash flow-to-price ratios. The latter tend to be less subject to managerial discretion—e.g., dividend smoothing policies leading firms to keep dividends low relative to prices to avoid to reduce the risk of having to reduce them subsequently (e.g., Wu, 2018). Whereas within-country location effects matter even for stock valuations (e.g., Garcia and Norli, 2012), the housing market is typically much more geographically segmented.

In the analysis below, we thus seek to explain the regional component of variation in rent-to-price ratios by using measures of local economic and social conditions, providing an upper bound for the role of geographical factors in the pricing of properties.[19]

3.1 Property characteristics

Heterogeneity across geographical areas may not only arise from differences in economic and social development, but also from mere differences in the local housing stock such as unit size, the number rooms, and so on. Moreover, our matching procedure—though imposing exact property matching at the zipcode-level, i.e., highly granular geographic units, especially in densely populated areas—is based on a parsimonious set of covariates,

[17]We identify global cities as those with a rating ranging between "Alpha" and "Gamma" according to the 2020 ranking by the Globalization and World Cities Research Network (see https://www.lboro.ac.uk/gawc/world2020t.html). Other major cities are those with a "Sufficiency" rating in the same ranking. Even starker within-city differences—though at generally lower levels of the rent-to-price ratio—emerge if we look at such cities in Appendix Figure A.2.

[18]As we highlight below, the actual rent-to-price ratio is available for a subsample of properties on sale for which a rental income is reported.

[19]In the remainder of the paper, we mostly focus on the logarithmic transformation of the ratio. Hence, we use interchangeably the expressions "natural logarithm of rent-to-price ratio" and "rent-to-price ratio".

making it possible that some of the variation in our synthetic rent-to-price ratios stems artificially from intrinsic differences between rental flats and matched flats for sale.

Before moving to district-level factors, we thus assess the role of observable property-specific characteristics in explaining equilibrium valuation ratios. We consider all property characteristics observable to us, including those not used to obtain the matched rent-to-price ratio from equation (1):

$$\ln\left(H/P_{f,t}\right) = \boldsymbol{\gamma}\mathbf{m}_{r,\bar{s},t} + \boldsymbol{\eta}\mathbf{x}_{r,t} + \boldsymbol{\theta}\mathbf{x}_{\bar{s},t} + \boldsymbol{\zeta}\mathbf{z}_{r,\bar{s},t} + \tau_t + \epsilon_{f,t}. \tag{6}$$

$\mathbf{m}_{r,\bar{s}}$ is the vector of covariates on which we match, either by minimizing distance (surface) or exactly (number of rooms, number of bedrooms, number of bathrooms, floor), with the addition of squared surface.[20] $\mathbf{x}_r$ contains covariates available only for rental flats (flat expenses, heating expenses, an indicator for rent inclusive of heating expenses, deposit). $\mathbf{x}_{\bar{s}}$ includes covariates available only for sale flats (housing benefits, an indicator for holiday properties, an indicator for rented out properties). $\mathbf{z}_{r,\bar{s}}$ is a set of distances between flat r and synthetic flat $\bar{s}$ for characteristics on which we do not match (number of floors in the building, energy source, flat expenses, etc.): $\mathbf{z}_{r,\bar{s}} = \mathbf{w}_r - \mathbf{w}_{\bar{s}}$. If a flat trait is missing, we set it to 0. To mitigate the bias potentially arising from this adjustment, for any incompletely reported variable, we include a corresponding missing value indicator. To absorb variation in nationwide macroeconomic conditions, we control for calendar quarter fixed effects τ_t.[21] The standard errors are clustered at the district level.

Table 3 presents coefficient estimates from specification (6). Column 1 includes the covariates on which we match properties. Total surface of the flat is significantly and negatively correlated with the rent-to-price ratio across all specifications, which is consistent with the evidence in Panel A of Figure 1. The correlation of the rent-to-price ratio with the number of rooms is positive. Since we control for the size of the property, this means that conditional on having similar size, the properties with a higher number of rooms tend to be valued less. To put this into perspective, a 40 sqm two-room flat has on average a lower rent-to-price ratio than a 40 sqm one-room studio. Interestingly, bedrooms and bathrooms exhibit a negative association with rent-to-price ratios, possibly pointing to a value-decreasing role of other types of rooms. The floor on which the flat is located does not load significantly.

Columns 2 and 3 introduce variables specific to rental and sale listings, respectively.

[20] We take the average between the surface of the rented flat and matched flats for sale.

[21] Regressions models throughout this section are estimated via the Stata package REGHDFE by Correia (2018).

Rental contracts requiring a higher deposit or higher heating expenses, and holiday properties come with lower rent-to-price ratios, possibly reflecting the negative correlation between ratios and flat size. Properties for sale that are already rented are valued significantly less: these properties, for instance, may have not undergone modernization for a longer time, may be occupied by a defaulting tenant, or the owner may be forced to sell the property at fire-sale price.[22]

The correlation patterns described so far remain robust once we include all the remaining observable flat characteristics in column 4, with adjusted R^2 raising by 18% to 31% (from 13% in column 3). In other words, heterogeneity in the rent-to-price ratio is unlikely to be uniquely a by-product of observable flat traits and of systematic matching errors. Below we explore several plausible channels through which district-level factors may be factored in house prices.

3.2 The role of regional factors

One of the most prominent features setting the housing market apart from other financial asset markets is its pronounced geographic segmentation. Therefore, a substantial fraction of cross-sectional heterogeneity in rent-to-price ratios may be explained by regional differences in factors such as age structure of the population, unemployment, income per capita, and the like.

The regional economic and social environment indeed feeds in households' expectations about discount rates and rent growth. Such expectations shaping the local rent-to-price ratio are inherently unobservable. But, provided that social and economic conditions in a given area are persistent, market participants can rely on current information on local factors to form rational expectations. Put differently, we use the historical record of social and economic indicators at the district level—the most granular administrative subdivision at which a comprehensive set of statistics is publicly available—to study their impact on rent-to-price ratio regional heterogeneity.

To this end, we estimate the following regression of rent-to-price ratios on regional

[22] We validate our baseline rent-to-price ratios against actual ones for properties for sale that are already rented. Indeed, for such properties—which constitute 21.9% of the sample (see Panel A of Table 1)—we do not only observe the sale price but also their rental income. This is admittedly a special sample, thus we do not use it for the main analysis, but it provides a valuable benchmark. The correlation between matched and actual logarithmic ratios is 63.5%. Similarly, the graphical inspection of the two measures in Appendix Figure A.3 supports the validity of the matching procedure.

factors:

$$\ln\left(H/P_{f,t}\right) = \boldsymbol{\nu}\mathbf{p}_{d,t} + \boldsymbol{\gamma}\mathbf{m}_{r,\bar{s},t} + \boldsymbol{\eta}\mathbf{x}_{r,t} + \boldsymbol{\theta}\mathbf{x}_{\bar{s},t} + \boldsymbol{\zeta}\mathbf{z}_{r,\bar{s}} + \tau_t + \epsilon_{f,t}, \tag{7}$$

where d denotes the district of the flat. The variables of interest in this step of the analysis are contained in $\mathbf{p}_d$, which is a vector of district-level covariates.

This specification nests the most saturated model of Table 3. In this way, we focus on the component of rent-to-price ratio variation that relates neither to the observable traits of the flat for rent r nor to those of the synthetic counterfactual flat for sale $\bar{s}$, and is therefore (at least partially) attributable to different conditions across regions.[23]

We group the regional factors in $\mathbf{p}_d$ in the two categories: 1) demographic and economic fundamentals, and 2) local housing market characteristics.

3.2.1 Demographic and economic fundamentals

In Table 4, we verify if housing yields associate with district-level demographic and economic fundamentals via regressions specified as in equation (7). Both standard overlapping generations models and existing evidence on the stock dividend-to-price ratio indicate that demographics, especially the age profile of the population, correlate with the valuation of assets (e.g., Geanakoplos, Magill, and Quinzii, 2004; Favero, Gozluklu, and Tamoni, 2011; Poterba, Weil, and Shiller, 1991). In column 1, we look at the age profile of the district, which is likely to capture slow-moving long-term expectations about the housing market due to life-cycle portfolio effects (Favero et al., 2011). The ratio of elderly dependent (above 65 years old) to working age population loads positively and significantly on rent-to-price ratios. This points to positive housing valuation effects of people in the prime of their careers coupled with a depressing effect of the elderly relative to active population (Takáts, 2012). At the same time, in column 2 we show that the district's total population correlates negatively with valuation ratios. This is arguably capturing a mere size effect, which could be driven by agglomeration economies (see Combes and Gobillon, 2015, and references therein).

Structural features of the local economy, such as disposable income, unemployment, and industry composition may also correlate with housing valuation ratios, both through expected rent growth and expected discount rates. For instance, income per capita ought

[23]The remaining rent-to-price variation may still reflect unobservable differences between the flats in each match. But estimates of $\boldsymbol{\nu}$ from equation (7) are less likely to suffer from this problem than those from a simple regression of the matched rent-to-price ratio on $\mathbf{p}_d$ alone.

to co-move with house prices in the long-run equilibrium (e.g., Abraham and Hendershott, 1996). The coefficient estimates in columns 3 and 4 appear to confirm this intuition: rent-to-price ratios are decreasing in both district-level disposable income per capita and GDP per capita.

A long-standing theoretical and empirical literature has uncovered rich interactions between local housing and labor markets (e.g., Cameron and Muellbauer, 2001; Branch, Petrosky-Nadeau, and Rocheteau, 2016; Zabel, 2012). In this spirit, in column 5 we find that districts with more unemployment display significantly higher rent-to-price ratios, consistently with Vermeulen and Van Ommeren (2009), who suggest that cheaper housing may compensate for lower wages in such regions and explain the persistence of across-region heterogeneity in unemployment rates.

In column 6, we observe that a higher importance of the manufacturing sector is associated with higher rent-to-price ratios. A possible story is that regions more reliant on traditional—eventually declining—industries exhibit a more pervasive presence of displaced workers (Case and Mayer, 1996), which is in line with the previous finding on unemployment.[24] The sheer number of registered business in the district correlates instead negatively with rent-to-price ratios. This correlation, like in the case of total population, is probably a manifestation of a size effect and is consistent with higher demand for dwellings (and by more highly paid workers) in metropolitan areas. Below, we investigate this point more in depth by studying the housing market of German global cities against that of peripheral regions.

3.2.2 Housing market characteristics

Substantial transaction costs in the housing market are prevalent and could lead to a wide no arbitrage interval. The wider the no arbitrage interval, the higher the price heterogeneity we will observe in the market. Thus, the liquidity of the local housing market is likely to be a non-negligible determinant of rent-to-price ratios (e.g. Krainer, 2001). Besides transaction costs, housing valuations should also hinge on the rigidity of local supply. Indeed, whereas standard finance theory relies on the assumption of stocks being close substitutes to each other, so that their demand curves are horizontal and single stocks' supply does not affect their price (Shleifer, 1986), a real estate property is not just a claim on a firm's cash flows, but offers a service to the (homeowner) investor. The

[24]The relation between house prices and industry structure can go both ways. For instance, Adelino, Schoar, and Severino (2015) show how increasing house prices favors self-employment by providing capital to new ventures through the collateral channel.

consumption value of residential properties makes their market akin to service markets, which are heavily driven by the demand and supply dynamics, and not just by investment motives. Property taxes may also affect valuations as a specific form of transaction costs (Poterba et al., 1991). For example, in the user cost model, the rent-to-price ratio is a function of the property tax rate (e.g., Hill and Syed, 2016).

In Table 5, we look at the relation of several measures capturing local housing market liquidity, supply rigidities, and property taxes. Columns 1 to 3 consider three different measures of liquidity. The first one captures the standardization of the properties listed on the online platform in a given calendar quarter: the district-level interquartile range of flat surface. The second measure is the average number of days a property listing stays online.[25] The third measure is number of flats posted on the online platform in a given district-calendar quarter, which also captures the size of the market. In line with intuition, the first and the third measure correlate negatively with rent-to-price ratios, the second positively. To put it differently, housing valuations are increasing in the liquidity of the local housing market, which in turn tends to be higher in growing urban areas.

Moving to supply factors, in columns 4 to 6 we study the role of rigidities in the provision of dwellings, which depend both on zoning laws and the geographical conformation of the area (e.g., Pogodzinski and Sass, 1991; Harari, 2020). Real estate valuations correlate negatively with the housing stock per capita, and positively with the price of construction land. Rent-to-price ratios, less intuitively, increase with completed living space per capita in a given year. The rationale for such a correlation may stem from reverse causality: property developers respond to higher demand (and prices) in areas where housing is in short supply by building more dwellings.

Finally, column 7 investigates whether property taxes feed into rent-to-price ratios. In Germany, a property tax is levied on all the land used for residential buildings, whereas the transaction tax is levied on traded properties. We focus on the former, for which each district's government can then apply a multiplier to the federal property assessment rate. Surprisingly, such rate exhibits a statistically insignificant coefficient.

3.2.3 A unified analysis of regional factors

The study of property-level rent-to-price ratios has so far highlighted the importance of numerous local factors. The signs of most correlations align well with intuition, thus enhancing the credibility of our synthetic rent-to-price ratios. In the remainder of the

[25] Note that these housing market liquidity measures are computed from RWI-GEO-RED data, not from regional data.

paper, to reduce the dimensionality of the analysis, we focus on selected factors and include them in the same specification (7). Specifically, we restrict the analysis to those that are statistically significant at the 1% level and exhibit low pairwise correlations among each other.

Table 6, in column 1, reports estimates from regressions of rent-to-price ratios on the selected local factors: the old-to-working age ratio, disposable income per capita, completed living space per capita, and the number of businesses. These factors heuristically condense information on the demographics, the economic fundamentals, the housing market, and the size of the district. All coefficient estimates display coefficient signs and magnitudes that are consistent with above results.[26] It is also instructive to compare the fraction of rent-to-price ratio variation explained by this specification inclusive of key regional factors against that of the specification in column 4 of Table 3, which features property-level covariates alone. The adjusted R^2 increases by a mere 6%, from 31% to 37%, pointing to the limited explanatory power of district-level factors, i.e., of factors at a still relatively aggregate level. Below, we expand this analysis to more granular local factors.

In columns 2 and 3, we decompose the rent-to-price ratio, investigating separately the rent price per sqm and the sale price per sqm. Both variables exhibit correlations with the four district-level factors of opposite sign relative to the rent-to-price ratio in column 1, namely the rent-to-price ratio appears to be mostly driven by the sale price at the denominator.

In Figure 4, we conduct an Blinder-Oaxaca decomposition of the rent-to-price ratio gap between districts belonging to the top quintile of each the four selected regional dimensions against all other districts. Such a decomposition allows us to distinguish between (i) the component of the gap due to differences in the observable characteristics of the housing stock across the two groups (e.g., in terms of facilities or size) and (ii) the component reflecting both unobservable differences and different sensitivities to the observable characteristics. Depending on the conditioning variable, mean differences in rent-to-price ratios (whose unconditional average is 5.52%) are economically relevant,

[26]We generally confirm these correlation patterns also in Appendix Table A.2, where we use both data aggregated at the district level (Panel A) and the pseudo-panel described in Section 2.1 (Panel B). Only the result on the number of registered businesses becomes insignificant. In addition, in column 2 of Panel A we use the standard deviation (SD) of the rent-to-price ratio in a given district-calendar quarter as the dependent variable. Cross-sectional dispersion in valuation ratios correlates with the selected local factors in the same way as their mean, except for district size. In other words, rent-to-price ratios are more dispersed in larger districts in terms of registered businesses, in line with our conjecture about the presence of relevant agglomeration effects in large cities.

as they hover around 1%. But observable characteristics in the housing stock account only for a small part of the rent-to-price ratio differential. In other words, we do not observe stark discrepancies in the housing stock between say cities with high disposable income per capita and other areas. By contrast, rent-to-price ratio sensitivities to some characteristics (like the flat surface, its number of rooms, its conditions, and quality of its facilities) greatly differ between the two groups of districts. These differences in sensitivities could drive the unexplained part of the gap, as long as they do not conflate the effect of unobservable property traits.

3.3 Unexplained heterogeneity

The analyses conducted so far suggest that cross-sectional variation in rent-to-price ratios increases over our sample period, it is substantial even at granular geographic level, and differences in observable traits of the housing stock across districts are likely to account only for a tiny part of it. At the same time, property- and district-level covariates together with time fixed effects explain only about a third of the variation of property-level rent-to-price ratios.

However, above we have only considered a limited number of district-level factors observed at annual frequency. In Table 7, we seek to provide an upper bound to the fraction of rent-to-price ratio variability that can be explained by local factors. In particular, we saturate the baseline specification in column 4 of Table 3—including all observable property-level covariates (with an adjusted R^2 of 31%)—with progressively finer geographical fixed effects. Relative to the baseline, federal state and federal state-calendar quarter fixed effects can explain 4% and 5% more rent-price ratio variation (columns 1 and 2), respectively. In columns 3 and 4, we examine fixed effects at the district or district-calendar quarter level, i.e., the same administrative level as our regional factors above. In this case, the fraction of explained variation rises to 41% and 44%, respectively. Hence, cross-sectional variation inside districts appears to account for the lion's share of variation. The role of time variation may be partially concealed by this relatively coarse fixed effect structure, though. In columns 5 and 6, we augment the specification with zipcode and zipcode-calendar quarter fixed effects, reaching an adjusted R^2 of 47% and 59%, respectively.[27]

[27] This upper bound for R^2 appears not to be an artifact of the chosen specification. First, to address possible concerns about the use of the log-transformation (e.g., Cohn, Liu, and Wardlaw, 2021), in Appendix Table A.3 we use the non-transformed rent-to-price ratio as dependent variable. Reassuringly, the sign and the statistical significance of the coefficients for property-specific and regional covariates remains generally unchanged (columns 1 and 2), and the R^2 in the presence of zipcode-calendar quarter fixed

Unobservable (at least in our setting) time-varying factors at the zipcode level together with observable property-level traits are thus *potentially* able to explain a large fraction of heterogeneity in rent-to-price ratios. But variation that cannot be even theoretically explained by those variables is still at a staggering 41%. Given the richness of the vector of observable property-level traits we control for, we reckon measurement error due to the matching procedure and/or unobservable flat "deep" characteristics to be unlikely to underlie such a housing yields' heterogeneity puzzle.

Nonetheless, in Table 8 we seek to indirectly assess the role of measurement issues. Matching errors could stem from several important unobservable (to us) property features such as exact location (micro-level, such as a south-facing window), distance from public transport network, view from the balcony, decoration aesthetics, and cultural heritage, which tilt pricing for different investors. Combined with search and match frictions in the housing market, this could amplify pricing heterogeneity. To work around this problem, column 1 estimates the most saturated specification from column 6 of Table 7, but using as dependent variable the actual rather than the synthetic rent-to-price ratio, which is available for a subset of flats (probably non-owner-occupied flats for sale). In this case, by construction, we do not control for traits observable only for flats for rent, for distances in observables between matched flats, or for the corresponding missing value indicators. The adjusted R^2 increases substantially to 72%, but the higher explanatory power appears to be linked to sample bias more than to the absence of matching errors. Indeed, in column 2 we consider the same subsample, but we go back to the synthetic rent-to-price ratio as dependent variable, obtaining a similar adjusted R^2 of 70%. Yet, it is possible that sample bias maps into lower matching errors. Thus, we further scrutinize the matching error story, first, by explicitly controlling for a matching quality score and, second, by removing any property possibly showing up more than once in a period. In column 3, we augment the specification with fixed effects for each percentile of a matching quality score. The score is an equally-weighted average of re-scaled distances between the flat for rent and the matched flat(s) for sale for the covariates on which we do not match, assigning maximum distance to cases in which the trait is missing for one or both of the properties. In column 4, we exclude properties that, while still having a unique identifier

effects is even lower than for the log-linear specification (column 3). Second, because in some instances zipcodes are not perfectly nested within districts, it is theoretically possible to estimate coefficients for regional factors even after the inclusion zipcode-calendar quarter fixed effects. The addition of regional factors leaves the explanatory power typically unchanged relative to column 6 of Table 7. In this setting, the estimation of such factors' coefficients relies on small discrepancies between districts and zipcodes, making their interpretation complicated. These results are available upon request.

in the RWI-GEO-RED database, exhibit the same characteristics in terms of pricing and facilities in a given zipcode-calendar quarter. These properties are "likely duplicates", i.e., properties that are listed more than once on the platform (or by multiple realtors) and could contribute to artificially inflate heterogeneity. In neither specification, the adjusted R^2 exceeds the baseline value of 59%. Overall, these findings underpin the idea that matching errors do not play a major role in generating unexplained variation in rent-to-price ratios.

We then investigate to which extent cross-sectional heterogeneity in rent-to-price ratios evolves as we move from finer to coarser geographic units. After removing likely duplicate property listings that may dampen variation as highlighted above, we compute the standard deviation of valuation ratios for each geographic unit-calendar quarter (with at least 30 observations, to obtain meaningful estimates). Starting from non-log-transformed ratios to favor interpretation of economic magnitudes, we carry out this procedure for both raw ratios and ratios filtered for property-specific observables and zipcode-calendar quarter fixed effects, and at different levels of aggregation: zipcode area, district, and federal state. Table 9 contains information about the distribution of these cross-sectional standard deviations. Mean dispersion of raw ratios increases by 30% (= 2.188/1.681 − 1) going from zipcode- to federal state-level data. After factoring out observables and zipcode-calendar quarter fixed effects, the degree of dispersion is de facto invariant to the coarseness of aggregation (as we would expect), with a mean standard deviation of roughly 1.4% (or around 25% of the mean raw ratio). We observe similar patterns if we look at another measure of dispersion like the interquartile range. This evidence corroborates the relevance of variation in rent-to-price ratios within small geographic areas, as it persists when we aggregate properties into coarser regions.

3.3.1 Possible drivers

Cross-sectional heterogeneity in German rent-to-price ratios goes hand in hand with the prominent role of idiosyncratic shocks for the variance of property-level returns documented in the US market (Giacoletti, 2021).[28] But unexplained variation in yields may not be strictly idiosyncratic (e.g., related to a random liquidity shocks to landlords selling or renting the property at fire prices, to investors' heterogeneity in terms of risk/consumption preferences, etc.) in its entirety.

A complementary explanation for high zipcode-level residual heterogeneity in yields

[28] Giacoletti (2021) does not rely on property-level rental agreements and thus assumes the rent-to-price ratio to be constant within a given zipcode area.

relates to agglomeration economies operating below the zipcode level or across zipcodes. Rosenthal and Strange (2020) discuss how such phenomena—due, for instance, to knowledge spillovers and access to concentrated skilled workforce—depend on proximity and can be present at different levels, even within neighborhoods, blocks or buildings. A substantial part of the observed heterogeneity in rent-to-price ratios may be driven by such spatial effects and not to stand-alone traits of properties or of landlords/households. Whereas access to information on properties' exact geographic coordinates could make it possible to conduct a proper spatial analysis and to abstract from after all arbitrary partitions like zipcode areas, it may not solve the problem of identifying actual local housing market segments within which agglomeration economies unfold. Future work is needed to investigate the link between local agglomeration economies and rent-to-price ratios.

Furthermore, the presence of rent leveling may exacerbate cross-sectional dispersion of housing yields. Zipcode-calendar quarter fixed effects arguably absorb variation in yields due to the introduction and modification of rent leveling, which since 2013 in Germany are typically set by federal states' governments for specific municipalities (Kholodilin, Mense, and Michelsen, 2016), but this holds true inasmuch such policies are equally binding across the properties in a given zipcode area. Suppose instead that the rent leveling binds only for certain properties and, at same time, that market participants expect the controls to be lifted in the future (this is the case in Germany, where they are typically introduced for only a limited number of years), then we could in principle observe a depressing effect on the current rent (the numerator) coupled with a mild-to-null effect on the price (the denominator), artificially driving down rent-to-price ratios for such properties. While it is true that nationwide second-generation (i.e., linking allowed rent hikes to the cost of living) rent leveling has been long in place in Germany, as in most European countries, they have relatively loose until the 2010s, when renewed regulatory effort led to introduction of local caps on rent and to the so-called rental brake in certain municipalities (Kholodilin et al., 2016; Kholodilin, 2020). However, cross-sectional heterogeneity in housing yields is substantial even in the early years of our sample (see Appendix Figure A.1), suggesting that rent leveling—though probably contributing to it—is hardly its main driver.

Informational frictions are another possible source of unexplained heterogeneity in housing markets. The asset pricing analogy between housing and stocks relies on the assumption that information assimilation or availability is similar on the two markets. However, in reality, the housing market is far from the idealized case of centralized public stock exchanges, nor does it have an abundance of professionals like equity research analysts constantly seeking, parsing, and generating information on listed stocks. By contrast, it is often costly for housing investors (especially private individuals) to acquire and process information, which leads to (rationally) incomplete search and insufficient trading for price discovery. For instance, an investor in Frankfurt is unlikely to search the entire nation for housing investment opportunities, but she could quickly check the entire Frankfurt Stock Exchange. Nonetheless, this type information frictions does not explain away within-city (or even more: within-zipcode) heterogeneity in rent-to-price ratios, as one should have less difficulty assessing several city neighborhoods during the search. Market segmentation with limited search (Piazzesi, Schneider, and Stroebel, 2020) and limited market mobility (Head and Lloyd-Ellis, 2012) can help explain why localized supply and demand shocks do not spread in the housing markets, giving rise to cross-sectional heterogeneity in property valuations that is otherwise difficult to rationalize.

To insulate the role of regulatory and informational frictions from that of pure idiosyncratic shocks, one could envision exploiting (plausibly exogenous) local variation in such frictions to estimate their effects on the cross-sectional dispersion of rent-to-price ratios. For instance, the rental brake implemented in several municipalities in the 2010s may offer a suitable setting to explore the regulatory frictions story. Similarly, local variation in the institutional features of real estate agents' registers or in the fraction of professional as opposed to private landlords could proxy for the degree of information asymmetry surrounding properties. The implementation of these tests goes beyond the scope of the present paper.

3.3.2 Implications for household wealth

The high degree of heterogeneity—both explained and unexplained—in rent-to-price ratios has relevant consequences for household wealth around Germany. We attempt to quantify such consequences by means of some back-of-the-envelope calculations.

The median flat in our dataset, which we use as a benchmark, has a surface of 66 sqm and its annual rent per sqm (exclusive of expenses) is EUR 81.04. The 10th and 90th percentiles of the rent-to-price ratio (non log-transformed) are 3.13% and 8.36%, respectively.[29] These ratios correspond to housing valuations of EUR 170,883 (= (81.04 ·

[29] The 10th and 90th percentile of the rent-to-price ratio refer to its unconditional distribution. One may be concerned that in this way we are simply picking up variation across flats with very different characteristics. Nonetheless, the degree of heterogeneity in rent-to-price ratios decreases only by little if we focus on flats very similar to the median one. For instance, looking only at flats with a surface

66)/0.0313) and EUR 63,979 (= (81.04·66)/0.0836) for the median flat, implying a variation in the housing wealth of its owners of EUR 106,904 (= EUR 170, 883 − EUR 63, 979).

If we repeat the same calculations on the unexplained component of rent-to-price ratios alone, a not-so-different picture emerges. To this end, based on the specifications in columns 4 and 6 of Table 7—where, besides controlling for property-level characteristics, we include district- and zipcode-calendar quarter fixed effects—, we obtain residuals of the log-transformed percentage rent-to-price ratio: $\hat{\epsilon}_{d,t}$ and $\hat{\epsilon}_{z,t}$ respectively. We then retrieve the filtered rent-to-price ratio as $\exp\left[\text{P50}_{\ln(H/P_{f,t})} + \hat{\epsilon}_{d,t}\right]$, where $\text{P50}_{\ln(H/P_{f,t})}$ is the unconditional median log-transformed percentage rent-to-price ratio. We proceed analogously for $\hat{\epsilon}_{z,t}$. This procedure allows us to abstract from any variation in the valuation ratio due to observable property characteristics and time-varying factors at the district or zipcode level. Looking at the difference between the 10th and 90th percentiles of the filtered rent-to-price ratio, the implied variation in valuations of the median flat stands at EUR 74,183 and EUR 60,778 if we absorb time-varying district- or zipcode-level factors, respectively. Magnitudes shrink—as we would expect—but remain sizable.

To put things in perspective, we retrieve information on German household balance sheets from Deutsche Bundesbank (2019). As of 2017, the mean (resp., median) net wealth across all households is EUR 202,541 (resp., EUR 61,597). For the 44% of home-owning households, the mean (resp., median) value of their main residence stands at EUR 225,161 (resp., EUR 173,308). Given these figures, explained and unexplained heterogeneity in rent-to-price ratios can greatly impact households. As conjectured above, agglomeration economies and (regulatory) frictions taking place at a geographically granular level may therefore underlie substantial disparities in household wealth via housing valuations. In turn, depending on households willingness and ability to consume out of housing wealth (e.g., Berger et al., 2018), this may translate in cross-sectional dispersion in the state of local business cycles and, ultimately, economic development.

4 Housing Returns

In this section, we rely on a pseudo-panel dataset and shift our attention to the time-series dimension of house prices while preserving non-trivial cross-sectional heterogeneity, as captured by cohorts of flats formed by district-rooms-surface category.

In Table 10, we provide some stylized facts on cohort-level returns, rent growth rates,

between 60 and 70 sqm and an annual rent per sqm between EUR 75 and EUR 85, the 10th and 90th percentiles of the ratio are 3.31% and 8.32%.

and rent-to-price ratios within our pseudo-panel. The average total housing return (r) stands at 1.98% per quarter, with an average price growth rate (r^*) per quarter of 0.78%. Rents grow at an average quarterly rate (Δh) of 0.35%. Total returns exhibit a correlation of 0.99 with price growth, whereas their correlation with rent growth is a mere 0.04. Excess returns (r^e) exceed total returns, because over our sample period risk-free rates are for long periods negative. The mean non-log-transformed rent-to-price ratio (H^q/P) is 1.21%: this is based on quarterly rather than annual rents, so its magnitude is consistent with the analysis above. However, it is worth noting that rent-to-price ratios in the pseudo-panel are substantially less heterogeneous than those from the matching exercise. The aggregation of flats in cohorts (by taking averages) absorbs a hefty fraction of cross-sectional variation. In other words, tests based on our pseudo-panel are most informative about time-series variation. The rent-to-price ratio is highly persistent with significant first- and fourth-order autocorrelation coefficients of 0.81 and 0.74. Return and rent growth are less persistent but still exhibit significant first-order autocorrelations. In both cases, moreover, such autocorrelations are negative at -0.36 and -0.34, respectively, pointing to the existence of some degree of overshooting in the pricing of properties.[30]

In Figure 5, we visualize (cumulative) housing returns. Panel A of Figure 5 plots the distribution of quarterly returns and growth by year. We do observe much more pronounced cross-sectional dispersion in r and r^* than in Δh. This may point to a primary role of expected discount rates in generating variation in rent-to-price ratios across properties. In Panel B, we then look at cumulative returns for buy-and-hold strategies of different housing assets against the cumulative returns from holding the 10-year German Bund between 2007 and 2017. The median housing asset delivers a cumulative return of about 90%. Even housing assets in the bottom percentile largely outperform the Bund, delivering a return of about 25% over the sample period (against less than 10%). Distinguishing among properties based on rooms-surface combinations, we obtain cumulative returns ranging approximately between 75% and 110% (Panel C). It is worth noting that small housing units tend to consistently outperform large ones.

[30] Appendix Table A.4 explores the distribution of cohort-level correlations between housing yield components—excess returns and rent growth rates—, and per capita growth rates of different components of district-level household consumption. In particular, we proxy for non-durable, durable, and housing consumption by means of household waste production, the number of registered vehicles and the housing stock per capita, respectively. Cross-district heterogeneity in the responsiveness of consumption to yield components appears to be substantial, with a surprisingly large fraction of negative correlations. However, these estimates should be interpreted with a grain of salt, because we can only observe local consumption very noisily, i.e., at annual frequency and in non-monetary terms.

4.1 Predictive regressions

Return and rent growth data from the pseudo-panel appear to constitute credible measures of housing market conditions. We make use of these data to examine return and rent growth predictability, with goal of further validating our empirical proxies.

The rent-to-price ratio, analogously to the dividend-to-price ratio for stocks (Campbell and Shiller, 1988), is a gauge of housing market participants' expectations about future returns and rent growth on properties (Plazzi et al., 2010). Since the rent-to-price ratio—the housing yield—captures investors' belief, then its current level must predict future housing returns and rent growth rates to the extent they are predictable. Thus, finding evidence of predictability would lend support to our estimates of housing yields and to our empirical findings.

Unlike Plazzi et al. (2010), we do not carry out a structural estimation of the predictive regressions. The analysis is instead based on reduced-form regressions and is thus correlational in nature. In particular, we estimate predictive regressions at the cohort-level

$$\begin{aligned} r^e_{c,t+1\to t+k} &= \beta_k \ln(H^q/P_{c,t}) + \tau_c + \nu_{c,t+1\to t+k}, \\ \Delta h_{c,t+1\to t+k} &= \lambda_k \ln(H^q/P_{c,t}) + \tau_c + \varsigma_{c,t+1\to t+k}, \end{aligned} \tag{8}$$

where $r^e_{c,t+1\to t+k}$ and $\Delta h_{c,t+1\to t+k}$ are the k-quarter-ahead excess return and rent growth rates, respectively. To focus on time-series variation, we include cohort-level fixed effects, τ_c, which capture any time-invariant difference across cohorts (and, therefore, districts). To account for heteroskedasticity and correlation in error terms $\nu_{c,t+1\to t+k}$and $\varsigma_{c,t+1\to t+k}$, we adjust standard errors following Driscoll and Kraay (1998).[31]

In Table 11, we estimate the predictive specifications in (8). In columns 1 to 3, we look at excess housing returns at horizons ranging from one quarter to three years. The rent-to-price ratio is invariably significant and positively related to the current rent-to-price ratio. The within R^2 suggests that the discount rate expectations impounded in the ratio explain up to 16% of the time-series variation in cohort-level housing premia. The rent-to-price ratio loads significantly and negatively in the case of rent growth specifications, explaining up to 9% of variation. Findings on both return and rent growth rate line up with the traditional present-value relationship: a higher rent-to-price ratio descends from higher expected discount rates and/or lower expected rent growth. Yet, the predictive

[31]In unreported tests, we verify that our findings are not sensitive to removing cohort fixed effects or to using alternative standard errors.

ability is lower than what found by Plazzi et al. (2010) for US commercial real estate properties.

Appendix Table A.5 shows that the relation between excess returns and rent growth, and the rent-to-price ratio remains qualitatively unchanged if we control for selected local factors. But the within R^2 substantially increases (up to 31% and 16%, respectively), i.e., current local factors contain information useful to predict future return and rent dynamics that is not fully incorporated in the rent-to-price ratio, even though the micro-level evidence of Section 3 unambiguously shows that these quantities co-move.

Overall, expectations about future discount rates and rents that are factored in rent-to-price ratios explain a statistically significant and economically meaningful part of realized return and rent dynamics. At the same time, this finding gives credibility to our empirical measures of housing yields.

5 Global cities vs. other districts

The existing literature across different fields points to a special role played by metropolitan areas and the agglomeration economies they come with (e.g., Combes and Gobillon, 2015; Rosenthal and Strange, 2020). These areas tend to attract affluent households and to have limited land availability, which—combined—lead to a surge in housing valuations (e.g., Gyourko et al., 2013; Van Nieuwerburgh and Weill, 2010). But not every large city is a global city: in this respect, the gap in housing valuations between Figure 3 (global cities) and Appendix Figure A.2 (other large cities) is telling. Here, we thus focus on seven globally relevant German cities.

Figure 6 visualizes the differences along some key dimensions between global cities and other districts. The Blinder-Oaxaca decomposition in Panel A points out that the large average gap in rent-to-price ratios is de facto unrelated to differences in observable characteristics of the housing stock in global cities vs. other districts. At the same time, Panel B illustrates that the cumulative return differential is sizable (around 20% over the sample period). These effects are to some extent present even when identifying large cities using the top quintile of the number of registered business (results available upon request). However, the aforementioned patterns become more pronounced once we use the narrower group of global cities and are arguably driven by such cities.

Table 12 hints that global cities are special also in terms of expectations about returns and rents. After re-estimating the predictive regressions of (8) over the pseudo-panel for such a subsample, we find that the rent-to-price ratio has no significant predictive ability for returns in global cities (within R^2 stands at a mere 3% vs. 17% in other districts). By contrast, an important fraction of variation in future rent growth rates can be explained by time-varying investors' expectations as captured by valuation ratios (within R^2 of 19% vs. 8% in other districts). Hence, the predictability patterns for global cities are reversed with respect to other regions. The economic mechanism underlying this phenomenon represents a new interesting avenue of research.

Together, these stylized facts support the idea that global cities feature peculiar characteristics, which have become more noticeable over our sample period.

6 Conclusion

Relying on sale and rental prices for flats from a major German online real estate platform, we study the distribution and the drivers of rent-to-price ratios (or housing yields). To this end, we compute a measure of the rent-to-price ratio at the property-level by matching flats for rental to those for sale. District-level factors (such as demographics, economic fundamentals, housing market features) appear to co-move with valuation ratios, but to explain only a limited part of their variation, even after accounting for a wide set of property-specific characteristics. The same continues to hold if we absorb all time-varying zipcode-level factors by means of fixed effects. In other words, much of the variation of rent-to-price ratios remains unexplained. Such a heterogeneity puzzle is arguably related to idiosyncratic factors and/or (regulatory) frictions and agglomeration economies operating within zipcode areas.

Because flat listings in our database are provided in monthly vintages akin to repeated cross-sections, we construct a pseudo-panel by aggregating cohort of properties, to track time-series variation in house prices and further validate our empirical approach to measuring housing yields by means of a predictability exercise. In this way, we show that rent-to-price ratios, which incorporate time-varying market participants' expectations about house pricing dynamics, indeed display significant and economically meaningful ability to predict returns and rent growth.

Overall, this paper points to the existence of a surprisingly large degree of unexplained heterogeneity in rent-to-price ratios. Given the importance of real estate valuations both for the distribution of wealth across households and for the amplification of business cycles through the consumption channel, further work is needed to pin down the origin(s) of such yet unexplained variation, be it related to idiosyncratic risk, to local agglomeration effects, informational and regulatory frictions, or to other economic mechanisms.

References

Abraham, J. M., and P. H. Hendershott. 1996. Bubbles in metropolitan housing markets. *Journal of Housing Research* 7:191–207.

Adelino, M., A. Schoar, and F. Severino. 2015. House prices, collateral, and self-employment. *Journal of Financial Economics* 117:288–306.

Ahlfeldt, G. M., and A. Feddersen. 2018. From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography* 18:355–390.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf. 2015. The economics of density: Evidence from the Berlin Wall. *Econometrica* 83:2127–2189.

an de Meulen, P., M. Micheli, and S. Schaffner. 2014. Documentation of German real estate market data. Working paper, RWI Essen.

Bednarek, P., D. M. Te Kaat, C. Ma, and A. Rebucci. 2020. Capital flows, real estate, and local cycles: Evidence from German cities, banks, and firms. *Review of Financial Studies* (Forthcoming).

Beraja, M., A. Fuster, E. Hurst, and J. Vavra. 2019. Regional heterogeneity and the refinancing channel of monetary policy. *Quarterly Journal of Economics* 134:109–183.

Berger, D., V. Guerrieri, G. Lorenzoni, and J. Vavra. 2018. House prices and consumer spending. *Review of Economic Studies* 85:1502–1542.

Bracke, P. 2015. House prices and rents: Microevidence from a matched data set in Central London. *Real Estate Economics* 43:403–431.

Branch, W. A., N. Petrosky-Nadeau, and G. Rocheteau. 2016. Financial frictions, the housing market, and unemployment. *Journal of Economic Theory* 164:101–135.

Breidenbach, P., and S. Schaffner. 2020. Real estate data for Germany (RWI-GEO-RED). *German Economic Review* 21:401–416.

Cameron, G., and J. Muellbauer. 2001. Earnings, unemployment, and housing in Britain. *Journal of Applied Econometrics* 16:203–220.

Campbell, J. Y., S. Giglio, and P. Pathak. 2011. Forced sales and house prices. *American Economic Review* 101:2108–31.

Campbell, J. Y., and R. J. Shiller. 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1:195–228.

Campbell, S. D., M. A. Davis, J. Gallin, and R. F. Martin. 2009. What moves housing markets: A variance decomposition of the rent-price ratio. *Journal of Urban Economics* 66:90–102.

Case, K. E., and C. J. Mayer. 1996. Housing price dynamics within a metropolitan area. *Regional Science and Urban Economics* 26:387–407.

Case, K. E., and R. J. Shiller. 1989. The efficiency of the market for single-family homes. *American Economic Review* pp. 125–137.

Chambers, D., C. Spaenjers, and E. Steiner. 2021. The rate of return on real estate: Long-run micro-level evidence. *Review of Financial Studies* (Forthcoming).

Chang, B., H.-S. Choi, H. G. Hong, and J. D. Kubik. 2017. Hedging and pricing rent risk with search frictions. Working paper, University of Wisconsin-Madison.

Clark, W. A., M. C. Deurloo, and F. M. Dieleman. 2000. Housing consumption and residential crowding in US housing markets. *Journal of Urban Affairs* 22:49–63.

Cohn, J. B., Z. Liu, and M. Wardlaw. 2021. Count data in finance. Working paper, University of Texas at Austin.

Combes, P.-P., and L. Gobillon. 2015. The empirics of agglomeration economies. In *Handbook of Regional and Urban Economics*, vol. 5, pp. 247–348. Elsevier.

Correia, S. 2018. REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Boston College.

Deaton, A. 1985. Panel data from time series of cross-sections. *Journal of Econometrics* 30:109–126.

Deutsche Bundesbank. 2019. Household wealth and finances in Germany: Results of the 2017 survey. *Monthly Report March* 57.

Driscoll, J. C., and A. C. Kraay. 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics* 80:549–560.

Duca, J. V., J. Muellbauer, and A. Murphy. 2011. House prices and credit constraints: Making sense of the US experience. *Economic Journal* 121:533–551.

Duca, J. V., J. Muellbauer, and A. Murphy. 2021. What drives house price cycles? International experience and policy issues. *Journal of Economic Literature* 59:773–864.

Eichholtz, P., M. Korevaar, T. Lindenthal, and R. Tallec. 2021. The Total Return and

Risk to Residential Real Estate. *Review of Financial Studies* (Forthcoming).

Eiling, E., E. Giambona, R. Lopez Aliouchkin, and P. Tuijp. 2019. The cross-section of expected housing returns. Working paper, University of Amsterdam.

Engsted, T., and T. Q. Pedersen. 2015. Predicting returns and rent growth in the housing market using the rent-price ratio: Evidence from the OECD countries. *Journal of International Money and Finance* 53:257–275.

Favero, C. A., A. E. Gozluklu, and A. Tamoni. 2011. Demographic trends, the dividend-price ratio, and the predictability of long-run stock market returns. *Journal of Financial and Quantitative Analysis* 46:1493–1520.

Flavin, M., and T. Yamashita. 2002. Owner-occupied housing and the composition of the household portfolio. *American Economic Review* 92:345–362.

Garcia, D., and Ø. Norli. 2012. Geographic dispersion and stock returns. *Journal of Financial Economics* 106:547–565.

Geanakoplos, J., M. Magill, and M. Quinzii. 2004. Demography and the long-run predictability of the stock market. *Brookings Papers on Economic Activity* 2004:241–307.

Giacoletti, M. 2021. Idiosyncratic risk in housing markets. *Review of Financial Studies* 34:3695–3741.

Granziera, E., and S. Kozicki. 2015. House price dynamics: Fundamentals and expectations. *Journal of Economic Dynamics and Control* 60:152–165.

Guerrieri, C., and C. Mendicino. 2018. Wealth effects in the euro area. Working paper, European Central Bank.

Gyourko, J., C. Mayer, and T. Sinai. 2013. Superstar cities. *American Economic Journal: Economic Policy* 5:167–99.

Han, L. 2013. Understanding the puzzling risk-return relationship for housing. *Review of Financial Studies* 26:877–928.

Harari, M. 2020. Cities in bad shape: Urban geometry in India. *American Economic Review* 110:2377–2421.

Head, A., and H. Lloyd-Ellis. 2012. Housing liquidity, mobility, and the labour market. *Review of Economic Studies* 79:1559–1589.

Hill, R. J. 2013. Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys* 27:879–914.

Hill, R. J., and I. A. Syed. 2016. Hedonic price-rent ratios, user cost, and departures from equilibrium in the housing market. *Regional Science and Urban Economics* 56:60–72.

Himmelberg, C., C. Mayer, and T. Sinai. 2005. Assessing high house prices: Bubbles, fundamentals and misperceptions. *Journal of Economic Perspectives* 19:67–92.

Howard, G., and C. Liebersohn. 2020. Regional divergence and house prices. Working paper, University of Illinois at Urbana-Champaign.

Jordà, Ò., K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor. 2019. The rate of return on everything, 1870–2015. *Quarterly Journal of Economics* 134:1225–1298.

Kaas, L., G. Kocharkov, E. Preugschat, and N. Siassi. 2021. Low homeownership in Germany—A quantitative exploration. *Journal of the European Economic Association* 19:128–164.

Kantak, P. 2019. Local prospects and housing dynamics: The asset pricing consequences of consumption commitments. Working paper, Indiana University-Bloomington.

Kholodilin, K. 2020. Long-term, multicountry perspective on rental market regulations. *Housing Policy Debate* 30:994–1015.

Kholodilin, K. A., A. Mense, and C. Michelsen. 2016. Market break or simply fake? Empirics on the causal effects of rent controls in Germany. Working paper, DIW Berlin.

Krainer, J. 2001. A theory of liquidity in residential real estate markets. *Journal of Urban Economics* 49:32–53.

Ling, D. C., J. T. Ooi, and T. T. Le. 2015. Explaining house price dynamics: Isolating the role of nonfundamentals. *Journal of Money, Credit and Banking* 47:87–125.

Mian, A., K. Rao, and A. Sufi. 2013. Household balance sheets, consumption, and the economic slump. *Quarterly Journal of Economics* 128:1687–1726.

Mian, A., and A. Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *Quarterly Journal of Economics* 124:1449–1496.

Piazzesi, M., M. Schneider, and J. Stroebel. 2020. Segmented housing search. *American Economic Review* 110:720–59.

Plazzi, A., W. Torous, and R. Valkanov. 2010. Expected returns and expected growth in rents of commercial real estate. *Review of Financial Studies* 23:3469–3519.

Pogodzinski, J. M., and T. R. Sass. 1991. Measuring the effects of municipal zoning regulations: A survey. *Urban Studies* 28:597–621.

Poterba, J. M., D. N. Weil, and R. Shiller. 1991. House price dynamics: the role of tax policy and demography. *Brookings Papers on Economic Activity* 1991:143–203.

Rosenthal, S. S., and W. C. Strange. 2020. How close is close? The spatial reach of agglomeration economies. *Journal of Economic Perspectives* 34:27–49.

Ruf, D. 2016. Agglomeration effects and liquidity gradients in local rental housing markets. Working paper, Goethe University Frankfurt.

Saadi, V. 2020. Role of the community reinvestment act in mortgage supply and the US housing boom. *Review of Financial Studies* 33:5288–5332.

Sagi, J. S. 2021. Asset-level risk and return in real estate investments. *Review of Financial Studies* 34:3647–3694.

Savills. 2019. Eigentümerstruktur am Wohnungsmarkt. SPOTLIGHT Savills Research.

Savov, A. 2011. Asset pricing with garbage. *Journal of Finance* 66:177–201.

Shleifer, A. 1986. Do demand curves for stocks slope down? *Journal of Finance* 41:579–590.

Sinai, T., and N. S. Souleles. 2005. Owner-occupied housing as a hedge against rent risk. *Quarterly Journal of Economics* 120:763–789.

Smith, M. H., and G. Smith. 2006. Bubble, bubble, where's the housing bubble? *Brookings Papers on Economic Activity* 2006:1–67.

Takáts, E. 2012. Aging and house prices. *Journal of Housing Economics* 21:131–141.

Tang, Y., T. Zeng, and S. Zhu. 2020. Bubbles and house price dispersion in the United States during 1975–2017. *Journal of Macroeconomics* 63:103163.

Van Nieuwerburgh, S., and P. O. Weill. 2010. Why has house price dispersion gone up? *Review of Economic Studies* 77:1567–1606.

Verbeek, M. 2008. Pseudo-panels and repeated cross-sections. In *The Econometrics of Panel Data*, pp. 369–383. Springer.

Vermeulen, W., and J. Van Ommeren. 2009. Compensation of regional unemployment in housing markets. *Economica* 76:71–88.

Vuolteenaho, T. 2002. What drives firm-level stock returns? *Journal of Finance* 57:233–264.

Wu, Y. 2018. What's behind smooth dividends? Evidence from structural estimation. *Review of Financial Studies* 31:3979–4016.

Zabel, J. E. 2012. Migration, housing market, and labor market responses to employment shocks. *Journal of Urban Economics* 72:267–284.

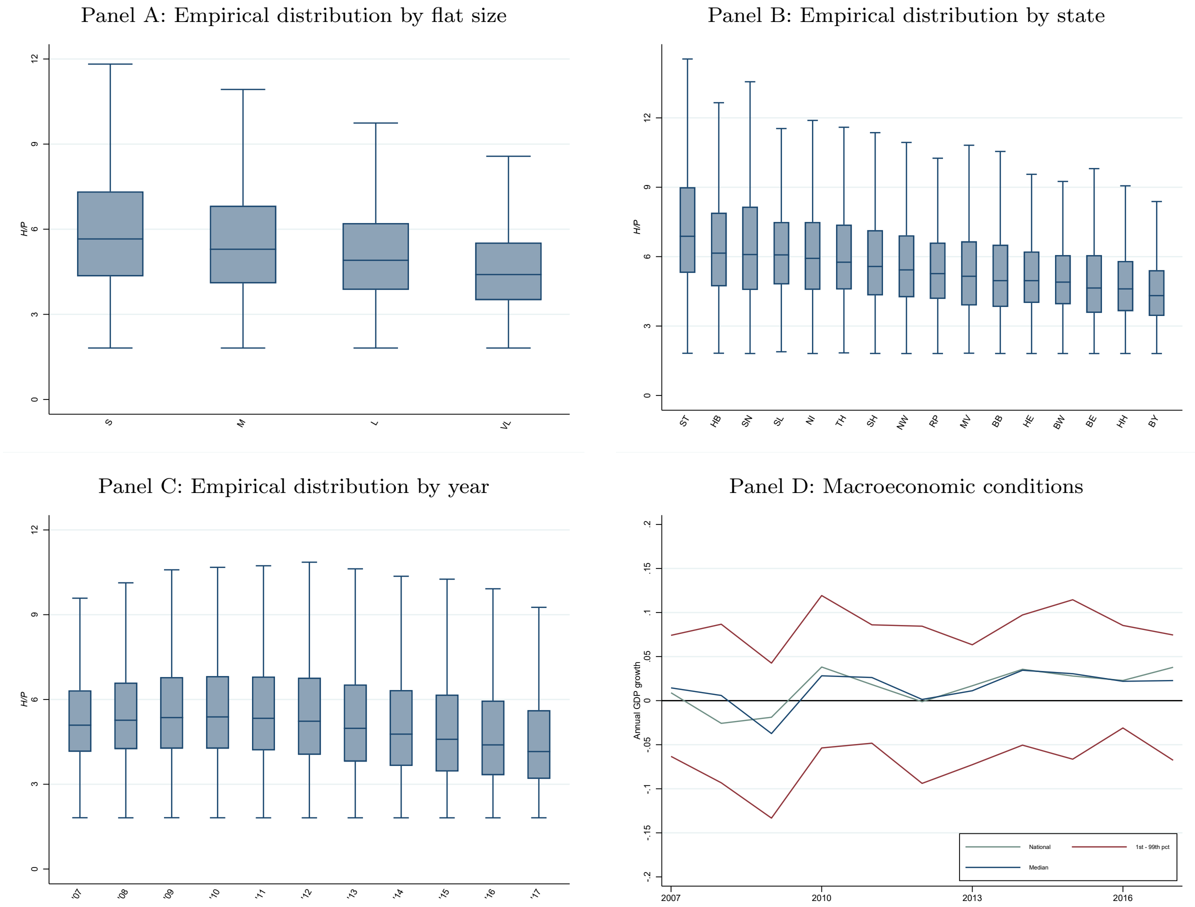


Figure 1: Empirical distribution of the rent-to-price ratio vs. macroeconomic conditions
This figure shows the conditional empirical distribution of the rent-to-price ratio (obtained via matching) through box plots (Panel A to C) against the evolution of macroeconomic conditions in Germany (Panel D) between 2007 and 2017. In Panel A, the conditioning variable is the flat's surface category, going from small (S) to medium (M), large (L), and very large (VL). In Panel B, the conditioning variable is the federal state where the flat is located (ordered by the median rent-to-price ratio): Saxony-Anhalt (ST), Bremen (HB), Saxony (SN), Saarland (SL), Lower Saxony (NI), Thuringia (TH), Schleswig-Holstein (SH), North Rhine-Westphalia (NW), Rhineland-Palatinate (RP), Mecklenburg-Vorpommern (MV), Brandenburg (BB), Hesse (HE), Baden-Württemberg (BW), Berlin (BE), Hamburg (HH), and Bavaria (BY). In Panel C, the conditioning variable is the year. Panel D shows the annual GDP growth rate at national level, as well as the median together with 1st and 99th percentile of the district-level GDP growth rate.

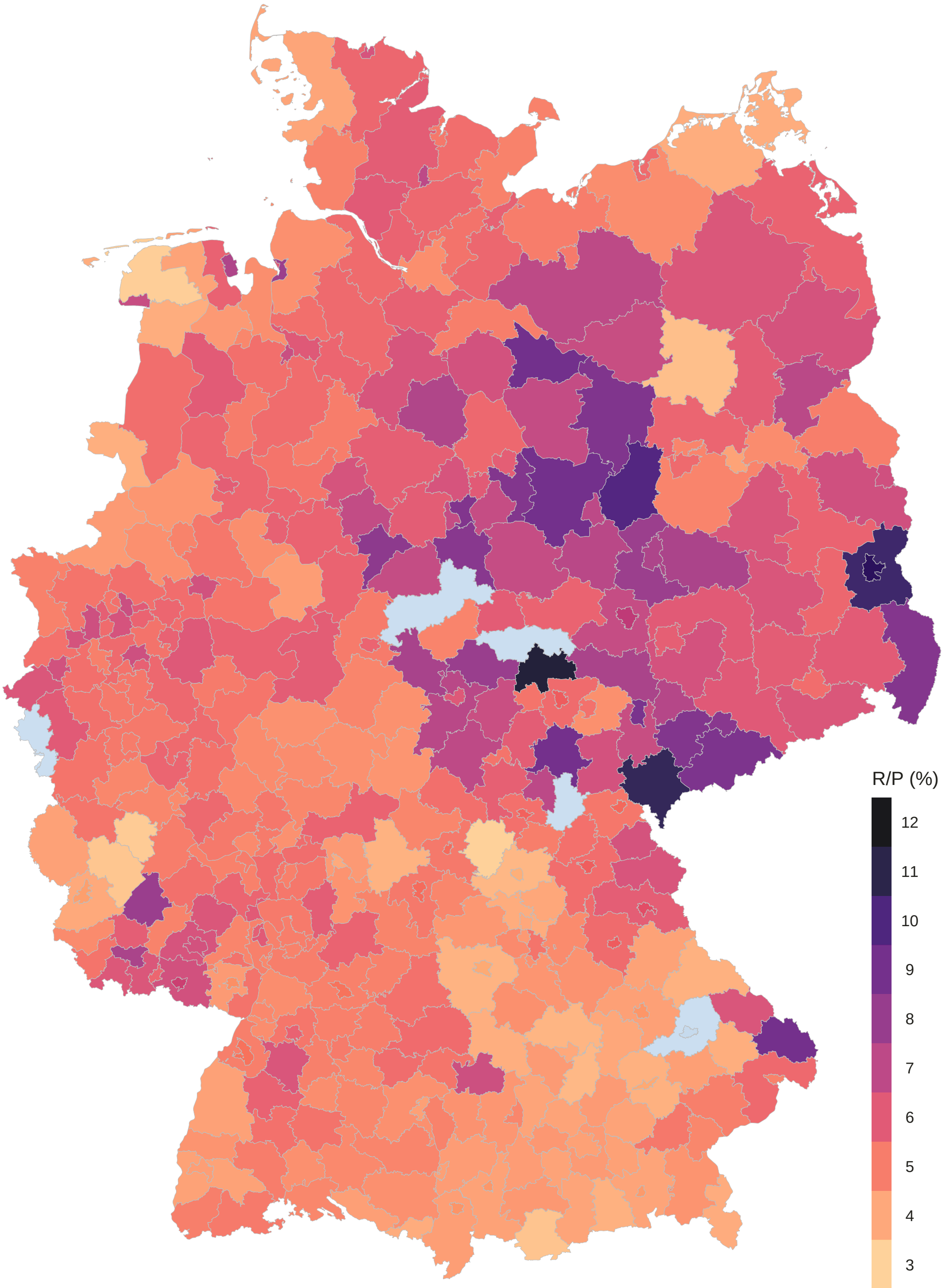


Figure 2: Median rent-to-price ratio at the district level
This figure visualizes the median rent-to-price ratio (obtained via matching) at the district level across Germany, pooling all periods between 2007 and 2017. Grey-colored districts have no observations.

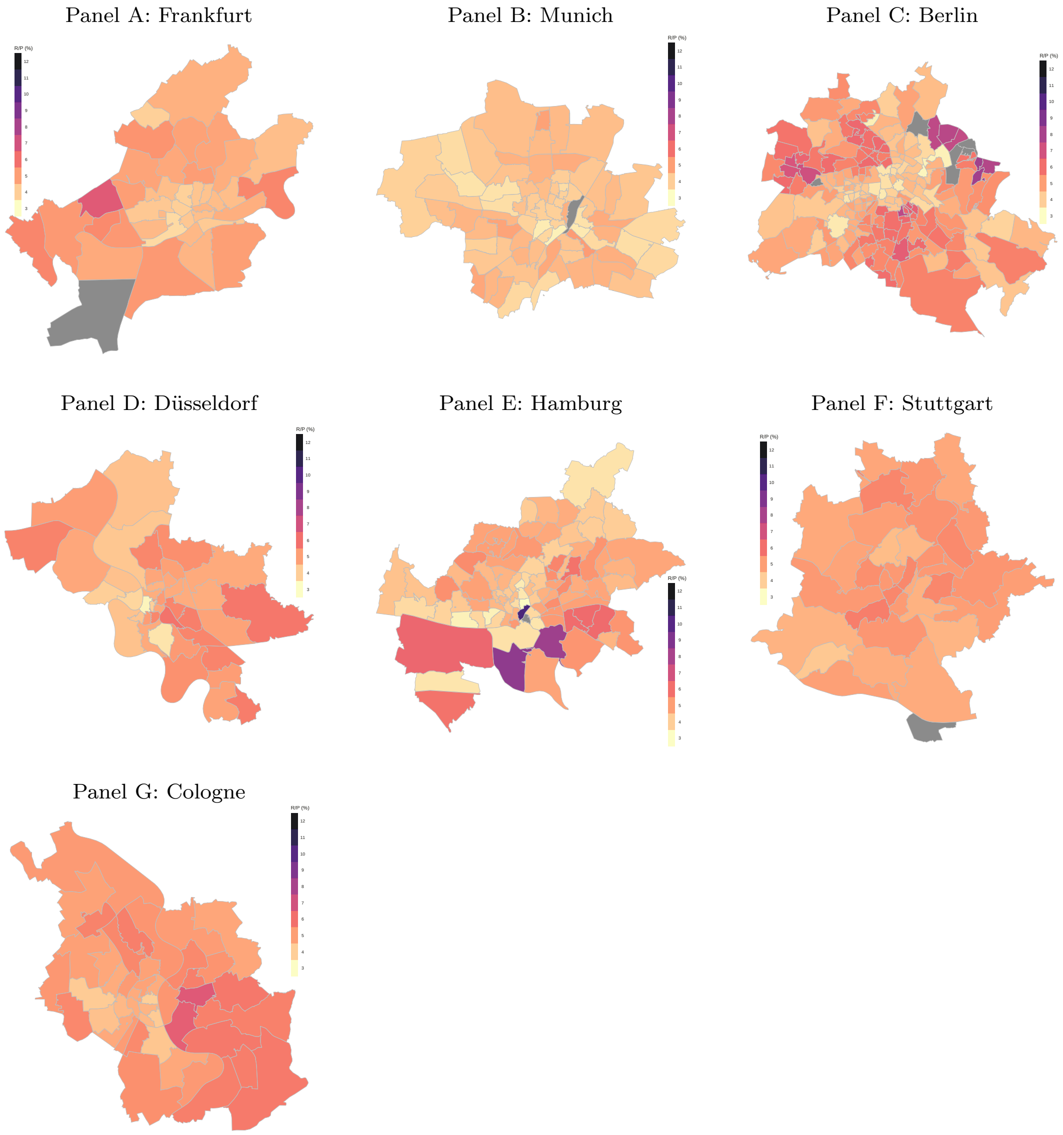
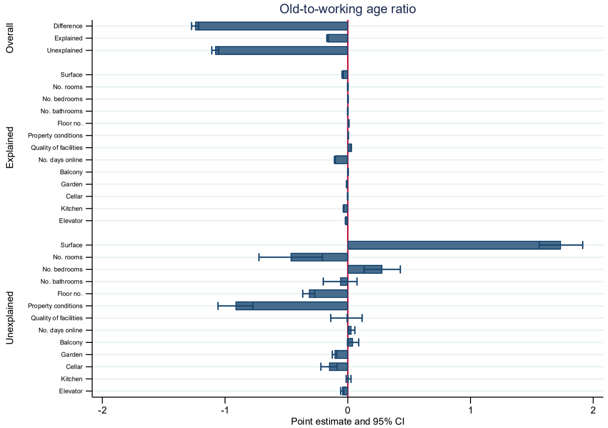
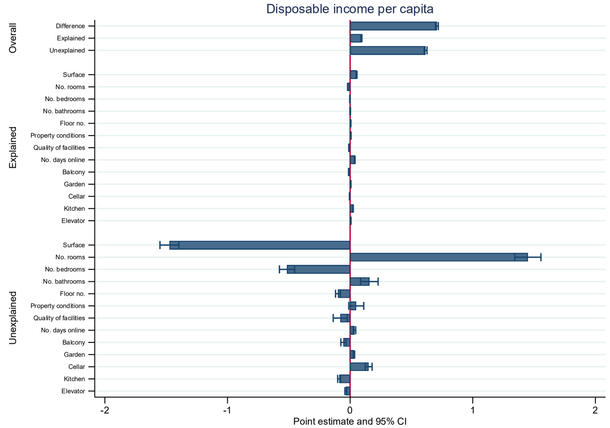


Figure 3: Median rent-to-price ratio at the zipcode level within global cities
This figure visualizes the median rent-to-price ratio (obtained via matching) at the five-digit zipcode level across German global cities, pooling all periods between 2007 and 2017. Grey-colored zipcode areas have no observations. Global cities are those assigned a rating ranging between "Alpha" and "Gamma" in the 2020 ranking by the Globalization and World Cities Research Network. Each of the panels from A to H corresponds to a different city.

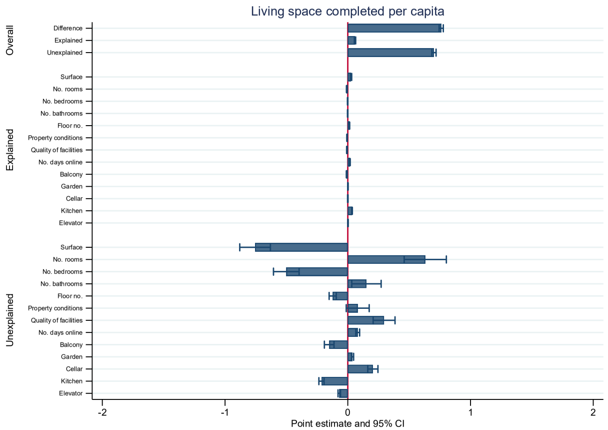
Panel A: Old-to-working age ratio



Panel B: Disposable income per capita



Panel C: Living space completed per capita



Panel D: No. businesses

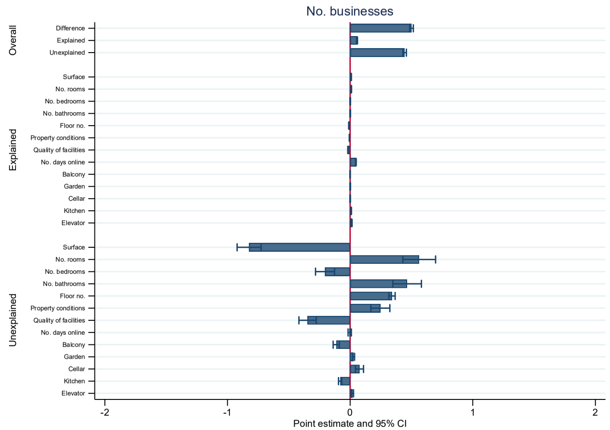


Figure 4: Blinder-Oaxaca decomposition of the difference in rent-to-price ratios across regions
Each panel in this figure shows the output of a Blinder-Oaxaca decomposition of the difference in rent-to-price ratios (obtained via matching) between properties located in districts belonging to the top quintile of a given regional trait (based on its average over all sample years) and the other areas. The conditioning district variable is indicated on the top of each panel.

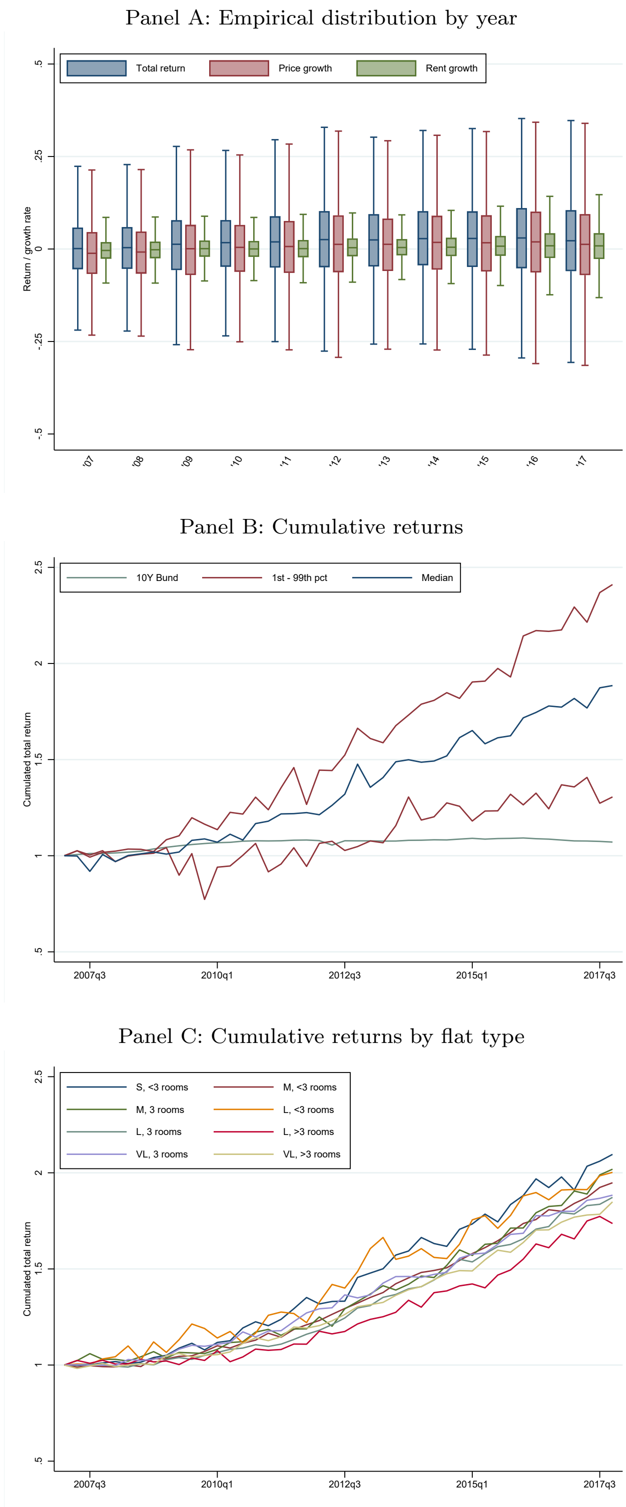
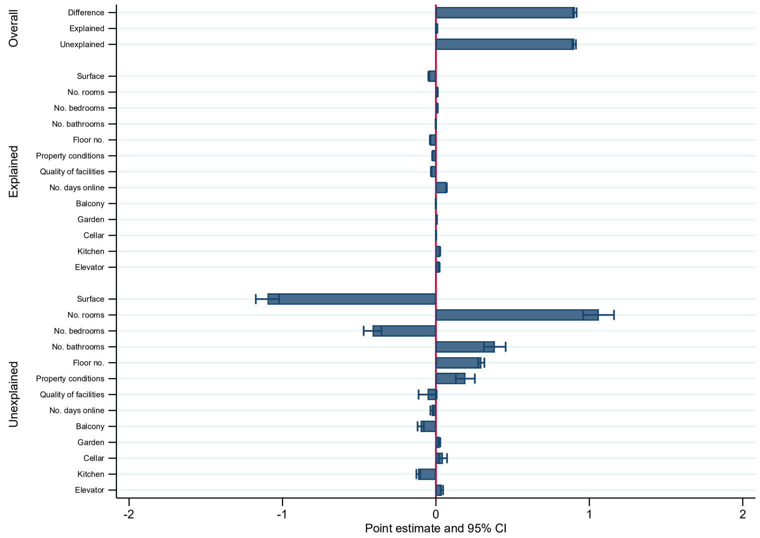


Figure 5: Housing returns
This figure shows the distribution of housing returns obtained via the pseudo-panel approach for the period 2007-2017. Panel A reports the empirical distribution of quarterly total returns (r), price growth rates (r^*), and rent growth rates (Δh) by means of year-by-year box plots. Panel B illustrates the evolution of the median as well as the 1st and 99th percentile of cumulative returns (focusing on cohorts with consecutive non-missing observations over the entire sample period). These time-series are plotted together with the cumulative return on the 10-year Bund. Panel C illustrates the evolution of the median of cumulative returns conditional on the rooms-surface category of properties. In each case, the quantiles of cumulative returns are computed as of 2017Q3.

Panel A: Blinder-Oaxaca decomposition



Panel B: Cumulative returns

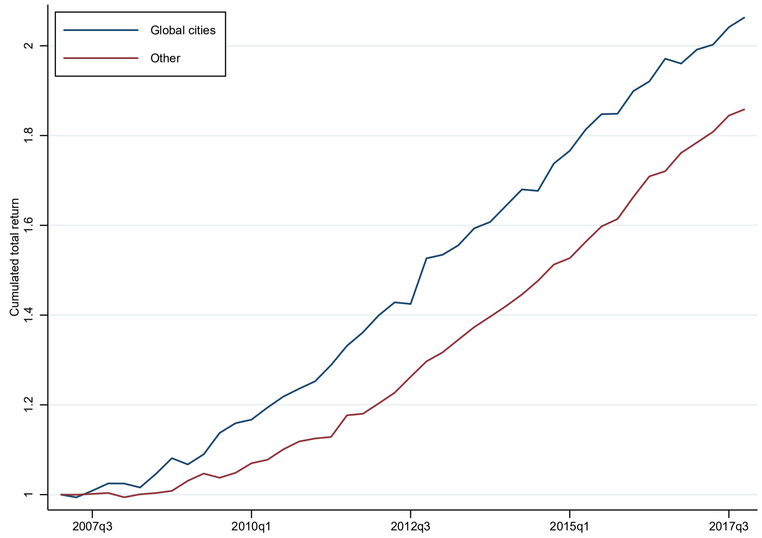


Figure 6: Global cities vs. other areas
This figure compares housing market conditions in global cities against other areas of Germany for the period 2007-2017. Panel A shows the output of a Blinder-Oaxaca decomposition of the difference in rent-to-price ratios (obtained via matching) between global cities and other areas. Panel B shows the median cumulative return (obtained from the pseudo-panel) for global cities and other areas.

Table 1: Summary statistics on flats and the local economy

This table shows summary statistics on property characteristics distinguishing between flats for rent and flats for sale (Panel A) and on socioeconomic and housing market conditions at the district level (Panel B) between 2007 and 2017. Refer to Appendix Table A.1 for variable definitions.

Panel A: Flat characteristics

	For rent			For sale (matched)			Mean-comparison test	
	Obs.	Mean	SD	Obs.	Mean	SD	Diff.	t-stat
Rental listing								
Annual rent (EUR)	1,615,217	5,854.832	2,467.206					
Annual inclusive rent (EUR)	1,623,237	7,321.903	2,314.765					
Expenses (EUR)	1,623,237	124.305	47.894					
Heating expenses (EUR)	1,623,237	62.268	12.978					
Heating included	1,623,237	0.729	0.445					
Deposit (EUR)	1,623,237	1,236.568	609.577					
Sale listing								
Sale price (EUR)				1,615,115	126,387.604	79,972.709		
Rented				1,623,237	0.219	0.411		
Housing benefits (EUR)				1,623,237	182.878	50.697		
Holiday property				1,623,237	-2.889	4.164		
Matching covariates								
Surface (sqm)	1,623,237	68.508	18.925	1,623,237	68.514	18.755	0.006	0.286
No. rooms	1,623,237	2.449	0.678	1,623,237	2.449	0.678	0.000	0.000
No. bedrooms	1,623,237	1.447	0.582	1,623,237	1.447	0.582	0.000	0.000
No. bathrooms	1,623,237	1.025	0.156	1,623,237	1.025	0.156	0.000	0.000
Floor no.	1,623,237	1.949	1.168	1,623,237	1.972	1.220	0.024	17.955

Panel A: Flat characteristics (continued)

	For rent			For sale (matched)			Mean-comparison test	
	Obs.	Mean	SD	Obs.	Mean	SD	Diff.	*t*-stat
Other characteristics								
No. building floors	1,310,657	3.396	1.358	1,324,507	3.741	1.902	0.345	169.057
Usable surface (sqm)	528,648	36.698	30.146	598,559	24.706	26.694	-11.992	-223.992
Construction year	1,006,827	1,967.503	39.168	1,432,602	1,971.472	36.386	3.969	81.263
Year last modernization	430,130	2,009.491	5.448	464,617	2,006.770	6.636	-2.722	-211.076
Property conditions	1,367,471	2.304	0.488	1,407,539	2.297	0.547	-0.007	-11.013
Quality of facilities	885,109	2.477	0.601	904,480	2.483	0.613	0.006	6.718
Protected building	87,068	0.008	0.090	1,125,763	0.052	0.220	0.044	58.018
Energy consumption (kWh/(sqm $\times$ y))	360,210	121.581	50.142	495,526	121.270	61.217	-0.310	-2.494
Energy rating	26,956	4.367	1.988	34,259	3.892	1.899	-0.475	-30.087
Hot water in energy consumption	878,746	0.150	0.357	888,372	0.225	0.414	0.075	129.169
Balcony	1,418,499	0.794	0.405	1,499,800	0.871	0.332	0.078	179.744
Available parking	159,211	0.934	0.248	198,328	0.957	0.201	0.023	30.470
Accessible with wheelchair	16,150	0.793	0.405	24,636	0.915	0.278	0.123	36.206
Guest WC	1,118,158	0.183	0.387	1,180,646	0.215	0.407	0.031	59.774
Garden	1,019,747	0.294	0.455	1,014,795	0.335	0.469	0.041	63.390
Cellar	1,283,667	0.806	0.396	1,329,429	0.830	0.371	0.025	51.817
Kitchen	1,211,684	0.563	0.496	1,215,469	0.563	0.492	-0.000	-0.627
Elevator	1,099,023	0.314	0.464	1,156,298	0.452	0.495	0.138	215.588
Assisted living	517,668	0.055	0.228	491,394	0.124	0.328	0.069	122.848
Immediate availability	1,526,724	0.316	0.465	1,138,168	0.344	0.469	0.028	48.570
No. of days online	1,615,180	26.637	33.399	1,623,237	62.898	90.884	36.262	476.096
No. of hits	1,615,106	905.636	1,024.022	1,623,237	876.336	1,486.243	-29.300	-20.648
No. of clicks (contact button)	1,615,112	11.033	22.451	1,623,237	6.164	14.017	-4.870	-234.251
No. of clicks (customer profile)	1,615,421	1.278	4.279	1,623,237	0.966	4.315	-0.312	-65.290
No. of clicks (share button)	1,616,085	0.861	1.544	1,623,237	0.730	1.621	-0.131	-74.464
No. of clicks (customer URL)	1,615,498	3.237	6.025	1,623,237	3.908	17.192	0.671	46.820

Panel B: Local conditions

	Obs.	Mean	SD	Min	Median	Max
Demographic and economic conditions						
Old-to-working age ratio (%)	3,469	34.940	4.509	22.704	34.682	52.192
Population (thousand)	3,469	22,527.278	25,528.577	3,394.400	17,061.400	361,349.500
Disposable income per capita (EUR, thousand)	3,438	18.760	2.214	14.050	18.735	28.394
GDP per capita (EUR, thousand)	3,440	30.307	11.921	14.134	27.196	88.045
Unemployment rate (%)	3,469	6.680	3.237	1.300	6.000	22.000
Manufacturing industry share (%)	3,457	7.279	2.396	2.518	7.013	21.360
No. businesses	3,469	10,689.176	13,707.748	1,537.000	7,800.000	189,177.000
Housing market conditions						
IQR of surface (sqm)	3,469	31.337	7.258	2.590	30.980	92.000
Average online listing time (days)	3,469	41.906	20.684	3.112	39.197	155.259
No. posted flats	3,469	445.268	1,356.146	5.000	163.000	29,504.000
Housing stock per capita (sqm)	3,442	44.408	4.016	35.457	44.323	57.357
Living space completed per capita (sqm)	3,446	0.269	0.137	0.045	0.250	0.747
Land price (sqm)	3,414	145.557	118.225	6.961	110.437	831.342
Property assessment rate (B) (%)	3,469	395.848	83.749	244.000	377.000	855.000

Table 2: Valuation ratios in the housing vs. stock market
This table shows summary statistics regarding valuation ratios in the German housing market and in the US stock market between 2007 and 2017. Housing market valuation ratios comprise the synthetic rent-to-price ratio (obtained via matching) as well as the actual one (available for a subsample of properties on sale for which a rental income is reported). Stock market valuation ratios comprise the dividend-to-price ratio (D/P), the earnings-to-price ratio (E/P), and the cash flow-to-price ratio (CF/P). The percentiles of stock market ratios are averages of annual Fama-French portfolio breakpoints over the sample period. Refer to Appendix Table A.1 for variable definitions.

	P5	P25	P50	P75	P95	P75−P25	P95−P5
German housing market							
H/P	2.740	3.947	5.077	6.545	9.844	2.598	7.103
Actual H/P	3.390	4.593	5.654	7.227	11.429	2.635	8.039
US stock market							
D/P	0.349	1.149	2.036	3.297	6.647	2.148	6.299
E/P	1.518	4.207	5.885	7.982	14.584	3.775	13.065
CF/P	2.304	5.892	8.169	11.394	19.827	5.502	17.524

Table 3: Rent-to-price ratio and property characteristics
This table reports estimates from regressions for property-level rent-to-price ratios on characteristics of the flat for a German sample between 2007 and 2017. The log-transformed rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. Column 1 includes the covariates on which the matching exercise is performed. Column 2 adds flat traits observed only for flats for rent. Column 3 adds flat traits observed only for flats for sale. The covariates added in columns 2 and 3 are set to 0 if missing. To control for that, for each of those variable a missing value indicator (equal to 1 if such a variable is missing and 0 otherwise) is added. Column 4 augments the specification with a set of covariates capturing the differences between the matched flats for rent and for sale with respect to a host of traits (number of floors in the building, usable surface, year of construction, year of the last modernization, property conditions, quality of facilities, protected building status, market segment, energy consumption, energy rating, inclusion of hot water in energy consumption, balcony, availability of parking place, accessibility with wheelchair, guest WC, garden, cellar, installed kitchen, elevator, living assistance, immediate availability, number of days online, and number of clicks on different items of the listing), together with the corresponding missing value indicators. All specifications include calendar quarter fixed effects. The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$\ln(H/P)$			
	(1)	(2)	(3)	(4)
Surface	-0.020***	-0.019***	-0.018***	-0.018***
	(-23.66)	(-14.63)	(-13.49)	(-16.66)
Surface squared	0.000***	0.000***	0.000***	0.000***
	(22.65)	(16.98)	(15.99)	(19.17)
No. rooms	0.146***	0.132***	0.129***	0.124***
	(10.98)	(12.99)	(13.25)	(15.21)
No. bedrooms	-0.030***	-0.023***	-0.021***	-0.014**
	(-2.91)	(-3.07)	(-2.85)	(-2.45)
No. bathrooms	-0.080***	-0.075***	-0.068***	-0.056***
	(-5.86)	(-5.88)	(-6.74)	(-5.71)
Floor no.	0.005	0.006	0.006	0.005
	(0.83)	(1.31)	(1.18)	(1.31)
Expenses		0.000	-0.000	-0.000***
		(0.43)	(-0.02)	(-2.68)
Heating expenses		-0.000***	-0.000***	-0.000***
		(-4.60)	(-4.84)	(-3.87)
Heating included		0.020	0.020	0.003
		(1.49)	(1.57)	(0.38)
Deposit		-0.000**	-0.000**	-0.000***
		(-2.41)	(-2.48)	(-3.78)
Rented			0.115***	0.084***
			(7.49)	(6.98)
Housing benefits			0.000***	0.000***
			(3.40)	(2.94)
Holiday property			-0.002***	-0.003***
			(-4.97)	(-8.66)
Missing indicators	No	Yes	Yes	Yes
Covariate distances	No	No	No	Yes
Time FE	Yes	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.63	1.63	1.63	1.63
SD(y)	0.38	0.38	0.38	0.38
Observations	1,606,962	1,606,962	1,606,962	1,606,962
Adjusted R^2	0.10	0.11	0.13	0.31

Table 4: Rent-to-price ratio and local demographic and economic factors

This table reports estimates from regressions for property-level rent-to-price ratios on selected district-level measures of demographic and economic conditions for a German sample between 2007 and 2017. The log-transformed rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. Each column augments the specification of column 4 of Table 3 with one district-level explanatory variable. The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$\ln(H/P)$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Old-to-working age ratio	0.021***						
	(7.62)						
ln(Population)		-0.033***					
		(-3.67)					
Disposable income per capita			-0.020***				
			(-3.09)				
GDP per capita				-0.003***			
				(-4.04)			
Unemployment rate					0.013**		
					(2.35)		
Manufacturing industry share						0.020***	
						(4.55)	
ln(No. businesses)							-0.046***
							(-4.69)
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.63	1.63	1.63	1.63	1.63	1.63	1.63
SD(y)	0.38	0.38	0.38	0.39	0.38	0.39	0.38
Observations	1,606,962	1,606,962	1,596,599	1,593,842	1,606,677	1,589,019	1,606,952
Adjusted R^2	0.34	0.31	0.32	0.32	0.32	0.32	0.32

Table 5: Rent-to-price ratio and local housing market conditions
This table reports estimates from regressions for property-level rent-to-price ratios on selected district-level measures of housing market conditions for a German sample between 2007 and 2017. The log-transformed rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. Each column augments the specification of column 4 of Table 3 with one district-level explanatory variable. The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$\ln(H/P)$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IQR of surface	-0.010***						
	(-6.41)						
No. of days online		0.006***					
		(10.04)					
ln(No. posted flats)			-0.023***				
			(-4.16)				
Housing stock per capita				0.009**			
				(2.54)			
Living space completed per capita					-0.499***		
					(-5.08)		
Land price						-0.000***	
						(-8.54)	
Property assessment rate							-0.000
							(-0.72)
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.63	1.63	1.63	1.63	1.63	1.64	1.63
SD(y)	0.38	0.38	0.38	0.39	0.39	0.39	0.38
Observations	1,606,962	1,606,962	1,606,962	1,596,654	1,592,340	1,473,808	1,606,196
Adjusted R^2	0.33	0.33	0.31	0.31	0.32	0.33	0.31

Table 6: Decomposition of the rent-to-price ratio and local conditions
This table reports estimates from regressions for property-level rent-to-price ratios (and the components thereof) on selected district-level characteristics for a German sample between 2007 and 2017. Column 1 uses as dependent variable the log-transformed rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. Column 2 (3) uses as dependent variable the log-transformed annual rental (sale) price per sqm listed for the matched flat for rent (for sale). The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$\ln(H/P)$	$\ln(H)$	$\ln(P)$
	(1)	(2)	(3)
Old-to-working age ratio	0.016***	-0.018***	-0.035***
	(7.18)	(-9.62)	(-10.55)
Disposable income per capita	-0.017***	0.029***	0.048***
	(-5.53)	(12.41)	(10.46)
Living space completed per capita	-0.308***	0.202***	0.527***
	(-4.95)	(3.55)	(5.11)
ln(No. businesses)	-0.035***	0.021***	0.059***
	(-3.36)	(4.08)	(4.76)
Flat covariates	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat
Mean(y)	1.63	4.41	7.37
SD(y)	0.39	0.30	0.56
Observations	1,581,967	1,589,856	1,589,724
Adjusted R^2	0.37	0.69	0.61

Table 7: Sources of variation of rent-to-price ratios
This table reports coefficients of determination from regressions for property-level rent-to-price ratios on different fixed effect structures for a German sample between 2007 and 2017. The log-transformed rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. Each column augments the specification of column 4 of Table 3 with progressively finer fixed effects (indicated below). The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$\ln(H/P)$					
	(1)	(2)	(3)	(4)	(5)	(6)
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	No	Yes	No	Yes	No
State FE	Yes	No	No	No	No	No
State×Time FE	No	Yes	No	No	No	No
District FE	No	No	Yes	No	No	No
District×Time FE	No	No	No	Yes	No	No
Zipcode FE	No	No	No	No	Yes	No
Zipcode×Time FE	No	No	No	No	No	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.63	1.63	1.63	1.63	1.63	1.63
SD(y)	0.38	0.38	0.38	0.38	0.38	0.38
Observations	1,606,962	1,606,962	1,606,960	1,606,096	1,606,683	1,593,367
Adjusted R^2	0.35	0.36	0.41	0.44	0.47	0.59

Table 8: The impact of potential matching errors on the variation of rent-to-price ratios
This table reports coefficients of determination from regressions for property-level rent-to-price ratios for a German sample between 2007 and 2017, exploring the impact of potential matching errors. Each column builds on the specification of column 4 of Table 3. In column 1, the dependent variable is the log-transformed actual rent-to-price ratio, as observed for a subsample of properties for sale. This specification, by construction of the dependent variable, does not control for traits observed only for flats for rent, for the differences between the matched flats, or for the corresponding missing value indicators. In columns 2-4, the dependent variable is the log-transformed rent-to-price ratio, as obtained via a matching procedure of flats for rent to flats for sale. Column 2 restricts the sample to properties for which also the actual rent-to-price ratio is available. Column 3 augments the specification with fixed effects for each percentile of a matching quality measure. Column 4 removes from the sample potential duplicate listings. The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	ln(Actual H/P)	ln(H/P)		
	(1)	(2)	(3)	(4)
Flat covariates	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes
Covariate distances	No	Yes	Yes	Yes
Zipcode × Time FE	Yes	Yes	Yes	Yes
Matching quality FE	No	No	Yes	No
Potential duplicate listings	Included	Included	Included	Excluded
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.78	1.74	1.63	1.63
SD(y)	0.40	0.37	0.38	0.38
Observations	329,471	327,417	1,593,367	1,346,486
Adjusted R^2	0.72	0.70	0.59	0.58

Table 9: Cross-sectional variation of rent-to-price ratios at different levels of geographic aggregation
This table shows summary statistics on cross-sectional standard deviations of rent-to-price ratios computed at different levels of geographic aggregation. The standard deviations are estimated at zipcode, district, and state-quarter level. The rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. The filtered rent-to-price ratio is the residual from a regression of non-log transformed rent-to-price ratios specified as in column 6 of Table 7. Suspect duplicate property listings are removed from the sample as well as geographic area-quarters with fewer than 30 observations. Refer to Appendix Table A.1 for variable definitions.

	Obs.	Mean	SD	Min	P25	Median	P75	Max
Zipcode-level aggregation								
SD(H/P)	12,693	1.681	0.698	0.273	1.191	1.538	2.007	5.896
SD(Filtered H/P)	12,693	1.434	0.615	0.341	1.008	1.280	1.683	5.958
District-level aggregation								
SD(H/P)	5,933	1.949	0.653	0.404	1.489	1.866	2.300	6.546
SD(Filtered H/P)	5,933	1.416	0.496	0.401	1.063	1.337	1.659	6.120
State-level aggregation								
SD(H/P)	663	2.188	0.524	0.867	1.821	2.115	2.491	4.557
SD(Filtered H/P)	663	1.443	0.370	0.636	1.203	1.381	1.657	3.515

Table 10: Summary statistics on the pseudo-panel
This table shows summary statistics on returns and rent-to-price ratios over the pseudo-panel between 2007 and 2017. Refer to Appendix Table A.1 for variable definitions.

	Obs.	Mean	SD	Min	P25	Median	P75	Max
r (%)	25,505	1.979	12.435	-41.044	-5.114	1.877	8.864	47.393
r^* (%)	25,505	0.778	12.518	-42.620	-6.380	0.682	7.725	46.575
r^e (%)	25,505	2.019	12.457	-40.735	-5.118	1.895	8.936	47.701
Δh (%)	25,505	0.355	4.308	-14.320	-2.151	0.263	2.816	16.341
H^q/P (%)	26,936	1.208	0.280	0.630	1.009	1.180	1.367	2.270

Table 11: Predictive regressions
This table reports estimates from predictive regressions for housing premium and rent growth on the log-transformed rent-to-price ratio. The regressions are estimated on a pseudo-panel constructed from a sample of German flats listed between 2007 and 2017. The unit of observation is at the cohort-calendar quarter level, where cohorts are defined by the district in which the flat is located, its number of rooms category, and its size category. The dependent variable in columns 1 to 3 (4 to 6) is the k-quarter ahead housing premium (rent growth), with k = 4, 8, 12. All specifications include cohort fixed effects. The t-statistics (in parentheses) are based on Driscoll-Kraay standard errors (number of lags equal to k). Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$r^e_{t+1\to t+k}$			$\Delta h_{t+1\to t+k}$		
	(1) $k=1$	(2) $k=4$	(3) $k=12$	(4) $k=1$	(5) $k=4$	(6) $k=12$
$\ln(H^q/P)$	0.328***	0.439***	0.595***	-0.048***	-0.104***	-0.168***
	(12.06)	(5.45)	(4.40)	(-16.64)	(-7.91)	(-4.46)
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Cohort-time	Cohort-time	Cohort-time	Cohort-time	Cohort-time	Cohort-time
Mean(y)	0.02	0.09	0.28	0.00	0.02	0.05
SD(y)	0.12	0.15	0.19	0.04	0.05	0.07
Observations	23,996	20,897	14,866	23,970	20,857	14,860
Within R^2	0.14	0.16	0.15	0.03	0.07	0.09

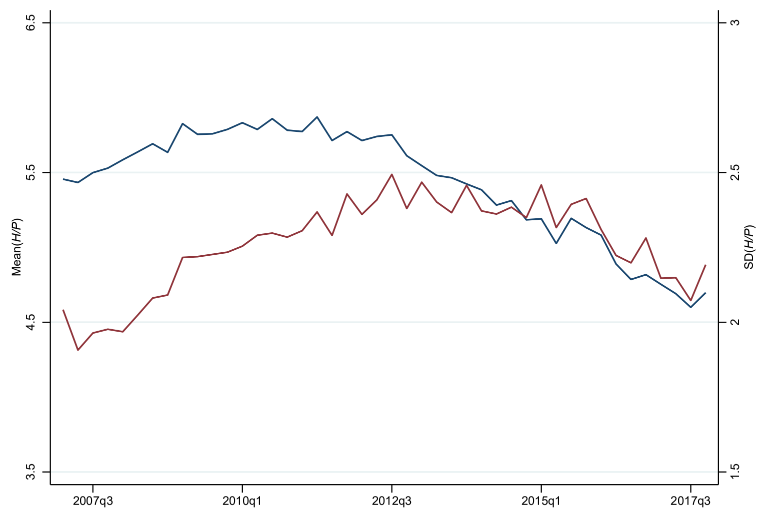
Table 12: Predictive regressions (global cities vs. other areas)

This table reports estimates from predictive regressions for housing premium and rent growth on the log-transformed rent-to-price ratio, conditional on the global status of the city where the properties are located. The regressions are estimated on a pseudo-panel constructed from a sample of German flats listed between 2007 and 2017. The unit of observation is at the cohort-calendar quarter level, where cohorts are defined by the district in which the flat is located, its number of rooms category, and its size category. The dependent variable in columns 1 and 2 (columns 3 and 4) is the 12-quarter ahead housing premium (rent growth). In odd columns (even columns), the sample is restricted to cohorts of flats located in (outside) global cities. All specifications include cohort fixed effects. The t-statistics (in parentheses) are based on Driscoll-Kraay standard errors (12 lags). Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$r^e_{t+1\to t+12}$		$\Delta h_{t+1\to t+12}$	
	(1) Global cities	(2) Other	(3) Global cities	(4) Other
$\ln(H^q/P)$	0.230	0.646***	-0.223***	-0.161***
	(1.20)	(5.16)	(-5.56)	(-4.33)
Cohort FE	Yes	Yes	Yes	Yes
Unit of observation	Cohort-time	Cohort-time	Cohort-time	Cohort-time
Mean(y)	0.34	0.27	0.09	0.04
SD(y)	0.15	0.19	0.07	0.07
Observations	1,706	13,160	1,603	13,257
Within R^2	0.03	0.17	0.19	0.08

Appendix for
"Housing Yields"

Panel A: Mean and SD trends



Panel B: Median and IQR trends

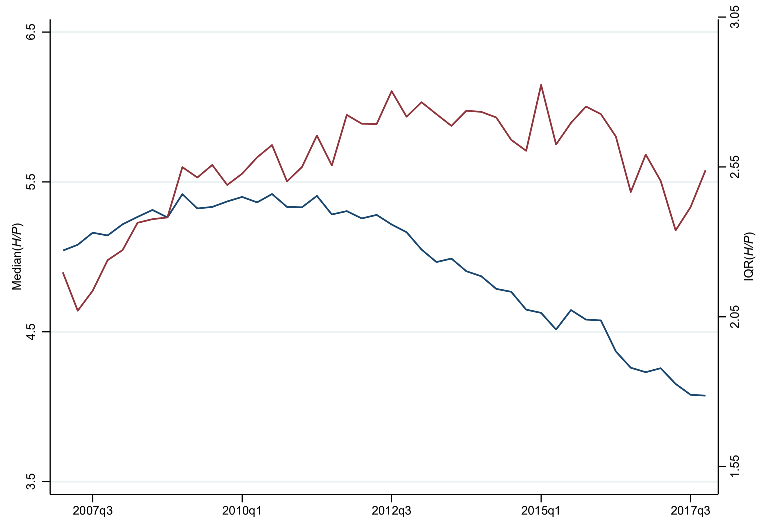


Figure A.1: Level and dispersion of rent-to-price ratios over time
This figure shows the evolution of the level and dispersion of rent-to-price ratios (obtained via matching) between 2007 and 2017. Panel A reports the mean and the standard deviation dynamics (blue and red line, respectively). Panel B reports the median and the interquartile range trends (blue and red line, respectively).

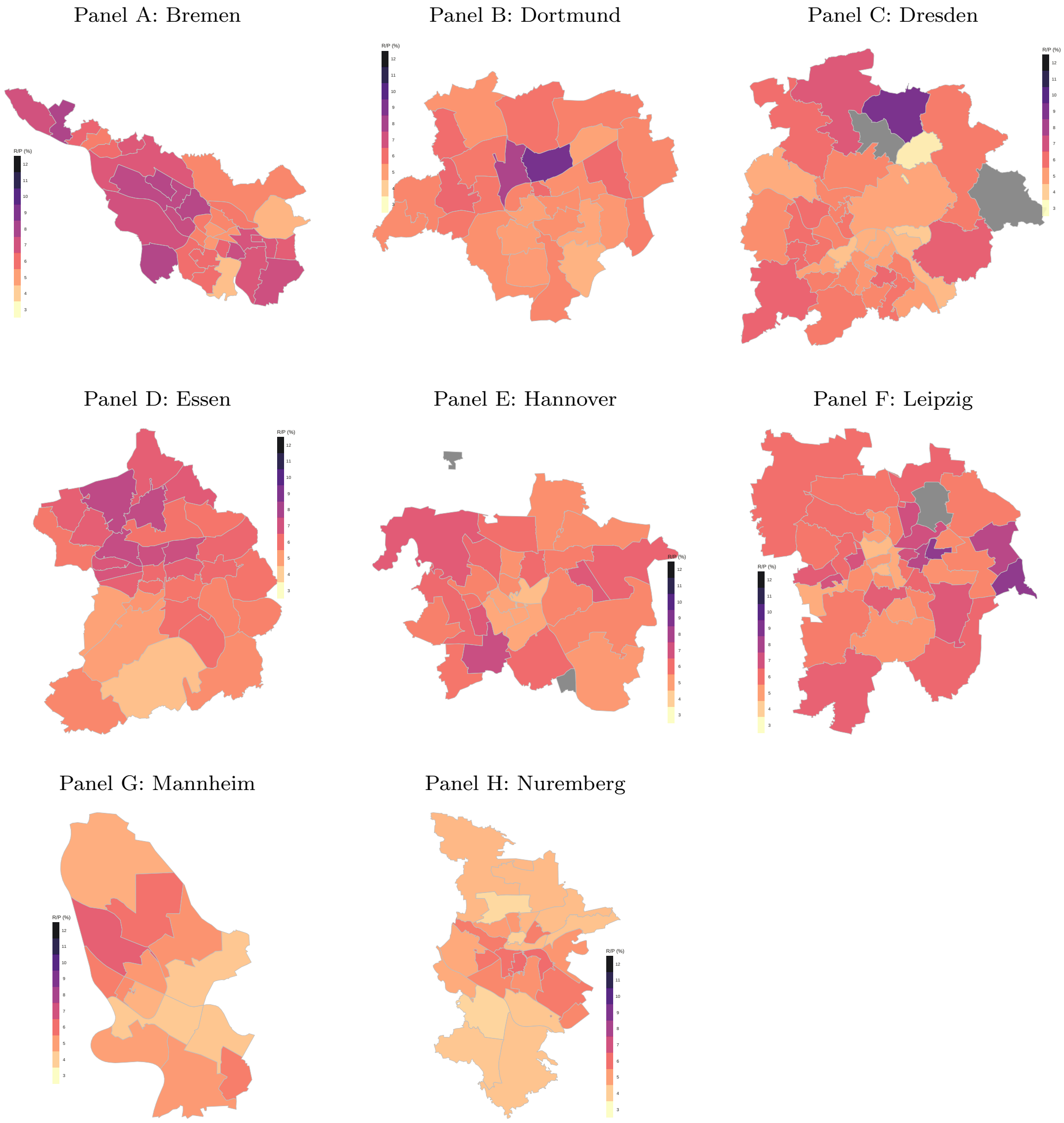


Figure A.2: Median rent-to-price ratio at the zipcode level within other major cities
This figure visualizes the median rent-to-price ratio (obtained via matching) at the five-digit zipcode level across selected German major cities, pooling all periods between 2007 and 2017. Grey-colored zipcode areas do not a sufficient number of observations. Reported cities cities are those assigned a "sufficiency" rating in the 2020 ranking by the Globalization and World Cities Research Network. Each of the panels from A to H corresponds to a different city.

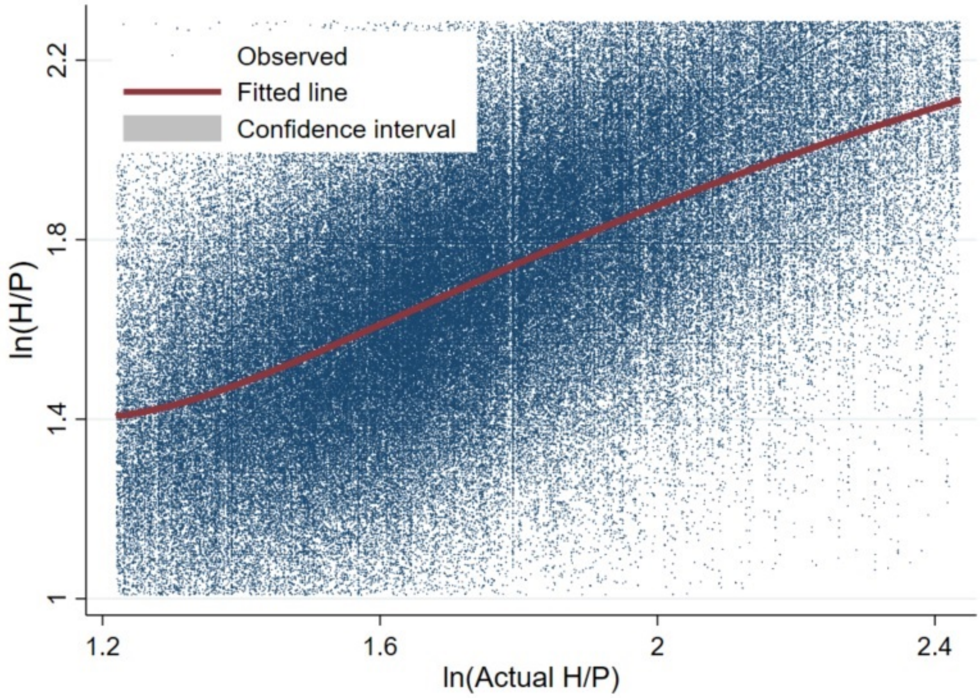


Figure A.3: Validation of rent-to-price ratios obtained via matching
This figure shows a scatter plot of rent-to-price ratios obtained via matching against their actual counterparts (all in natural logarithm). Actual rent-to-price ratios refer to properties on sale for which a rental income is reported. Both matched and actual rent-to-price ratios are trimmed at the 5% and 95% level in the figure to favor readability. A line fitted with a fractional polynomial is also plotted (together with 95% confidence bands).

Table A.1: Definition of variables

Variable	Databases	Definition
Flat characteristics		
Rent-to-price ratio (H/P)	RWI-GEO-RED	Ratio of the annual rent of a listed rental flat (H) to the sale price of a matched counterfactual flat for sale (P). Details on the matching procedure are provided in Section 2.1.
Annual rent (H)	RWI-GEO-RED	Listed annual rent exclusive of expenses in EUR. Available only for flats for rent.
Annual inclusive rent	RWI-GEO-RED	Annual rent inclusive of expenses in EUR. Available only for flats for rent.
Expenses	RWI-GEO-RED	Expenses for utilities in EUR. Available only for flats for rent.
Heating expenses	RWI-GEO-RED	Heating expenses in EUR. Available only for flats for rent.
Heating included	RWI-GEO-RED	Indicator equal to 1 if heating expenses are comprised in the inclusive rent, and 0 otherwise. Available only for flats for sale.
Deposit	RWI-GEO-RED	Deposit in EUR. Available only for flats for rent.
Sale price (P)	RWI-GEO-RED	Listed sale price in EUR. Available only for flats for sale.
Rented	RWI-GEO-RED	Indicator equal to 1 if the flat is rented, and 0 otherwise. Available only for flats for sale.
Housing benefits	RWI-GEO-RED	Housing benefits in EUR. Available only for flats for sale.
Holiday property	RWI-GEO-RED	Indicator equal to 1 if the flat can be used as a holiday property, and 0 otherwise. Available only for flats for sale.
Surface	RWI-GEO-RED	Surface of the flat in sqm.
No. rooms	RWI-GEO-RED	Number of rooms in the flat.
No. bedrooms	RWI-GEO-RED	Number of bedrooms in the flat.
No. bathrooms	RWI-GEO-RED	Number of bathrooms in the flat.
Floor no.	RWI-GEO-RED	Floor on which the flat is located.
No. building floors	RWI-GEO-RED	Number of floors of the building where the flat is located
Usable surface	RWI-GEO-RED	Usable surface of the flat in sqm.
Construction year	RWI-GEO-RED	Year of construction of the property.
Year last modernization	RWI-GEO-RED	Year in which the last modernization of the property took place.
Property conditions	RWI-GEO-RED	Categorical variable (eleven categories) indicating the conditions of the flat.
Quality of facilities	RWI-GEO-RED	Categorical variable (four categories) indicating the quality of the facilities in the flat.
Protected building	RWI-GEO-RED	Indicator equal to 1 if the flat is located in listed building, and 0 otherwise.
Energy consumption	RWI-GEO-RED	Annual energy consumption in kWh per sqm.
Energy rating	RWI-GEO-RED	Categorical variable capturing the rating of the flat based on the Energy Performance Certificate.
Hot water in energy consumption	RWI-GEO-RED	Indicator equal to 1 if hot water is included in energy consumption, and 0 otherwise.
Balcony	RWI-GEO-RED	Indicator equal to 1 if the flat has a balcony, and 0 otherwise.
Available parking	RWI-GEO-RED	Indicator equal to 1 if the flat comes with a parking place, and 0 otherwise.
Access with wheelchair	RWI-GEO-RED	Indicator equal to 1 if the flat is accessible with a wheelchair, and 0 otherwise.
Guest WC	RWI-GEO-RED	Indicator equal to 1 if the flat has a guest WC, and 0 otherwise.
Garden	RWI-GEO-RED	Indicator equal to 1 if the flat gives access to a garden, and 0 otherwise.
Cellar	RWI-GEO-RED	Indicator equal to 1 if the flat gives access to a cellar, and 0 otherwise.
Kitchen	RWI-GEO-RED	Indicator equal to 1 if a kitchen is already installed in the flat, and 0 otherwise.
Elevator	RWI-GEO-RED	Indicator equal to 1 if there is an elevator in the building in which the flat is located, and 0 otherwise.
Assisted living	RWI-GEO-RED	Indicator equal to 1 if the flat provides assisted living services, and 0 otherwise.
Immediate availability	RWI-GEO-RED	Indicator equal to 1 if the flat is immediately available, and 0 otherwise.
No. of days online	RWI-GEO-RED	Number of days the flat listing stays online on the platform.

(Continued)

Table A.1: – *Continued*

No. of hits	RWI-GEO-RED	Number of hits the flat listing on the online platform.
No. of clicks (contact button)	RWI-GEO-RED	Number of clicks on the “Contact” button of the flat listing.
No. of clicks (customer profile)	RWI-GEO-RED	Number of clicks on the customer profile linked to the flat listing.
No. of clicks (share button)	RWI-GEO-RED	Number of clicks on the “Share” button of the flat listing.
No. of clicks (customer URL)	RWI-GEO-RED	Number of clicks on the customer URL linked to the flat listing.
Housing return and rent growth		
r	RWI-GEO-RED	Quarterly logarithmic total return (i.e., reflecting also rental income) for the cohort of properties of the pseudo-panel dataset. Details on the construction of the pseudo-panel are provided in Section 2.1.
r^*	RWI-GEO-RED	Quarterly logarithmic ex-rent return for the cohort of properties of the pseudo-panel dataset. Details on the construction of the pseudo-panel are provided in Section 2.1.
r^e	RWI-GEO-RED, FRED	Quarterly logarithmic total return (i.e., reflecting also rental income) in excess of the nationwide 3-month interbank rate for the cohort of properties of the pseudo-panel dataset. Details on the construction of the pseudo-panel are provided in Section 2.1.
Δh	RWI-GEO-RED	Quarterly logarithmic rent growth rate for the cohort of properties of the pseudo-panel dataset. Details on the construction of the pseudo-panel are provided in Section 2.1.
H^q/P	RWI-GEO-RED	Ratio of the average quarterly rent per sqm to the average sale price per sqm within the cohort of properties of the pseudo-panel dataset. Details on the construction of the pseudo-panel are provided in Section 2.1.
District-level characteristics		
Old-to-working age ratio	Federal Statistical Office	Ratio of population older than 65 years to working age population (20-65 years) in a given district-year.
Population	Federal Statistical Office	Total population in a given district-year.
Disposable income per capita	Federal Statistical Office	Disposable income per capita in thousand EUR in a given district-year.
GDP per capita	Federal Statistical Office	GDP per capita in thousand EUR in a given district-year.
Unemployment rate	Federal Statistical Office	Unemployment rate in a given district-year.
Manufacturing industry share	Federal Statistical Office	Ratio of the number of manufacturing firms to the number of registered businesses across all industries in a given district-year.
No. businesses	Federal Statistical Office	Number of registered businesses across all industries in a given district-year.
IQR of surface	RWI-GEO-RED	Interquartile range of the surface of the flats listed on the online platform in a given district-calendar quarter.
Average online listing time	RWI-GEO-RED	Average number of days online of the flat listings on the platform in a given district-calendar quarter.
No. posted flats	RWI-GEO-RED	Number of flats listed on the online platform in a given district-calendar quarter.
Housing stock per capita	Federal Statistical Office	Residential housing stock per capita in sqm in a given district-year.
Living space completed per capita	Federal Statistical Office	Living space completed per capita in sqm in a given district-year.
Land price	Federal Statistical Office	Average ready-for-building land price per sqm in a given district-year.
Property assessment rate	Federal Statistical Office	Property tax multiplier (type B) in a given district-year.

(Continued)

Table A.1: – *Continued*

Global city	Globalization and World Cities Research Network	Indicator equal to one if a district has a rating between "Alpha" and "Gamma".
Household consumption		
Δc^{nd}	Federal Statistical Office	Annual logarithmic growth rate of district-level household waste production per capita (in the spirit of Savov, 2011).
Δc^{d}	Federal Motor Transport Authority, Federal Statistical Office	Annual logarithmic growth rate of passenger vehicles registrations (by employees or inactive population) per capita (in the spirit of Mian et al., 2013).
Δc^{h}	Federal Statistical Office	Annual logarithmic growth rate of district-level housing stock per capita (in the spirit of Clark, Deurloo, and Dieleman, 2000).

Table A.2: Rent-to-price ratio and local conditions (further tests)
This table reports estimates from regressions for rent-to-price ratios (and the components thereof) on selected district-level characteristics for a German sample between 2007 and 2017. In Panel A, the log-transformed rent-to-price ratio is obtained via a matching procedure of flats for rent to flats for sale. Column 1 (column 2) uses its district-calendar quarter-level mean (standard deviation) as dependent variable. Specifications in Panel B are estimated on a pseudo-panel. The unit of observation is at the cohort-calendar quarter level, where cohorts are defined by the district in which the flat is located, its number of rooms category, and its size category. Column 1 uses the log-transformed annual rent-to-price ratio as the dependent variable. Columns 2 and 3 use the log-transformed annual rental and sale price per sqm as the dependent variable, respectively. The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

Panel A: District-level

	Mean($\ln(H/P)$)	SD($\ln(H/P)$)
	(1)	(2)
Old-to-working age ratio	0.011***	0.001
	(6.34)	(1.56)
Disposable income per capita	-0.031***	-0.003**
	(-9.06)	(-2.10)
Living space completed per capita	-0.416***	-0.067***
	(-8.06)	(-3.58)
ln(No. businesses)	0.004	0.032***
	(0.46)	(8.25)
Time FE	Yes	Yes
Unit of observation	Match. flat	Match. flat
Mean(y)	1.65	0.31
SD(y)	0.27	0.11
Observations	13,938	13,093
Adjusted R^2	0.30	0.05

Panel B: Pseudo-panel

	$\ln(H/P)$	$\ln(H)$	$\ln(P)$
	(1)	(2)	(3)
Old-to-working age ratio	0.009***	-0.026***	-0.035***
	(4.23)	(-10.40)	(-8.86)
Disposable income per capita	-0.009**	0.043***	0.050***
	(-2.00)	(10.40)	(6.74)
Living space completed per capita	-0.395***	0.157*	0.605***
	(-5.73)	(1.86)	(4.26)
ln(No. businesses)	-0.028	0.064***	0.091***
	(-1.55)	(5.45)	(4.19)
Time FE	Yes	Yes	Yes
Unit of observation	Cohort-time	Cohort-time	Cohort-time
Mean(y)	1.55	3.00	7.44
SD(y)	0.23	0.23	0.37
Observations	26,411	26,529	26,537
Adjusted R^2	0.23	0.62	0.56

Table A.3: Re-estimating the main specifications with non-transformed rent-to-price ratios
This table re-estimates the specifications from column 4 of Table 3, column 1 of Table 6, and column 6 of Table 7, using non-transformed property-level rent-to-price ratio as the dependent variable for a German sample between 2007 and 2017. The t-statistics (in parentheses) are based on standard errors clustered by district. Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	H/P		
	(1)	(2)	(3)
Surface	-0.104***	-0.118***	-0.117***
	(-13.60)	(-24.70)	(-25.26)
Surface squared	0.000***	0.000***	0.001***
	(14.19)	(21.63)	(23.97)
No. rooms	0.655***	0.454***	0.154***
	(13.21)	(10.46)	(4.77)
No. bedrooms	-0.084**	0.067**	0.082***
	(-2.17)	(2.02)	(4.20)
No. bathrooms	-0.310***	-0.301***	-0.275***
	(-5.86)	(-7.19)	(-5.76)
Floor no.	0.042**	0.041**	0.048***
	(2.13)	(2.39)	(4.25)
Expenses	-0.001**	-0.000	0.000
	(-2.22)	(-0.11)	(0.81)
Heating expenses	-0.002***	-0.002***	-0.000
	(-3.15)	(-2.75)	(-1.47)
Heating included	0.023	-0.009	0.070***
	(0.38)	(-0.29)	(4.47)
Deposit	-0.000***	0.000***	0.001***
	(-4.51)	(5.41)	(11.54)
Community charge	0.001**	0.001***	0.001***
	(2.56)	(4.33)	(5.02)
Holiday property	-0.016***	-0.018***	-0.016***
	(-6.72)	(-8.97)	(-6.91)
Rented	0.400***	0.330***	0.358***
	(5.78)	(5.35)	(6.59)
Old-to-working age ratio		10.306***	
		(6.25)	
Disposable income per capita		-0.000***	
		(-6.40)	
Living space completed per capita		-1.736***	
		(-4.79)	
ln(No. businesses)		-0.182***	
		(-2.94)	
Missing indicators	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes
Time FE	Yes	Yes	No
Zipcode×Time FE	No	No	Yes
Unit of observation	Match. flat	Match. flat	Match. flat
Mean(y)	5.52	5.52	5.52
SD(y)	2.30	2.30	2.30
Observations	1,606,962	1,581,967	1,593,367
Adjusted R^2	0.25	0.32	0.55

Table A.4: Housing yield components and consumption growth
This table shows the distribution of cohort-level correlations of housing premia and rent growth with different components of per capita consumption growth over the pseudo-panel between 2007 and 2017. At least 8 non-missing observations of both variables are required to compute each cohort-level correlation. Refer to Appendix Table A.1 for variable definitions.

	Obs.	Mean	SD	Min	P25	Median	P75	Max
Corr$(r^e, \Delta c_{nd})$	479	-0.034	0.360	-0.880	-0.299	-0.047	0.246	0.823
Corr$(r^e, \Delta c_d)$	487	-0.075	0.335	-0.867	-0.303	-0.070	0.158	0.859
Corr$(r^e, \Delta c_h)$	474	-0.060	0.337	-0.804	-0.312	-0.073	0.179	0.818
Corr$(\Delta h, \Delta c_{nd})$	459	-0.004	0.372	-0.883	-0.282	0.010	0.254	0.861
Corr$(\Delta h, \Delta c_d)$	468	0.011	0.312	-0.832	-0.219	0.002	0.235	0.792
Corr$(\Delta h, \Delta c_h)$	457	-0.136	0.294	-0.874	-0.347	-0.141	0.080	0.716

Table A.5: Predictive regressions controlling for local conditions
This table reports estimates from predictive regressions for housing premium and rent growth on the log-transformed rent-to-price ratio, controlling for selected district-level characteristics. The regressions are estimated on a pseudo-panel constructed from a sample of German flats listed between 2007 and 2017. The unit of observation is at the cohort-calendar quarter level, where cohorts are defined by the district in which the flat is located, its number of rooms category, and its size category. The dependent variable in columns 1 to 3 (4 to 6) is the k-quarter ahead housing premium (rent growth), with $k = 4,\ 8,\ 12$. All specifications include cohort fixed effects. The t-statistics (in parentheses) are based on Driscoll-Kraay standard errors (number of lags equal to k). Significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Refer to Appendix Table A.1 for variable definitions.

	$r^e_{t+1\to t+k}$			$\Delta h_{t+1\to t+k}$		
	(1) $k=1$	(2) $k=4$	(3) $k=12$	(4) $k=1$	(5) $k=4$	(6) $k=12$
$\ln(H^q/P)$	0.372***	0.516***	0.693***	-0.049***	-0.094***	-0.153***
	(17.22)	(7.48)	(6.72)	(-17.55)	(-10.13)	(-4.81)
Old-to-working age ratio	0.016***	0.040***	0.081***	0.001***	0.007***	0.025***
	(6.45)	(6.48)	(12.06)	(3.16)	(3.98)	(10.40)
Disposable income p.c.	0.016**	0.019	-0.035**	-0.000	0.006**	0.002
	(2.09)	(1.13)	(-2.23)	(-0.52)	(2.68)	(0.28)
Living space completed p.c.	0.165***	0.236**	0.038	-0.017***	0.003	-0.001
	(4.83)	(2.57)	(0.33)	(-2.80)	(0.15)	(-0.01)
ln(No. businesses)	0.039	0.484	1.908***	0.008	0.136*	0.254*
	(0.23)	(1.03)	(3.89)	(0.33)	(1.96)	(1.80)
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Cohort-time	Cohort-time	Cohort-time	Cohort-time	Cohort-time	Cohort-time
Mean(y)	0.02	0.09	0.28	0.00	0.02	0.05
SD(y)	0.12	0.15	0.19	0.04	0.05	0.07
Observations	23,547	20,522	14,634	23,516	20,472	14,611
Within R^2	0.17	0.23	0.31	0.03	0.09	0.16