ISSN 2279-9362

Housing Yields

Stefano Colonnello, Roberto Marfè and Qizhou Xiong

No. 716
February 2024

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Housing Yields*

Stefano Colonnello† Roberto Marfè‡ Qizhou Xiong§

This Draft: February 5, 2024

First Draft: February 16, 2021

Abstract

We build a granular dataset of residential property yields using rental and sale listings from a major German real estate platform. Equipped with more than 1.5 million property-level rent-to-price ratios, we document a novel *heterogeneity puzzle*. About one-third of dispersion in yields can be explained neither by an extensive array of property-specific observable features, nor by accounting for any possible below-zip code-level time-varying factor through a rich fixed effects structure. Unexplained yield heterogeneity is sizable and economically significant. Whereas property yields predict returns and rent growth rates, we show that their time-series variation largely originates at a highly local level. Our evidence may point to the importance of heterogeneity in investors' beliefs and preferences, as opposed to a battery of alternative explanations for which we directly test.

JEL Classification: G12, G51, R31

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

*We thank Ufuk Akcigit, Pietro Dindo, Dmitry Kuvshinov (discussant), Francesco Lippi, Yiran Mao (discussant), Juan S. Morales, Tommaso Oliviero, Loriana Pelizzon, Alberto Plazzi, Geert Rouwenhorst, Marti G. Subrahmanyam, 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, the 2021 Paris December Finance Meeting, the workshop "The socioeconomics of housing and finance" at FU Berlin, and the "Finance workshop" at Ca' Foscari University of Venice for insightful comments. We are grateful to staff at FDZ Ruhr of RWI Essen for granting us access to the RWI-GEO-RED database. The views in the paper are those of the authors and do not necessarily reflect those of Immo Capital or its staff.

†Corresponding author. Ca' Foscari University of Venice and Halle Institute For Economic Research (IWH), Cannaregio 873, Fondamenta San Giobbe, 30121 Venezia, Italy. Phone: +39 041 234 9198. E-mail: stefano.colonnello@unive.it.

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

§Immo Capital, United Kingdom. E-mail: leo.xiong@immo.capital.

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 homeownership 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; Guren, McKay, Nakamura, and Steinsson, 2021) 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 are a metric widely used by investors and policy makers to gauge the conditions of the housing market because it incorporates 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 variation in housing yields can be explained neither by zip code-level time-varying factors nor by a rich set of property-level characteristics. Such heterogeneity is economically sizable, with the the 90th−10th per-

[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 homeownership (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. Kohl (2016) and Kohl and Sørvoll (2021) provide an historical perspective on the roots of the homeownership gap between Germany and other countries like the US and the Nordic ones.

[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 measure 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).

centile range of unexplained yields in our dataset corresponding to a variation of EUR 73,729 in the value of the median flat, against a mean household net wealth estimated at EUR 202,541 according to Deutsche Bundesbank (2019). Whereas neighborhood amenities explain away a non-negligible part of this residual heterogeneity, a large fraction of it may stem from unobservable property-specific traits as well as from dispersion in investors' belief and preferences. These forces, in turn, presumably drive heterogeneity in both risk premia and expected rent growth rates. By contrast, local agglomeration effects, housing supply rigidities, informational frictions, and regulation do not appear to command sizable variation in housing yields over and above zip code-level time-varying factors.

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 gross 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. The rent-to-price ratios so obtained vary greatly across regions and their dispersion is remarkable not only across states or districts, but even across zip codes within the same city. Time-series variation in rent-to-price ratios is largely property-specific or at the zip code level, with macroeconomic fluctuations accounting only for about one-third of it.

At the same time, cross-sectional dispersion in housing yields is remarkably stable with respect to the level of regional aggregation considered (federal state-, district-, or zip code-level). Even after controlling for an extensive set of observable property-level characteristics and fine fixed effects at the zip code-calendar quarter level in a regression framework, unexplained heterogeneity in yields remains large at 41% based on the adjusted R^2. Such a heterogeneity originates from a remarkably symmetric distribution of residuals, in which outliers appear to have a limited role.

Given the high degree of regional segregation of the housing market, we seek to pin down the part of variation in rent-to-price ratios captured by local fixed effects to precise economic factors, focusing on local economic and social conditions as plausible determinants. We find that district-level demographics, industry and economic fundamentals, rigidities in housing supply, and liquidity and size of the housing market co-move significantly with valuation ratios. However, these specific forces span only 42% of district-level variation in yields. Furthermore, 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.

Then, we verify that neither potential matching errors inherent to our synthetic yields nor the unobservability—in our setting—of costs borne by owners are likely to drive the 41% of variation that we cannot explain. We also implement an alternative procedure improving the matching precision with respect to the location of properties. At the cost of obtaining a significantly smaller sample, we are indeed able to include an even finer layer of fixed effects capturing neighborhood-level time-varying factors. As a result, we can explain an additional 12% of variation in yields, which speaks to the importance of local amenities. By contrast, we do not detect any quantitatively relevant role of agglomeration economies, informational frictions, rigidities in housing supply, or regulation in explaining within-zip code-quarter dispersion in yields.

The 41% of variation in rent-to-price ratios—or 29% when accounting for neighborhood amenities over the alternative matched sample—that we cannot explain is noteworthy, even more so if we think that observable property- and district-level characteristics deliver an adjusted R^2 of just 36%. In other words, leaving aside ultimately unobservable (at least in our dataset) factors subsumed in fixed effects, we are not able to trace back to precise economic forces more than 60% of dispersion in yields. Taking one step back, 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 zip code- or neighborhood-level traits (e.g., distance from schools, hospitals or restaurants, quality of local services, number of nearby shops, etc.), one may expect that rent-to-price ratios exhibit little variation.[4]

From an asset pricing perspective, a property's rent-to-price ratio should respond to changes in expected returns and rent growth (e.g., Plazzi et al., 2010). Although theory does not predict that yields should be equal across properties, our results pose a challenge about the drivers of house valuations. The substantial variation in yields at the very local level after accounting for a host of property-level traits may stem from heterogeneity in properties' risk premia and/or expected rent growth. Dispersion in investors' beliefs and preferences, in this respect, emerges as a possible crucial force.

Ideally, one would observe single properties over time to investigate to what extent

[4]Moreover, we document that the total variation of housing yields is comparable to, if not smaller than, that of yields for a widely researched asset class like US equities, pointing to the importance of the granular structure of our data and the documented unexplained heterogeneity.

dispersion in risk premia and rent growth expectations—and in investors' beliefs and preferences about properties—induce substantially different yields after filtering out the impact of observable characteristics. Previous studies show that, possibly due to underdiversification of housing investors, idiosyncratic risk is priced in housing returns (Eiling, Giambona, Lopez Aliouchkin, and Tuijp, 2020; Eichholtz, Korevaar, Lindenthal, and Tallec, 2021), posing a question with regards to the extent to which within-neighborhood variation in expected returns emanates from properties' idiosyncratic risk as opposed to their heterogeneous exposure to systematic risk factors. At the same time, investors' heterogeneity could produce sizable dispersion in yields, possibly leading to biases in prices in the presence of market frictions. We leave the identification of these economic mechanisms for future research. At each point in time, in fact, our matching approach readily provides property-level rent-to-price ratios that are informative about expected returns and rents, and that can be of interest to investors and policy makers, but it does not allow to track properties over time because we only observe listings as repeated cross-sections. More generally, computing comprehensive asset-level returns and rent growth rates is challenging for residential properties that typically transact rarely.

Against a substantial loss in terms of local granularity, we can nonetheless cross-validate our yield measure by means of a time series analysis. Specifically, we 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 indeed predict future excess returns and rent growth, in line with the theory and existing evidence (Plazzi et al., 2010). By controlling for aggregate shocks, we illustrate that local fluctuations in expectations significantly contribute to predictability. These results further corroborate the main analysis based on synthetic property-level rent-to-price ratios.

Finally, it is worth discussing the generalizability of our findings. The German housing market has a number of peculiarities that set it apart from those of other advanced economies, such as the US and the UK. In particular, homeownership is particularly low and unequally distributed; private contracts dominate the rental market over our sample period, against the backdrop of the declining importance of—historically widespread—social housing (e.g., Dustmann, Fitzenberger, and Zimmermann, 2022; Voigtländer, 2009); property taxation is comparatively low (e.g., Bach and Eichfelder, 2021). Moreover, real house prices exhibited a distinctively stable pattern over the last three decades, without a boom-bust cycle around the 2007–2008 financial crisis (e.g., Dustmann et al., 2022). Possibly as a result of these market features, yield dispersion is arguably lower in Germany than in other markets—Voigtländer (2009) show it is indeed the case for aggregate time-series variation. Moreover, flats—the type of properties we focus on—are probably characterized by smaller yield dispersion than detached houses or commercial properties. Then, the fraction of unexplained variation in yields we find may qualify as a lower bound. Put differently, the heterogeneity puzzle is probably more pronounced in other economies and/or other segments of the real estate market.

Related literature. 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). We mainly follow the latter—the standard in the finance literature—that regards 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 et al. (2021), relying on historical data from Paris and Amsterdam, highlight the primary role of property-level yields in explaining total housing returns.[5,6] A

[5]Eichholtz et al. (2021) also study housing yields dispersion. Differently from our 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.

[6]Other studies consider long-time series on the housing market. Eichholtz, Korevaar, and Lindenthal (2023) use more than 500 years of data to construct rental price indexes for seven European cities, documenting that they exhibited modest growth. Amaral, Dohmen, Kohl, and Schularick (2022) analyze returns for 27 large cities around the world over 150 years, showing that properties located in those cities are safer than those from more peripheral areas.

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 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 zip code-level US data on housing returns, Eiling et al. (2020) document relevant dispersion across MSAs with respect to which sources of risk are priced in housing returns, with idiosyncratic risk playing a sizable role. 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). Closer to this paper is Kantak (2022), 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. Then, Feng, Jaimovich, Rao, Terry, and Vincent (2023) demonstrate the importance of local industry structure—and of the manufacturing share, in particular—for explaining inequality in housing valuations. 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 and rent fluctuations in the US (e.g., Mian and Sufi, 2009; Duca, Muellbauer, and Murphy, 2011; Saadi, 2020; Reher, 2021). 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, Kaat, Ma, and Rebucci, 2021). More relevant for the German case are potentially 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. Gyourko, Mayer, and Sinai (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. Unequal access to amenities in different locations and the ensuing spatial sorting of households are another potentially important mechanism behind dispersion in housing valuations (e.g., Couture and Handbury, 2020; Couture, Gaubert, Handbury, and Hurst, 2021). Thanks to the size and granularity of our dataset, we complement this body of work by examining the impact of within-city agglomeration economies and neighborhood amenities on the heterogeneity of housing yields.

From a methodological perspective, this paper adds to a strand of research aimed at recovering rent-to-price ratios for single properties or for granular geographic areas. A number of studies use a matching approach similar to ours over more limited datasets. Smith and Smith (2006) focus on ten US metropolitan statistical areas (MSAs) in 2005. Bracke (2015) focuses on the London area between 2006 and 2013 and implements an algorithm that matches rental and sale information for the very same property. Clark and Lomax (2020) apply the same methodology to a sample of listings from England in 2014-2015. The approach used in these two studies is highly precise but restricts the analysis to properties that transact both on the sale and the rental market within a short period of time, thus potentially increasing sample bias. Hill and Syed (2016) adopt instead hedonic imputation to obtain property-level rent-to-price ratios for the Sydney area in Australia. Again for the Sydney area, Waltl (2018) combines the matching and the hedonic imputation approach to develop quality-adjusted rent-to-price ratios over the cross-section. Ahlfeldt, Heblich, and Seidel (2022) use the same listing data for Germany as in this paper to develop rental and sale price indexes for arbitrarily defined geographic areas, highlighting some stylized facts about ongoing trends in rent-to-price ratios across such areas and calling for more research on their drivers. Our paper responds to this call by looking into the determinants of dispersion of property-level rent-to-price ratios for a wide sample of matched properties.

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 speaks to 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, or location (district, municipality, zip code, 1-km raster cell). We also remove observations with sale (monthly rental) price below EUR 10,000 (EUR 50) and surface below 10 square meters (sqm). We exclude observations in the top 0.5% of price, surface, number of rooms, bathrooms, bedrooms, and floor number. As a result of this screening, the sample goes down to 5,100,753 rental listings, and 1,982,461 sale listings.

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 some 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. Indeed, although transaction prices are a more credible gauge of property valuations, it is not uncommon to supplement them with listed prices to build longer time series (e.g. Amaral et al., 2022). Throughout the paper, we consistently control for property-level demand pressure—as proxied by the number of visits, clicks, and time online of the listings—to account for those listed prices that are more likely to be upward biased. Below, we also validate our measure of rent-to-price ratios against actual rent-to-price ratios available for a subset of properties for sale that are already rented.

If we abstract from these relatively infrequent cases, the rent-to-price ratio of a given property is an inherently unobservable quantity. We use multiple approaches to compute rent-to-price ratios when we do not simultaneously observe the rental and sale price of the same property.

2.1.1 Matching approach

We follow different matching schemes to obtain counterfactual sale prices for flat for rent. Equipped with such a counterfactual, we then compute synthetic rent-to-price ratios.

Baseline matching scheme. For the baseline analysis, we adopt a parsimonious set of matching covariates aimed at ensuring a large number of matches, namely: flat surface (distance minimization), number of rooms, number of bedrooms, number of bathrooms, floor category, five-digit zip code, 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. Note that we do not observe whether a flat is furnished or not, potentially a key information for rent determination. However, the German rental market is largely dominated by unfurnished flats, which should mitigate this limitation. At the same time, we do not match on the year of construction of the property or for the quality of facilities because this pieces of information, while important for housing valuations, are often missing in our sample. Nonetheless, below we perform robustness tests in which we discard those matches characterized by sizable distances in terms of construction year or quality of facilities.

Alternative matching schemes. We then implement two alternative, more restrictive matching exercises. In the first one, we augment the baseline exact matching covariates with a categorical variable capturing the conditions of the property–going from flats needing renovation to new ones. Requiring that the flat for rent and the matched one for sale are in the same conditions allows us to better control for quality differences.

Second, whereas in Germany five-digit zip codes are pretty narrowly defined—especially in large cities—, in several instances they cut through different municipalities or even different labor market regions (as defined by Kosfeld and Werner, 2012). In an alternative scheme, we thus impose a more precise matching in terms of flat location. Specifically, we exploit geo-referenced information on the location of properties at the level of 1-km raster cells, which we supplement with information on the five-digit zip code, municipality, and labor market region of properties. By considering all possible combinations of such location traits, we define highly granular geographic areas. Imposing exact matching on these areas greatly improves the precision in terms of property location, reducing the errors in rent-to-price ratios stemming from potential differences in neighborhood-level amenities within matched flats. The dataset resulting from this alternative matching scheme features matched properties from over 20,000 of these highly granular geographic areas as opposed to less than 5,000 different zip codes. Based on Germany's population of 81.7M as of 2015, these geographic areas have on average around 4,000 inhabitants, a size that is pretty close to that of US census tracts, but, unlike the latter, are not specifically defined to be homogeneous in terms of socioeconomic characteristics.[7] Nonetheless, forcing matched between properties to be in the same narrow area should ensure that they have access to similar amenities (e.g., restaurants, shops, entertainment, parks), which have been shown to be important determinants of households' location choices and house prices (e.g., Couture and Handbury, 2020; Couture et al., 2021).

Our two alternative matching schemes unambiguously improve matching precision, but pose a trade-off when it comes to sample bias by restricting the analysis to those flats for which a very high quality match is available. Because of this, below we use these

[7] See, e.g., https://www2.census.gov/geo/pdfs/education/CensusTracts.pdf.

alternative matched datasets only for supplementary analyses.

Admittedly, any matching scheme relies on the assumption that properties for rent and for sale are ultimately comparable. However, the choice to rent or to sell a flat is not random but is instead likely to reflect also its unobservable characteristics. Hence, we conduct additional tests on a subsample of properties for which both the rental and the sale price are observable.

Computation of housing yields. For each of the different matching schemes, we remove any 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 (resp., maximum) number of matches is 1.56 (resp., 20).[8] 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. Following the notation of Plazzi et al. (2010), $H_{r,t}$ is the monthly rent exclusive of 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 (e.g., utilities), which we nonetheless use for robustness purposes below. 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.[9]

It is worth pointing out that the valuation ratio in equation (1) is a gauge of the owner's *gross* yield on the property because it does not reflect maintenance costs (not covered by the RWI-GEO-RED database), which can be sizable and typically increase the volatility of the owner's income stream (Chambers et al., 2021). By the same token, we are not able to reliably adjust the gross yield for taxation on the property itself, inheritance, capital gains, or rental income (e.g., in a progressive income tax system such

[8] We remove the few properties with more than 20 matches.

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

as the German one, the effective tax rate on rental income depends on the overall—and unobservable to us—income of the specific owner). Yet, below we also conduct robustness tests on *net* yield measures accounting for nontax costs, vacancy losses, and property taxes, which we approximate by means of regional variables and available estimates from the literature (e.g., Chambers et al., 2021; Eichholtz et al., 2021; Demers and Eisfeldt, 2022).

2.1.2 Pseudo-panel approach

As noted above, the RWI-GEO-RED database is de facto (repeated) cross-sectional in nature. Hence, a relevant drawback of the 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, we resort to a pseudo-panel in the spirit of Deaton (1985) by creating cohorts of properties. Whereas our main analysis in Section 3 builds on the synthetic rent-to-price ratios described above, in Section 4 we conduct complementary time-series analyses using such pseudo-panel returns.

To obtain unbiased estimates from a pseudo-panel analysis, the cohorts must be defined using time-invariant attributes that are observed in all periods for 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.[10] Despite having a large number 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

[10] Whereas the location of a property is unambiguously time-invariant, its number of rooms or surface 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.

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 than three rooms (1.06 % of the sample). Removing rooms-surface combinations with few observations (accounting for less than 2% of total observations) leaves us 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 combinations. Besides the sheer difference in population, another reason is that we focus on flats instead of houses, which naturally tilts the 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.[11] 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 (corresponding to 5,236,418 micro-level listings).

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

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

growth over the same horizon:

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

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

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

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.

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.[12]

The housing market is often regarded as highly segmented 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 categorized as rural districts (*Landkreis*) or urban districts (*Stadtkreis* or *Kreisfreie Stadt*).

To ensure consistency of regional variables, we account for those instances in which districts changed codes 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 Federal Reserve Economic

[12] 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.

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 synthetic rent-to-price ratios from equation (1) at the matched property-level. Table 1 reports summary statistics for flat characteristics under the baseline matching approach (Panel A) and district-level variables (Panel B).[13] The final sample in Panel A contains 1,613,889 flats for rent matched to counterfactual flats for sale.[14] By construction, the differences between 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 in magnitude. Nonetheless, below we augment rent-to-price ratio regression 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 synthetic rent-to-price ratio 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 in-

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

[14] The number of observations for matched flats for sale is mechanically equal to that of flats for rent. As we allow for matching with replacement, the number of distinct sale listings that are actually used is lower at 753,386 units, reflecting their lower frequency in the raw data relative to flats for rent. In our baseline matching exercise, the average (median) sale listing is used 3.6 (2) times as a match.

crease around the 2008-2009 recession, and displays an increasing trend in within-period heterogeneity.[15] 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 whereas around the European debt crisis a mere 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, possibly 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. 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., focusing on Bavaria, the Munich area vs. the districts on the Czech border). If we zoom in on the seven "global" German cities—i.e., those with an advanced service sector and that serve as hubs of international transportation networks—and consider variation at the zip code level in Figure 3, we observe that median ratios greatly vary even within some of the most thriving metropolitan areas like Hamburg or Frankfurt.[16]

In the spirit of Piazzesi, Schneider, and Tuzel (2007), we decompose time-series variation of valuation ratios into zip code-, district-, federal state-, and national level variation. For each geographic area, we compute the average rent-to-price ratio quarter-by-quarter so to obtain a time-series, for which we estimate standard deviation and interquartile range. Table 2 reports the average of time-series standard deviation and interquartile range estimates across geographic units. Similarly to Piazzesi and Schneider (2016), we find that national variation in prices only explains around one-third (two-fifths) of zip code-level standard deviation (interquartile range). These results do not change qualitatively whether, in computing average measures of variation, we weight geographic areas equally or by their number of listed properties. While we cannot evaluate property-specific time variation in ratios as our data are not longitudinal, we can still compare the average time-series standard deviation across zip code areas (around 1%) against property-level

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

[16] 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.

overall standard deviation of rent-to-price ratios (around 2%), which suggests that idiosyncratic patterns in pricing are pervasive. The decomposition of yield variation points to the relevance of looking at single properties within fine geographic subdivisions to investigate house pricing, which resonates with the above finding that cross-sectional heterogeneity is prevalent. Our analysis below centers on providing bounds to the fraction of variation in yields that can be explained by local as opposed to idiosyncratic factors.

Our estimates of variation in rent-to-price ratios for Germany can be then compared to those from other countries. Halket, Loewenstein, and Willen (2023), in an exercise akin to our Table 2, compute zip code-level time-series standard deviations of yields on single-family homes (net of property taxes) for 21 US cities, obtaining estimates that range between 0.60% and 2.21%. Our zip code-level estimate for Germany is around 1%, which hints at a similar degree of variability in yields, but on the low-end of the distribution across US cities. Also the levels of our yield estimates appear to be comparable to the US ones: Demers and Eisfeldt (2022) estimate an average yield (net of property taxes, maintenance costs, and vacancy losses) of 4.2% for major cities; the average yields obtained by Halket et al. (2023) range between 4.41% and 8.62%. The average gross yield in our sample is 5.53%. Below, when adjusting for likely—yet unobservable to us—costs, we obtain an average net yield of around 3.5%, only slightly below US levels. Moving to other economies, Voigtländer (2009) provides a direct comparison, based on aggregate data, between Germany and selected European countries (UK, Netherlands, and Spain) for the period 1970-2009, illustrating that time-series variability of both returns and yields is lowest in Germany, which, contrary to other countries, experienced negative real growth of house prices over such a period. It is possible that in the subsequent period characterized by house price appreciation—the one we consider—the distribution of German yields became more aligned with other European countries. Overall, it seems safe to argue that variation in rent-to-price ratios is not particularly pronounced in Germany. Put differently, despite its features contributing to a peculiarly low homeownership rate, the German housing market appears to be a useful setting to learn about general patterns in the pricing of residential properties.

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 3 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.[17] Focusing on the interquartile range and on the spread between the 95th and the 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 more geographically segmented.

In the analysis below, we seek to establish an upper bound for the role of geographical factors—after controlling for observable property characteristics—in the variation in rent-to-price ratios, to tell them apart from the component driven by idiosyncratic factors. To this end, we mostly focus on the logarithmic transformation of the ratio, using interchangeably the expressions "natural logarithm of the rent-to-price ratio" and "rent-to-price ratio".

3.1 Observable property characteristics

Heterogeneity across geographic 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 zip code-level, i.e., granular geographic units, especially in densely populated areas—is based on a parsimonious set of covariates, 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 the analysis of geographic factors, we thus assess the role of observable property-specific hedonic 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 category),

[17] 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.

with the addition of squared surface.[18] $\mathbf{x}_r$ 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.[19] The standard errors are clustered at the district level.

Table 4 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. However, pairwise correlations among surface, number of rooms, and number of bedrooms are above 70% (this does not extend to the number of bathrooms), which suggests caution in interpreting their regression coefficients. 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. Rental contracts requiring a higher deposit or higher heating expenses, and holiday properties for sale 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.[20]

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

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

[20] 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 roughly one fourth of the raw sample of sale listings—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

The correlation patterns described so far remain robust once we include all the remaining observable flat characteristics in column 4, with adjusted R^2 rising by 17% to 30% (from 13% in column 3). In Appendix Table A.2, we augment this specification with interaction terms between hedonic matching covariates and key national macroeconomic variables—such as the 10-year Bund yield-to-maturity, the growth of GDP per capita, and the growth of housing stock per capita—to allow for time-varying coefficients and thus better capture changes in expectations about the housing market. Even after including these terms, the adjusted R^2 remains at 30%.

All in all, 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 Local vs. idiosyncratic factors

The analyses conducted so far suggest that cross-sectional variation in rent-to-price ratios is substantial and tends to increase over our sample period. The importance of cross-sectional variation is corroborated by the fact that most of time-series variation in yields originates within highly granular geographic units, such as zip code areas. The idiosyncratic component of variation also appears to be crucial, with observable (to us) property-specific covariates together with time fixed effects explaining about one-third of the variation of property-level rent-to-price ratios.

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 heterogeneity in rent-to-price ratios may be explained by local differences in factors such as the age structure of the population, unemployment, income per capita, and the like. In Table 5, we thus seek to provide an upper bound to the fraction of rent-to-price ratio variability that can be explained by local geographic factors. In particular, we saturate the baseline specification in column 4 of Table 4—including all observable property-level covariates—with progressively finer geographical fixed effects. Relative to the baseline (adjusted R^2 of 30%), federal state and federal state-calendar quarter fixed effects can explain 5% and 6% more rent-price ratio variation (columns 1 and 2), respectively. In columns 3 and 4, we examine fixed effects at the district or district-quarter level. In this case, the fraction of explained variation rises to 41% and 44%, respectively.

correlation between matched and actual logarithmic ratios is 62.9%. Similarly, the graphical inspection of the two measures in Appendix Figure A.3 supports the validity of the matching procedure.

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 zip code and zip code-quarter fixed effects, reaching an adjusted R^2 of 47% and 59%, respectively.

Figure 4 sheds light on the distribution of residuals from these regressions, starting from a specification without any fixed effect up to one featuring zip code-quarter fixed effects. No matter the fixed effects scheme, residuals are symmetrically distributed if we consider the full regression sample. As we saturate the specification with finer fixed effects, however, the distribution becomes less dispersed and more peaked at the mean of 0. If we split the regression sample based on rent-to-price ratios quartiles, we observe that richer specifications produce more centered residuals. Especially for the top and bottom quartile yield subsamples, the residuals from the models without zip code-quarter fixed effects have substantially negative (positive) mean and median. Hence, the 41% unexplained variation of yields from column 6 of Table 5 stems from a remarkably symmetric distribution, in which outliers seem to play a limited role.[21]

Time-varying factors at the zip code level, together with observable property-level traits, are *potentially* able to explain a sizable fraction of heterogeneity in rent-to-price ratios. But variation that cannot be even potentially explained by those variables is still at a staggering 41%. The challenge is not only to identify the possible sources of the 41% of unexplained variation, but also to identify the *actual* local time-varying factors behind the 29% (= 59% − 30%) of variation captured by zip code-quarter fixed effects.

[21] In Appendix Figure A.4, we further explore the role of outliers in residuals. In Panel A we compare the density of actual rent-to-price ratios against that of fitted values (excluding estimated zip code-quarter fixed effects) from the model in column 6 of Table 5. The latter density is more peaked but, at the same time, exhibits the presence of notable outliers on the right tail. To understand the origin of these outliers, we re-estimate the model excluding property-specific covariates, namely $\mathbf{m}_{r,\bar{s}}$, $\mathbf{x}_r$, $\mathbf{x}_{\bar{s}}$, and $\mathbf{z}_{r,\bar{s}}$ in equation (6). Panels B, C, and D show the densities of estimated zip code-quarter fixed effects, fitted values (including estimated fixed effects), and residuals, respectively, for both the baseline specification and the one without flat covariates. It emerges that outliers in all these quantities originate from the inclusion of covariates. Next, we identify outliers as those observations for which either the estimated fixed effect or the residual are more than five standard deviations away from their mean. Appendix Figure A.5 confirms that the exclusion of such observations allows us to obtain fixed effects (Panel A) and residuals (Panel B) that are not plagued by outliers. Finally, Appendix Table A.3 repeats the analysis of Table 5 but excluding observations generating such extreme values. Reassuringly for the robustness of our baseline analysis, their exclusion has no appreciable effect on the adjusted R^2.

3.2.1 *Explaining local fixed effects*

The analysis above, by means of the inclusion of geographic area-quarter fixed effects, provides possible upper bounds to the fraction of variation in housing yields that can be explained by local time-varying factors. In this section, we make a step forward in trying to pin down actual local factors that matter for housing valuations. To this end, we use the historical record of social and economic indicators at the district level to study their impact on rent-to-price ratio regional heterogeneity. We estimate the following regression of rent-to-price ratios on regional 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 4. 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.[22]

To select the district-level covariates to be included in $\mathbf{p}_d$, we look at the relation between yields and single regional factors, discussing their possible role in shaping expectations about discount rates and rent growth in Appendix A. In particular, we group such factors in two categories: 1) demographic and economic fundamentals (Appendix Table A.4), and 2) local housing market characteristics (Appendix Table A.5). This analysis highlights 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. To reduce the dimensionality of the analysis, in $\mathbf{p}_d$ we include those factors that are statistically significant at the 1% level and exhibit low pairwise correlations among each other: 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.

Based on the adjusted R^2, district-level time-varying factors can potentially explain 14% (= 44% − 30%) of yield variation, as documented in Table 5. Column 1 of Table

[22] 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.

6 shows that directly controlling for the covariates in $\mathbf{p}_d$—rather than including district-quarter fixed effects—produces an adjusted R^2 of 36%, meaning that we can trace back to specific factors only around 42% (= 6%/14%) of the district-quarter fixed effect. All coefficient estimates display coefficient signs and magnitudes that are consistent with the discussion in Appendix A.[23]

In Appendix Figure A.6, we conduct a Blinder-Oaxaca decomposition of the rent-to-price ratio gap between districts belonging to the top quintile of each of the four selected regional dimensions in $\mathbf{p}_d$ 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. Irrespective of the conditioning variable, mean differences in rent-to-price ratios (whose unconditional average is 5.53%) are economically relevant, 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 characteristics between say cities with high disposable income per capita and other areas. By contrast, rent-to-price ratio sensitivities to some characteristics (like flat surface, number of rooms, property conditions, and quality of 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 capture the mere effect of unobservable property traits.

In Table 6, we then decompose the rent-to-price ratio, investigating separately the rent price per sqm (columns 2-5) and the sale price per sqm (columns 6-9). The specifications in columns 2-4 (6-8) illustrate that time-varying district-level factors can potentially explain up to 17% (23%) of the variation in rent (sale) prices, and that the covariates in $\mathbf{p}_d$ capture 65% (61%) of this incremental explanatory power. Moreover, both the rent and the sale price exhibit correlations with the four district-level factors of opposite sign relative to the rent-to-price ratio in column 1. In other words, the rent-to-price ratio co-moves more strongly with the sale price at the denominator than with the rent price at the numerator—which is consistent with the greater variability of the former. The

[23]We generally confirm these correlation patterns also in Appendix Table A.6, 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.

inclusion of zip code-quarter fixed effects in column 5 (9) raises the adjusted R^2 for the rent (sale) price to 82% (79%), as opposed to the 59% found for the rent-to-price ratio.

To sum up, although specifications featuring zip code-quarter fixed effects explain around two-thirds of variation in yields—and around four-fifths of variation in rent and sale prices—the part that we can re-conduct to specific property- or regional-level factors is substantially lower.[24] To capture the latter, we have only considered covariates at the district level, the most granular administrative subdivision at which a comprehensive set of statistics is publicly available (including, e.g., local output measures, industry structure, or new building completion). However, a narrower set of statistics exists also at municipality level and, more interestingly, for finely-defined 1-km raster cells (Breidenbach and Eilers, 2018). Below, we consider fixed effects for such cells, but we leave the study on the actual determinants of rent-to-price ratios at this level of geographic granularity to future research.

3.2.2 Exploring unexplained variation

Observable property-level characteristics and fine geographic fixed effects cannot span around two-fifths of variation in housing yields. This unexplained 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).[25] 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.). Apart from idiosyncratic factors, we thus evaluate several possible non-mutually exclusive stories behind unexplained cross-sectional variation in rent-to-price ratios, such as measurement errors, local amenities and agglomeration economies, and housing market frictions (illiquity, supply rigidities, regulation).

Measurement issues. Despite the richness of the vector of observable property-level traits we control for, we reckon measurement error due to unobserved flat characteristics, costs

[24]These upper bounds for adjusted R^2 appear not to be an artifact of the chosen specification. To address possible concerns about the use of the log-transformation (e.g., Cohn, Liu, and Wardlaw, 2022), in Appendix Table A.7 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 zip code-quarter fixed effects is even lower than for the log-linear specification (column 3).

[25]Giacoletti (2021) does not rely on property-level information on rents, thus abstracting from this source of variation in housing yields within zip codes.

borne by property owners, and the matching procedure could still underlie the puzzling degree of heterogeneity in housing yields that we document. In Panel A of Table 7 we seek to assess this potential source of "artificial" heterogeneity. In column 1, we start by considering rent-to-price ratios adjusted for expenses faced by tenants—namely by using rent inclusive of utilities at the numerator—in a specification otherwise identical to the most saturated one of Table 5. Not surprisingly, the adjusted R^2 decreases from 59% to 52% because the type and contractual features of included expenses are likely to introduce a further layer of heterogeneity in our valuation ratios, possibly exacerbating measurement problems. In other words, the choice to focus on rents exclusive of expenses for the computation of housing yields is warranted. Moreover, we do not observe the costs faced by owners. Although such costs are likely to exacerbate rather than mitigate idiosyncratic dispersion in yields by inducing higher volatility in rental income (Chambers et al., 2021), it is useful to tentatively quantify them. In Appendix Table A.8, we compute measures of net yields by approximating for maintenance costs, vacancy losses, and property taxes in a variety of ways. Net yields are on average roughly 30% lower than gross ones and exhibit a very similar fraction of unexplained variation in specifications including zip code-quarter fixed effects.[26]

Moving to matching errors, these could stem from several important unobservable (to us) property features such as the exact layout (e.g., number of windows south- vs. north-facing windows, etc.), distance from public transport network, view from the balcony, decoration aesthetics, cultural heritage, and the like, 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 2 (Panel A of Table 7) estimates the specification with zip code-quarter fixed effects, 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 to 71%, but the higher explanatory power appears to be linked to sample bias more than to the absence of matching errors. Indeed, in column 3 we use 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

[26] Appendix B provides a detailed discussion of the approach we follow to compute net yields.

possibly showing up more than once in a period. In column 4, 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 5, we exclude properties that, while still having a unique identifier in the RWI-GEO-RED database, exhibit the same characteristics in terms of pricing and facilities in a given zip code-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%.

Next, we impose additional restrictions in the matching algorithm to improve comparability within matched properties in terms of quality—besides controlling for covariate distances. In column 6, we consider only matches with an absolute surface difference smaller or equal than 5 sqm (the threshold in the baseline analysis is at 10 sqm). In column 7, the sample comprises only matched properties whose year of construction is at most 10 years apart from each other. In column 8, we require matched properties to have the same quality of facilities based on a categorical variable with four categories (from "simple" to "deluxe"). The more restrictive filter on the surface difference excludes 464,815 matched properties, leaving the adjusted R^2 unchanged at 59%. The adjusted R^2 increases to 63% and 66% when filtering on the year of construction and the quality of facilities, respectively. At the same time, in both cases the sample shrinks to around 300,000 observations. However, the notable decline in sample size is largely the result of the lack of information on either of the two variables for the flat for rent and/or the matched flat for sale, and does not necessarily reflect genuine matching errors of our algorithm. Therefore, the modest increase in explanatory power may just signal an exacerbation of sample bias.

In Panel B of Table 7, we implement two alternative matching schemes to better assess the role of matching errors related to (a) flat quality (columns 1 and 2) and (b) location (columns 3 and 4), getting in both cases rent-price ratios that map closely those from the baseline scheme, with correlations of 81% and 83%, respectively. In scheme (a), we extend the baseline algorithm by imposing exact matching also on flat conditions, thus obtaining a sample of 702,185 matched properties. In column 1, we re-estimate the specification in column 4 of Table 4, whereas in column 2 we saturate it with zip code-quarter fixed effects as in column 6 of Table 5. The adjusted R^2 goes from 23% to 56%, i.e., lower than in the baseline sample in both cases, suggesting that matching errors with respect to flat quality are not a prominent driver of the dispersion of housing yields.

Scheme (b), instead, improves the baseline one by imposing exact matching on highly granular geographic areas corresponding to all possible combinations of 1-km raster cells, five-digit zip codes, municipalities, and labor market regions. The resulting dataset features 610,120 matched properties: this approach minimizes spatial distance between each flat for rent and the corresponding flat for sale in terms of exposure to local amenities (e.g., restaurants) and/or agglomeration economies (due, e.g., to pooling of highly-skilled human capital). In spite of more accurate matching with regards to these dimensions, the adjusted R^2 in columns 3 and 4 stays at 30% and 61% without and with zip code-quarter fixed effects, respectively.

Overall, these findings underpin the idea that matching errors do not play a major role in generating unexplained variation in rent-to-price ratios under the baseline matching scheme. Put differently, their dispersion is unlikely to be an artifact of our methodology.

Local amenities and agglomeration economies. Two complementary explanations for high residual heterogeneity in yields relate to amenities and agglomeration economies operating below the zip code level. A growing body of evidence highlights how better access to amenities induces households to sort into certain neighborhoods based on their income or skills (e.g., Handbury and Weinstein, 2015; Ferreira and Wong, 2020; Couture and Handbury, 2020; Couture et al., 2021; Diamond and Moretti, 2021), potentially affecting property valuations. Rosenthal and Strange (2020) discuss how agglomeration economies—due, for instance, to knowledge spillovers and access to concentrated skilled workforce—are spatial in nature and can be present at different levels, even within neighborhoods, blocks or buildings. Leonardi and Moretti (2023) uncover agglomeration effects in the within-city location choices of amenities. Therefore, 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.

To explore this possibility, we continue to work with the matching scheme (b). In column 5 of Panel B of Table 7, we saturate the rent-to-price ratio regression with geographic area-quarter fixed effects. Relative to the specification with zip code-quarter, the adjusted R^2 increases by 10% reaching 71%. Time-varying factors at the level of these narrow geographic areas thus meaningfully impact dispersion in housing valuations. This is consistent with a relevant role of local amenities as well as of local agglomeration economies. We seek to disentangle the two stories by splitting the sample between global cities (columns 6 and 7) and other regions (columns 8 and 9), conjecturing that the former are more conducive to agglomeration effects. Across the two subsamples, the adjusted R^2 exhibits analogous patterns when moving from zip code-quarter to geographic area-quarter fixed effects. More intuitively, the explanatory power of these neighborhood-level factors appears to be more consistent with within-city disparities in access to amenities—a phenomenon likely at work also outside big cities—than with local agglomeration effects. Nonetheless, around 30% of variation in housing yields remains unexplained even after accounting for any possible difference in amenities available across our highly granular geographic areas. Research is needed to verify whether this residual variation is at least partially spatial in nature—i.e., at even more local level, like streets or blocks—or rather mostly reflects the property- or investor-specific factors we discuss below, as the study of Eichholtz et al. (2021) on Amsterdam and Paris suggests.

Housing market frictions. A variety of frictions in local housing markets could contribute to generate unexplained heterogeneity in housing valuations. A first class of frictions hinders the diffusion of information across markets. The asset pricing analogy between housing and stocks relies on the assumption that assimilation and availability of information 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. Market segmentation with limited search (Piazzesi, Schneider, and Stroebel, 2020) and limited liquidity of properties (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.

A second class of frictions pertains to rigidities in local housing supply. Such rigidities could stem, for instance, from local zoning laws or from the geographical conformation of the area. A highly inelastic supply together with a robust demand for housing could amplify market attention to small differences in the quality of properties and turn them into sizable wedges in valuation ratios.

Also local regulations can exacerbate heterogeneity in housing yields. Regulation can take a number of forms in the housing market, from the aforementioned zoning laws to taxation and rent controls. Inasmuch these regulations differentially impact properties within the same zip code, they may qualify as a driver of the unexplained variation of rent-to-price ratios.

In Panel C of Table 7, we provide suggestive evidence on the role of these housing market frictions by estimating the specification with zip code-quarter fixed effects on subsamples of the baseline matched dataset. We define such subsamples based on the median district in terms of (i) median online listing time (columns 1 and 2), (ii) housing stock per capita (columns 3 and 4), (iii) living space completed per capita (columns 5 and 6), and (iv) the property assessment rate B (column 7 and 8). Case (i) aims at capturing differences in informational frictions as proxied by the liquidity of the local housing market. Cases (ii), (iii), and (iv) gauge the role of supply rigidities and regulations in the form of zoning laws and property taxation. In each case, we do not uncover any economically significant difference in the explanatory power of our regression specification relative to the baseline in Table 5. This is not surprising because zip code-quarter fixed effects arguably absorb all cross-zip code variation in yields due to time-varying local housing liquidity, supply rigidities, or regulation, making it effectively a test of within-zip code differential effects of such frictions. Our results point to their limited quantitative importance.

In Appendix Table A.9, we then analyze the impact of a specific type of housing market regulation such as rent controls on the dispersion of rent-to-price ratios. To this end, we use a difference-in-differences design based on a reform of rent controls implemented in Germany during our sample period, finding no statistically significant impact on the heterogeneity of yields. We provide details on this exercise in Appendix C.

To sum up, we fail to find compelling evidence of a major role of measurement issues, agglomeration economies, informational frictions, supply rigidities, or regulation for zip code-level unexplained dispersion of rent-to-price ratios.[27] By contrast, within-zip code differences in access to amenities seem to account for a relevant share of dispersion. Nonetheless, even after controlling for those differences, about 30% of variation in housing yields cannot be captured.

Abstracting from rent growth expectations—arguably driven by the regional and local factors discussed above (demographics, economic growth, access to amenities, agglomera-

[27]Despite the lack of prima facie evidence, we acknowledge that these economic channels may still be at work within zip codes. More powerful testing strategies are needed to assess them in detail.

tion effects, etc.) rather than property-specific characteristics—, rent-to-price ratios may exhibit substantial variation within narrow geographic areas because of heterogeneous risk premia.[28] In turn, these risk premia may reflect properties' heterogeneous exposure to systematic risk factors but, due to widespread underdiversification of homeowners, also idiosyncratic risk, which in turn crucially depends on yield risk (Eichholtz et al., 2021; Eiling et al., 2020). The cross-sectional nature of our data does not allow us to credibly distinguish between idiosyncratic and systematic risk as determinants of such risk premia (and housing yields dispersion).

Moreover, idiosyncratic variation in yields links to the growing literature on heterogeneity across housing market investors (e.g., Fischer, Füss, Ruf, and Stehle, 2022; Gurun, Wu, Xiao, and Xiao, 2023; Lambie-Hanson, Li, and Slonkosky, 2019). In this sense, it seems natural to conjecture that unexplained local dispersion in yields in the presence of frictions may also reflect dispersion in investors' beliefs, both about discount rates and future rent growth. For instance, a particularly pessimistic seller may ask for a low price, inducing a high rent-to-price ratio. In other words, investor disagreement about future rent growth can generate idiosyncratic dispersion in yields.

Property-specific risk premia and expected rent growth rates—potentially originating from dispersion in investors' beliefs and preferences—are thus likely to underlie unexplained variation in yields. At the same time, spatial effects related to amenities and agglomeration economies may be at work even below the neighborhood level. Future work may seek to disentangle the role of all these potential drivers of granular yield variation.

3.2.3 The role of geographic aggregation

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-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 zip code-quarter fixed effects, and at different levels of aggregation: zip code area, district, and federal state.

[28] According to the static Gordon growth model, the rent-to-price ratio of flat f can be decomposed as $H/P_f = r^f + \mathbb{E}(r^e_f) - \mathbb{E}(g_{H,f})$, where r^f is the risk-free rate, $\mathbb{E}(r^e_f)$ is the expected risk premium, and $\mathbb{E}(g_{H,f})$ is the expected rent growth.

Table 8 contains information about the distribution of these cross-sectional standard deviations. Mean dispersion of raw ratios increases by 30% (= 2.177/1.667−1) going from zip code- to federal state-level data. After factoring out observables and zip code-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 persistence of cross-sectional variation in rent-to-price ratios at different levels of geographic aggregation.

3.2.4 Implications for household wealth

The high degree of heterogeneity—both explained and unexplained—in rent-to-price ratios maps into the distribution of household wealth around Germany. We attempt to put this relationship in perspective 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 80.88. The 10th and 90th percentiles of the rent-to-price ratio (non log-transformed) are 3.16% and 8.38%, respectively.[29] These ratios correspond to housing valuations of EUR 168,927 (= $(80.88 \cdot 66)/0.0316$) and EUR 63,700 (= $(80.88 \cdot 66)/0.0838$) for the median flat, implying a variation in the housing wealth of its owners of EUR 105,227 (= EUR $168,927 -$ EUR $63,700$).

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 5—where, besides controlling for property-level characteristics, we include district- and zip code-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 zip code 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

[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 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 3.32% and 8.32%.

EUR 90,556 and EUR 73,729 if we absorb time-varying district- or zip code-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 span a substantial fraction of dispersion in household wealth. The economic forces at play at geographically granular level may therefore translate into 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 transmit to local business cycles and, ultimately, economic development.

4 Housing returns

In this section, we rely on the pseudo-panel dataset described in Section 2.1.2 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 9, we provide some stylized facts on cohort-level returns, rent growth rates, and rent-to-price ratios within our pseudo-panel. Panel A shows that 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 0.36%. 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.[30] 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)

[30] These results are not far from those of Demers and Eisfeldt (2022), who recover city- and zip code-level returns for single-family rental properties in the US and find that price growth and rental yields constitute about 50% of total returns each. In our case, the split is 39% vs. 61%, respectively. Given that Demers and Eisfeldt (2022) use *net* rental yields while we use *gross* ones, our decomposition of total returns is probably even closer to theirs.

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.

Following Piazzesi et al. (2007), Panel B studies time-series variation in at different levels of geographic aggregation, namely at the cohort-, district-, state-, and national level. In line with Table 2, we find that national variation accounts only for a small fraction of cohort-level variation, especially in the case of returns and yields, whereas this pattern is less pronounced for rent growth. This is even more remarkable because, as pointed out above, the construction of the pseudo-panel per se absorbs a significant part of variation. It is also apparent that price growth rates are substantially more volatile than rent growth rates.

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 return from holding the DAX index, a proxy for the German stock market. The median housing asset delivers a cumulative return of about 90%, in line with the German stock market. Even housing assets in the bottom percentile generate a non-negligible return of about 25% over the sample period. 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.

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

In Table 11, we perform predictive regressions augmented with calendar quarter fixed effects, which absorb macroeconomic fluctuations in the spirit of Fama and French (2023). The inclusion of time fixed effects increases the R^2 across the board, in particular for returns, pointing to the relevance of aggregate shocks. Yet, the rent-to-price ratio retains

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

its predictive ability. Put differently, the part of yield variation driven by local factors significantly contributes to the predictability of returns and rent growth rates. This in turn lines up well with the decomposition in Panel B of Table 9, where we highlight how time-series variation in yields largely originates at a granular level.

Then, Appendix Table A.10 shows that the relation between excess returns and rent growth, and the rent-to-price ratio remains qualitatively unchanged if we directly 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, (local) 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 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. After accounting for a wide set of property-specific characteristics, 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. A lot of the variation of rent-to-price ratios remains unexplained even if we absorb all time-varying zip code-level factors by means of fixed effects or if we account for measurement errors in a variety of ways. This residual heterogeneity appears to relate to disparities in amenities across neighborhoods as well as, arguably, to property-specific dispersion in risk premia and rent growth expectations, possibly driven by heterogeneity in investors' beliefs and preferences.

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 (local) fluctuations in 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 heterogeneity in rent-to-price ratios, with the interdecile range of their unexplained component accounting for more than one-third of the mean German household's wealth. 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 rigorously pin down the origin(s) of such yet unexplained variation, be it related to local amenities and/or agglomeration effects, heterogeneous exposures to systematic risk factors, idiosyncratic risk, investors' disagreement about future rent growth, 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. Heblich, and T. Seidel. 2022. Micro-geographic property price and rent indices. *Regional Science and Urban Economics* p. 103836.

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.

Amaral, F., M. Dohmen, S. Kohl, and M. Schularick. 2022. Superstar returns. Working paper.

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

Bach, S., and S. Eichfelder. 2021. Real estate taxation reform: Tax land values, abolish privileges. *DIW Weekly Report* 11:199–205.

Barrios, S., C. Denis, V. Ivaškaitė-Tamošiūnė, A. Reut, and E. Vázquez Torres. 2019. Housing taxation: A new database for Europe. Working paper.

Bednarek, P., D. M. t. Kaat, C. Ma, and A. Rebucci. 2021. Capital flows, real estate, and local cycles: Evidence from German cities, banks, and firms. *Review of Financial Studies* 34:5077–5134.

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 L. Eilers. 2018. RWI-GEO-GRID: Socio-economic data on grid level. *Jahrbücher für Nationalökonomie und Statistik* 238:609–616.

Breidenbach, P., L. Eilers, and J. Fries. 2022. Temporal dynamics of rent regulations–The case of the German rent control. *Regional Science and Urban Economics* 92:103737.

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

Callaway, B., and P. H. Sant'Anna. 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics* 225:200–230.

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

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* 34:3572–3607.

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

Clark, S., and N. Lomax. 2020. Rent/price ratio for English housing sub-markets using matched sales and rental data. *Area* 52:136–147.

Cohn, J. B., Z. Liu, and M. I. Wardlaw. 2022. Count (and count-like) data in finance. *Journal of Financial Economics* 146:529–551.

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.

Couture, V., C. Gaubert, J. Handbury, and E. Hurst. 2021. Income growth and the distributional effects of urban spatial sorting. Working paper.

Couture, V., and J. Handbury. 2020. Urban revival in America. *Journal of Urban Economics* 119:103267.

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

Demers, A., and A. L. Eisfeldt. 2022. Total returns to single-family rentals. *Real Estate Economics* 50:7–32.

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

Diamond, R., and E. Moretti. 2021. Where is standard of living the highest? Local prices and the geography of consumption. Working paper.

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.

Dustmann, C., B. Fitzenberger, and M. Zimmermann. 2022. Housing expenditure and income inequality. *Economic Journal* 132:1709–1736.

Eichholtz, P., M. Korevaar, and T. Lindenthal. 2023. Limited growth: 500 years of urban rents. Working paper.

Eichholtz, P., M. Korevaar, T. Lindenthal, and R. Tallec. 2021. The total return and risk to residential real estate. *Review of Financial Studies* 34:3608–3646.

Eiling, E., E. Giambona, R. Lopez Aliouchkin, and P. Tuijp. 2020. Homeowners' risk premia: Evidence from zip code housing returns. Working paper.

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.

Fama, E. F., and K. R. French. 2023. House prices and rents. Working paper.

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.

Feng, X., N. Jaimovich, K. Rao, S. J. Terry, and N. Vincent. 2023. Location, location, location: Manufacturing and house price growth. *Economic Journal* 133:2055–2067.

Ferreira, F. V., and M. Wong. 2020. Estimating preferences for neighborhood amenities under imperfect information. Working paper.

Fischer, M., R. Füss, D. Ruf, and S. Stehle. 2022. Betting on homes. Working paper.

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.

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

Guren, A. M., A. McKay, E. Nakamura, and J. Steinsson. 2021. Housing wealth effects:

The long view. *Review of Economic Studies* 88:669–707.

Gurun, U. G., J. Wu, S. C. Xiao, and S. W. Xiao. 2023. Do Wall Street landlords undermine renters' welfare? *Review of Financial Studies* 36:70–121.

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

Hahn, A. M., K. A. Kholodilin, S. R. Waltl, and M. Fongoni. 2023. Forward to the past: Short-term effects of the rent freeze in Berlin. *Management Science* (Forthcoming).

Halket, J., L. Loewenstein, and P. S. Willen. 2023. The cross-section of housing returns. Working paper.

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

Handbury, J., and D. E. Weinstein. 2015. Goods prices and availability in cities. *Review of Economic Studies* 82:258–296.

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.

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. 2022. Consumption commitments and housing dynamics. Working paper.

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

Kholodilin, K. A. 2017. Quantifying a century of state intervention in rental housing in Germany. *Urban Research & Practice* 10:267–328.

Kholodilin, K. A., A. Mense, and C. Michelsen. 2019. Rent control, market segmentation, and misallocation: Causal evidence from a large-scale policy intervention. Working paper.

Kohl, S. 2016. Urban history matters: Explaining the German–American homeownership gap. *Housing Studies* 31:694–713.

Kohl, S., and J. Sørvoll. 2021. Varieties of social democracy and cooperativism: Explaining the historical divergence between housing regimes in Nordic and German-speaking countries. *Social Science History* 45:561–587.

Kosfeld, R., and A. Werner. 2012. Deutsche Arbeitsmarktregionen–Neuabgrenzung nach den Kreisgebietsreformen 2007–2011. *Raumforschung und Raumordnung—Spatial Research and Planning* 70:49–64.

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

Lambie-Hanson, L., W. Li, and M. Slonkosky. 2019. Institutional investors and the US housing recovery. Working paper.

Leonardi, M., and E. Moretti. 2023. The agglomeration of urban amenities: Evidence from Milan restaurants. *American Economic Review: Insights* 5:141–157.

Mense, A., C. Michelsen, and K. A. Kholodilin. 2019. The effects of second-generation rent control on land values. *AEA Papers and Proceedings* 109:385–88.

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., and M. Schneider. 2016. Housing and macroeconomics. *Handbook of Macroeconomics* 2:1547–1640.

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

Piazzesi, M., M. Schneider, and S. Tuzel. 2007. Housing, consumption and asset pricing. *Journal of Financial Economics* 83:531–569.

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.

Reher, M. 2021. Finance and the supply of housing quality. *Journal of Financial Economics* 142:357–376.

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.

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.

Sant'Anna, P. H., and J. Zhao. 2020. Doubly robust difference-in-differences estimators. *Journal of Econometrics* 219:101–122.

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

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.

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.

Voigtländer, M. 2009. Why is the German homeownership rate so low? *Housing Studies* 24:355–372.

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

Waltl, S. R. 2018. Estimating quantile-specific rental yields for residential housing in Sydney. *Regional Science and Urban Economics* 68:204–225.

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.

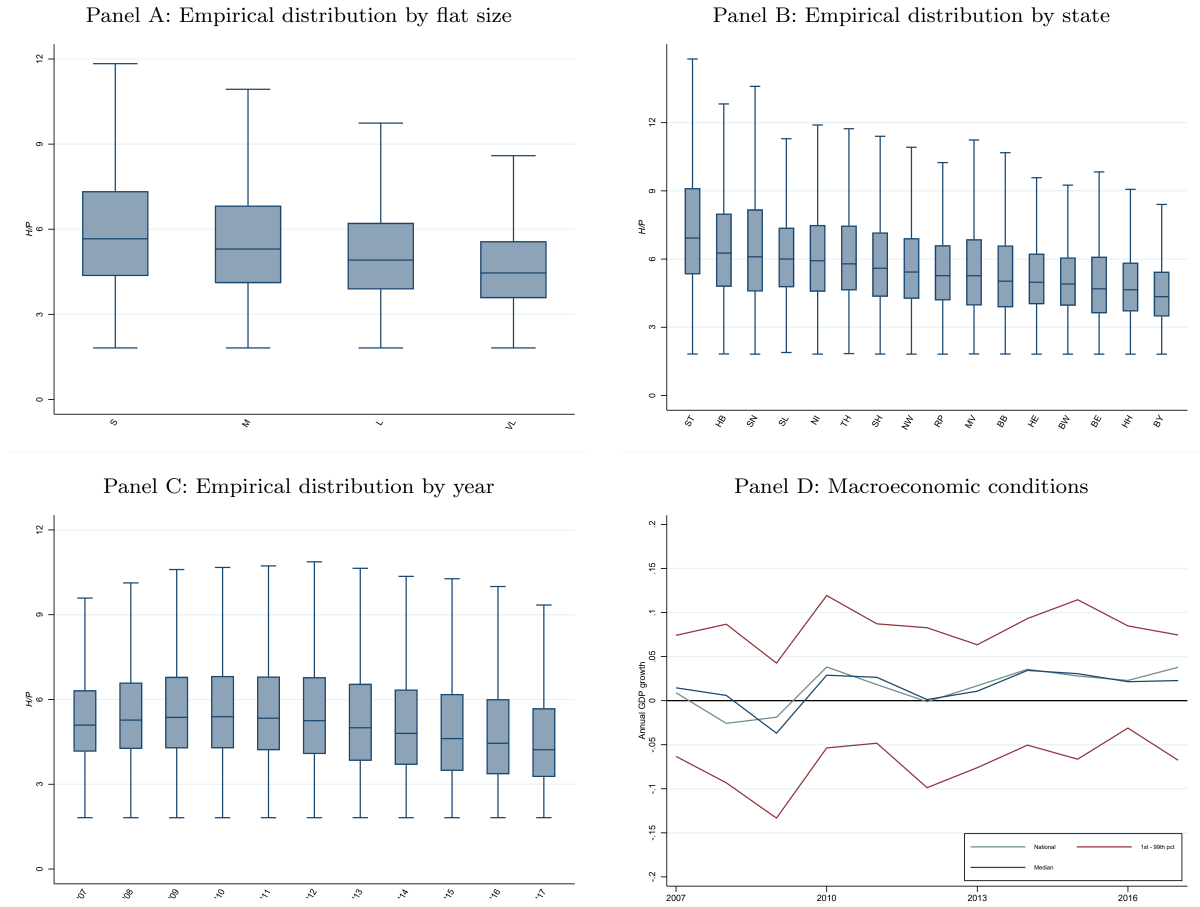


Figure 1: Empirical distribution of the rent-to-price ratio vs. macroeconomic conditions
This figure shows the conditional empirical distribution of the rent-to-price ratio (obtained via the baseline matching) through box plots (Panel A to C) against the evolution of macroeconomic conditions in Germany (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.

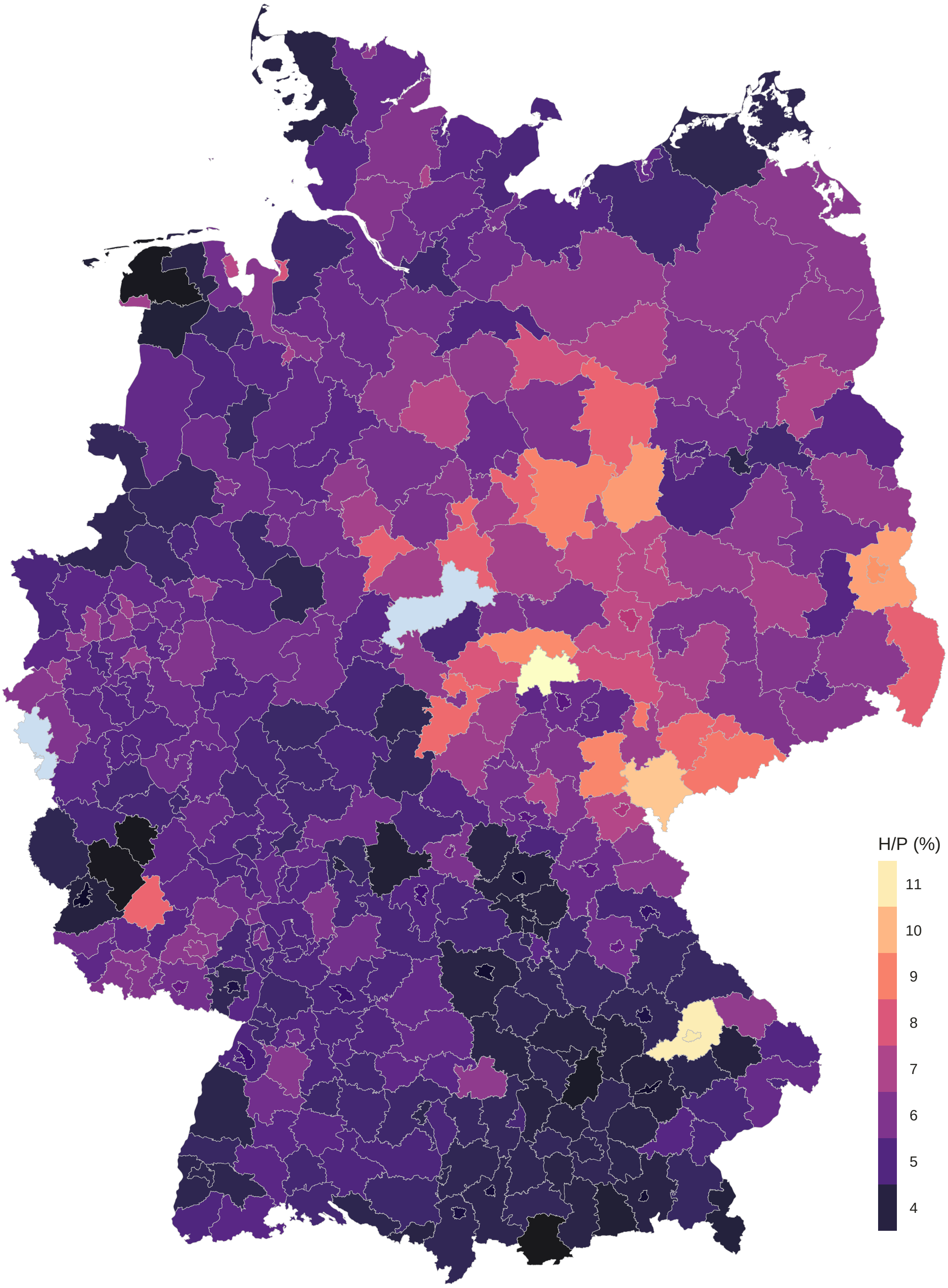


Figure 2: Median rent-to-price ratio at the district level
This figure visualizes the median rent-to-price ratio (obtained via the baseline matching) at the district level across Germany, pooling all periods between 2007 and 2017. Grey-colored districts have no observations.

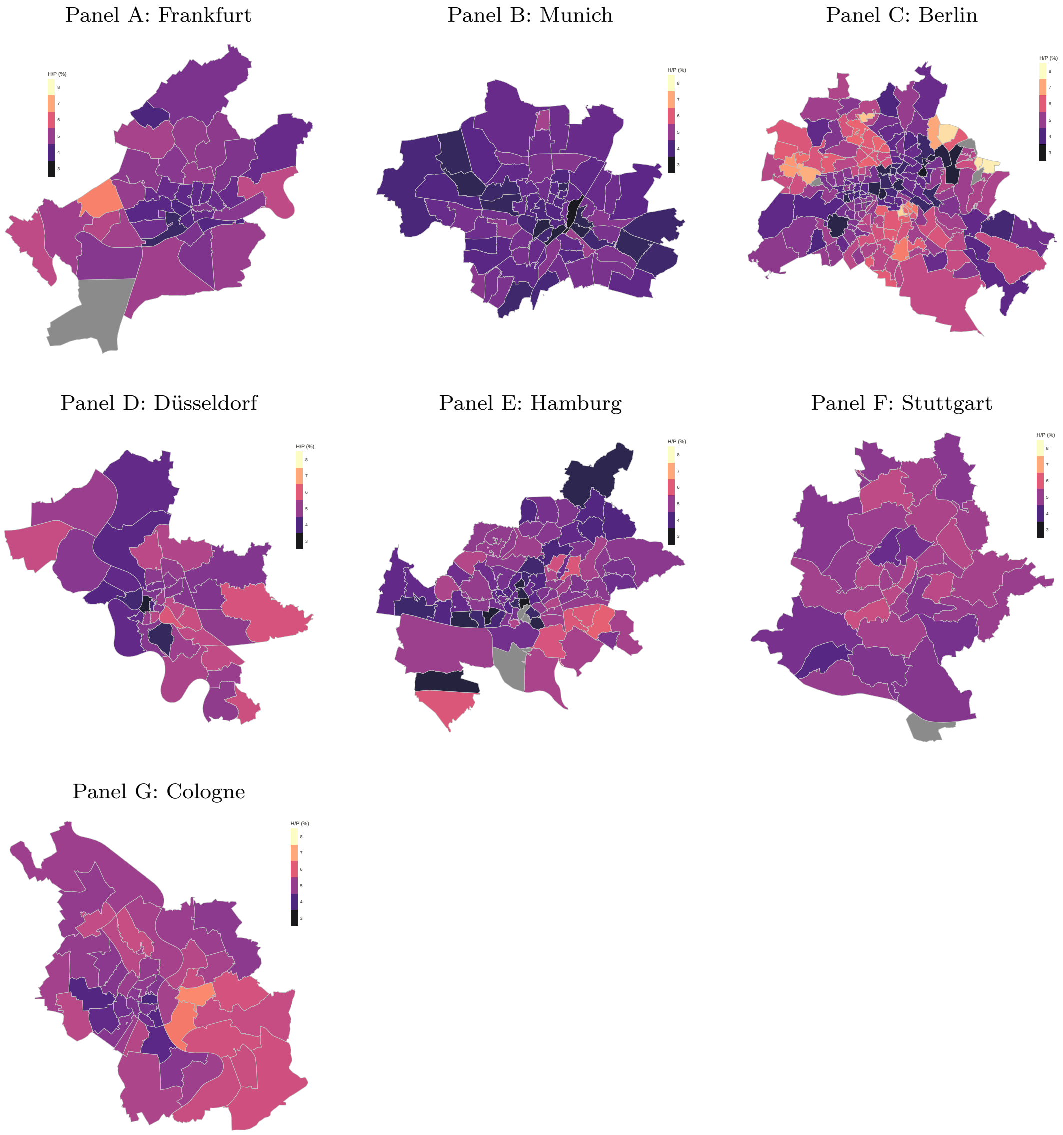


Figure 3: Median rent-to-price ratio at the zip code level within global cities
This figure visualizes the median rent-to-price ratio (obtained via the baseline matching) at the five-digit zip code level across German global cities, pooling all periods between 2007 and 2017. Grey-colored zip code areas 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.

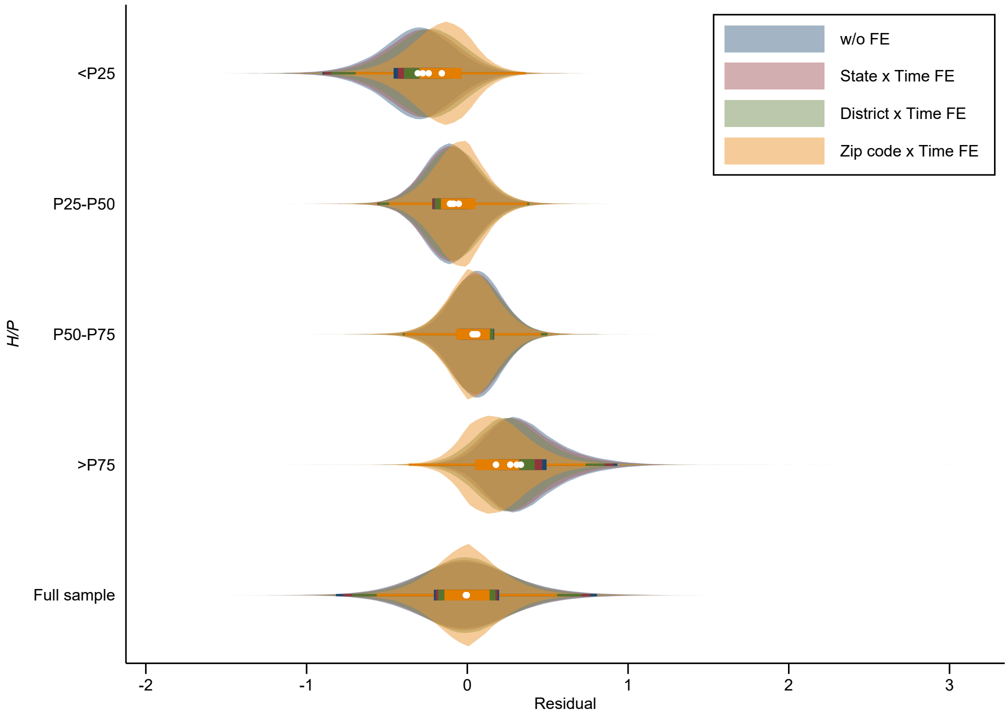


Figure 4: Distribution of residuals from rent-to-price ratio regressions
This figure visualizes the distribution of residuals from rent-to-price ratio regressions under different fixed effects schemes by means of violin plots. The violin plot of each variable features a superimposed box plot. The residuals are first obtained from estimating equation (6) without calendar quarter fixed effects. The other specifications augment equation (6) with (i) state-quarter fixed effects as in column 2 of Table 5, (ii) district-quarter fixed effects as in column 4 of Table 5, and (iii) zip code-quarter fixed effects as in column 6 of Table 5. The violin plots are reported for the entire distribution of the rent-to-price ratio as well as conditioning on the quartile of the latter.

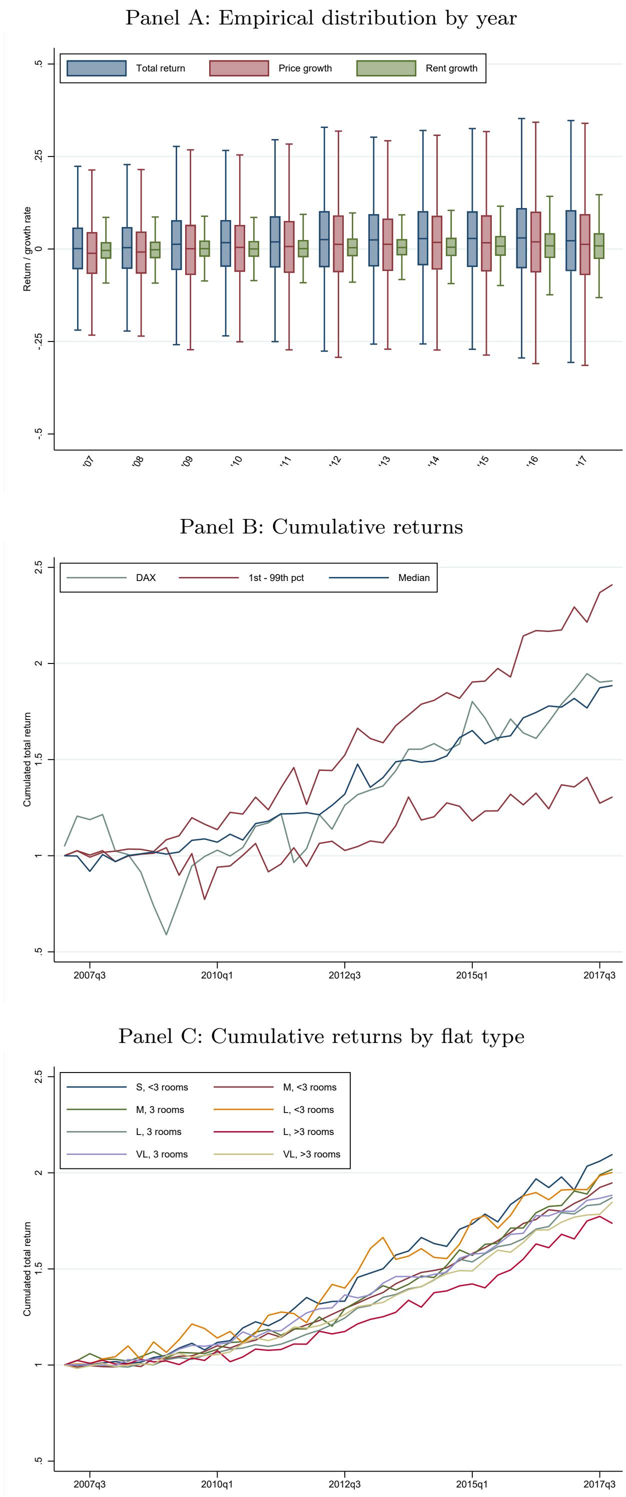


Figure 5: Housing returns
This figure shows the distribution of housing returns obtained via the pseudo-panel approach for the period 2007-2017. Panel A reports the empirical distribution of quarterly total returns (r), price growth rates (r^*), and rent growth rates (Δh) by means of year-by-year box 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 DAX stock index. 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.

Table 1: Summary statistics on flats and the local economy

This table shows summary statistics on property characteristics distinguishing between flats for rent and matched 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)	1,613,889	5,822.831	2,417.087					
Annual inclusive rent (EUR)	1,542,229	7,328.041	2,752.795					
Expenses (EUR)	1,539,334	123.834	48.294					
Heating expenses (EUR)	350,499	63.530	27.084					
Heating included	1,422,135	0.693	0.461					
Deposit (EUR)	1,447,416	1,244.801	636.028					
Sale listing								
Sale price (EUR)				1,613,889	124,811.464	76,866.112		
Housing benefits (EUR)				1,041,156	184.236	62.559		
Rented				1,013,531	0.352	0.474		
Holiday property				1,613,889	-2.904	4.170		
Matching covariates								
Surface (sqm)	1,613,889	68.122	18.225	1,613,889	68.129	18.043	0.007	0.370
No. rooms	1,613,889	2.444	0.671	1,613,889	2.444	0.671	0.000	0.000
No. bedrooms	1,613,889	1.443	0.577	1,613,889	1.443	0.577	0.000	0.000
No. bathrooms	1,613,889	1.021	0.143	1,613,889	1.021	0.143	0.000	0.000
Floor no.	1,613,889	1.939	1.159	1,613,889	1.963	1.210	0.023	17.673

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	1,301,809	3.379	1.356	1,315,864	3.719	1.886	0.340	167.522
Usable surface (sqm)	527,012	36.579	30.093	598,913	24.656	26.504	-11.924	-223.547
Construction year	999,921	1,967.616	39.058	1,425,880	1,971.613	35.989	3.998	82.196
Year last modernization	425,567	2,009.449	5.464	462,030	2,006.736	6.614	-2.713	-209.646
Property conditions	1,359,308	2.671	0.687	1,399,923	2.612	0.787	-0.060	-67.000
Quality of facilities	876,449	2.464	0.592	895,645	2.472	0.603	0.008	8.562
Protected building	87,259	0.008	0.089	1,117,558	0.051	0.218	0.043	57.242
Energy consumption (kWh/(sqm $\times$ y))	355,908	121.907	49.992	493,104	121.681	61.092	-0.226	-1.814
Energy rating	26,376	4.408	1.983	33,204	3.944	1.911	-0.464	-28.948
Hot water in energy exp.	869,821	0.151	0.358	881,597	0.228	0.416	0.077	131.277
Balcony	1,408,992	0.793	0.405	1,490,899	0.872	0.331	0.079	182.345
Available parking	158,962	0.934	0.248	198,145	0.957	0.201	0.023	30.183
Accessible with wheelchair	15,784	0.787	0.410	24,122	0.914	0.281	0.127	36.735
Guest WC	1,107,219	0.178	0.383	1,169,994	0.210	0.404	0.032	60.997
Garden	1,010,590	0.295	0.456	1,007,113	0.336	0.469	0.041	63.197
Cellar	1,273,428	0.805	0.396	1,319,378	0.831	0.370	0.026	54.150
Kitchen	1,200,599	0.561	0.496	1,207,938	0.565	0.492	0.004	6.541
Elevator	1,086,959	0.309	0.462	1,143,600	0.445	0.494	0.136	212.363
Assisted living	516,664	0.055	0.229	490,551	0.124	0.328	0.069	122.810
Immediate availability	1,518,491	0.317	0.465	1,133,943	0.346	0.469	0.029	50.189
No. of days online	1,605,830	26.764	33.474	1,613,889	63.379	91.683	36.616	475.534
No. of hits	1,605,910	903.414	1,020.385	1,613,889	871.727	1,476.745	-31.687	-22.389
No. of clicks (contact button)	1,606,036	10.947	22.277	1,613,889	6.086	13.674	-4.861	-236.105
No. of clicks (customer profile)	1,606,154	1.286	4.287	1,613,889	0.957	4.275	-0.329	-69.028
No. of clicks (share button)	1,606,865	0.858	1.539	1,613,889	0.724	1.611	-0.134	-76.071
No. of clicks (customer URL)	1,606,246	3.246	6.030	1,613,889	3.879	17.190	0.632	43.996

Panel B: Local conditions

	Obs.	Mean	SD	Min	Median	Max
Demographic and economic conditions						
Old-to-working age ratio (%)	3,661	35.009	4.520	22.704	34.716	52.192
Population (thousand)	3,661	22,005.841	24,991.681	3,394.400	16,500.000	361,349.500
Disposable income per capita (EUR, thousand)	3,630	18.692	2.207	14.050	18.649	28.394
GDP per capita (EUR, thousand)	3,630	30.037	11.853	14.134	26.925	88.045
Unemployment rate (%)	3,661	6.672	3.239	1.200	6.000	21.200
Manufacturing industry share (%)	3,649	7.364	2.418	2.518	7.133	21.360
No. businesses	3,661	10,416.132	13,410.078	1,537.000	7,577.000	189,177.000
Housing market conditions						
IQR of surface (sqm)	3,661	31.195	7.416	2.590	30.750	92.000
Median online listing time	3,661	22.795	13.115	1.000	20.000	160.000
Living space completed per capita (sqm)	3,636	0.266	0.138	0.045	0.246	0.747
No. posted flats	3,661	424.677	1,323.048	4.000	153.000	29,504.000
Housing stock per capita (sqm)	3,632	44.534	4.059	35.457	44.453	57.357
Land price (sqm)	3,604	140.706	117.261	6.961	105.187	831.342
Property assessment rate B (basis points)	3,661	393.470	82.563	244.000	374.000	855.000

Table 2: Time-series variation of rent-to-price ratios at different levels of geographic aggregation
This table shows summary statistics on time-series variation of rent-to-price ratios—obtained via the baseline matching procedure of flats for rent to flats for sale—at zip code-, district-, and state- and national level. To this end, for each geographic area a time series of the rent-to-price ratio is obtained by computing its average quarter-by-quarter. Then, the standard deviation and the interquartile range of each of these series is computed. Finally, based on these estimates, the average standard deviation and interquartile range across geographic areas are obtained and reported below. These averages are computed either giving equal weight to the geographic areas or by weighting them by the number of listed flats. Geographic area-quarters with fewer than 10 observations are removed from the sample. Refer to Appendix Table A.1 for variable definitions.

Level of aggregation:	Zip code	District	State	National
Equally weighted				
SD(H/P)	1.050	0.887	0.571	0.356
P75-P25(H/P)	1.215	1.118	0.872	0.561
Weighted by no. properties				
SD(H/P)	0.987	0.631	0.442	0.356
P75-P25(H/P)	1.283	0.968	0.747	0.561

Table 3: 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 the baseline matching procedure of flats for rent to flats for sale) 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	2.766	3.969	5.094	6.557	9.839	2.587	7.073
Actual H/P	3.407	4.605	5.662	7.211	11.294	2.606	7.887
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 4: 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 the baseline 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	-0.021***	-0.020***	-0.020***	-0.020***
	(-26.14)	(-17.97)	(-16.86)	(-20.37)
Surface squared	0.000***	0.000***	0.000***	0.000***
	(31.51)	(29.10)	(28.07)	(33.65)
No. rooms	0.150***	0.136***	0.134***	0.128***
	(11.51)	(13.45)	(13.82)	(17.10)
No. bedrooms	-0.036***	-0.030***	-0.028***	-0.020***
	(-3.72)	(-4.25)	(-4.13)	(-3.59)
No. bathrooms	-0.069***	-0.060***	-0.054***	-0.040***
	(-5.63)	(-5.33)	(-6.15)	(-4.95)
Floor no.	0.005	0.007	0.006	0.006*
	(0.97)	(1.53)	(1.39)	(1.69)
Expenses		0.000	-0.000	-0.000***
		(0.38)	(-0.14)	(-2.68)
Heating expenses		-0.000***	-0.000***	-0.000***
		(-3.98)	(-4.16)	(-3.00)
Heating included		0.020	0.021	0.005
		(1.52)	(1.62)	(0.54)
Deposit		-0.000**	-0.000***	-0.000***
		(-2.52)	(-2.61)	(-4.05)
Housing benefits			0.000***	0.000***
			(3.77)	(2.66)
Holiday property			-0.002***	-0.003***
			(-4.90)	(-8.42)
Rented			0.114***	0.075***
			(7.59)	(6.52)
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)	1.64	1.64	1.64	1.64
SD(y)	0.38	0.38	0.38	0.38
Observations	1,613,889	1,613,889	1,613,889	1,613,889
Adjusted R^2	0.09	0.11	0.13	0.30

Table 5: Local 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 the baseline matching procedure of flats for rent to flats for sale. Each column augments the specification of column 4 of Table 4 with progressively finer fixed effects (indicated below). Standard errors are clustered by district. 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
Zip code FE	No	No	No	No	Yes	No
Zip code×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)	1.64	1.64	1.64	1.64	1.64	1.64
SD(y)	0.38	0.38	0.38	0.38	0.38	0.38
Observations	1,613,889	1,613,889	1,613,887	1,612,893	1,613,243	1,595,086
Adjusted R^2	0.35	0.36	0.41	0.44	0.47	0.59
Adjusted within R^2	0.28	0.28	0.29	0.30	0.31	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 and fixed effects for a German sample between 2007 and 2017. Column 1 uses as dependent variable the log-transformed rent-to-price ratio is obtained via the baseline matching procedure of flats for rent to flats for sale. Columns 2-5 (6-9) use 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)	(4)	(5)	(6)	(7)	(8)	(9)
Old-to-working age ratio (%)	0.016***		-0.018***				-0.034***		
	(7.54)		(-9.62)				(-10.44)		
Disposable income per capita (EUR, thousand)	-0.018***		0.029***				0.047***		
	(-5.84)		(12.54)				(10.55)		
Living space completed per capita (m^2)	-0.295***		0.198***				0.493***		
	(-4.87)		(3.56)				(5.08)		
ln(No. businesses)	-0.034***		0.021***				0.054***		
	(-3.31)		(3.94)				(4.82)		
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	No	No	Yes	Yes	No	No
District×Time FE	No	No	No	Yes	No	No	No	Yes	No
Zip code×Time FE	No	No	No	No	Yes	No	No	No	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.64	4.41	4.40	4.41	4.41	7.38	7.37	7.38	7.38
SD(y)	0.38	0.30	0.30	0.30	0.30	0.51	0.51	0.51	0.51
Observations	1,588,823	1,613,889	1,588,823	1,612,893	1,595,086	1,613,889	1,588,823	1,612,893	1,595,086
Adjusted R^2	0.36	0.58	0.69	0.75	0.82	0.42	0.56	0.65	0.79
Adjusted within R^2	0.35	0.57	0.69	0.31	0.28	0.40	0.55	0.22	0.21

Table 7: Possible drivers of unexplained 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 role of potential measurement errors in rent-to-price ratios and local housing market characteristics. Each column builds on the specification of column 4 of Table 4 augmented with fixed effect structures (indicated below). Panel A focuses on measurement errors in rent-to-price ratios. In column 1, the dependent variable is the log-transformed inclusive rent-to-price ratio (i.e., adjusted for utilities), as obtained via the baseline matching procedure of flats for rent to flats for sale. In column 2, 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 3 to 8, the dependent variable is the log-transformed rent-to-price ratio, as obtained via the baseline matching procedure of flats for rent to flats for sale. Column 3 restricts the sample to properties for which also the actual rent-to-price ratio is available. Column 4 augments the specification with fixed effects for each percentile of a matching quality measure. Column 5 removes from the sample potential duplicate listings. In columns 6 to 8, additional restrictions in the matching procedure are introduced. The sample in column 6 discards any match for which the absolute difference in terms of surface between the flat for sale and the matched flat for sale is larger than 5 sqm. The sample in column 7 discards any match for which the absolute difference in terms of construction year between the flat for sale and the matched flat for sale is larger 10 years or for which the construction year is missing for any of the flats. The sample in column 8 discards any match for which the flat for sale and the matched flat for sale have different reported quality of facilities or for which this variable is missing for any of the flats. Both in Panel B and in Panel C, the dependent variable is the log-transformed rent-to-price ratio, as obtained via the baseline matching procedure of flats for rent to flats for sale. Panel B considers two alternative matching schemes. The alternative matching scheme (a) in columns 1 and 2 augments the baseline matching procedure by imposing exact matching also on flat conditions. The alternative matching scheme (b) in columns 3 and 9 imposes exact matching at a finer geographic level than the five-digit zip code, namely on the areas resulting from combinations of labor market regions, municipalities, five-digit zip codes, and 1-km raster cells. Columns 6 and 7 restrict the sample to global cities, whereas columns 8 and 9 focus on the remaining regions. Panel C estimates the specification in column 6 of Table 5 over different subsamples defined with respect to local housing market characteristics. The considered housing market characteristics (indicated below) comprise the median online listing time (i), housing stock per capita (ii), living space completed per capita (iii), and the property assessment rate B (iv). Odd (even) columns restrict the sample to districts that are above (below) the median of the considered housing market characteristic. Standard errors are clustered by district. Refer to Appendix Table A.1 for variable definitions.

Panel A: Measurement errors

	ln(Inclusive H/P)	ln(Actual H/P)	ln(H/P)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Zip code×Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matching quality FE	No	No	No	Yes	No	No	No	No
Likely duplicates	Included	Included	Included	Included	Excluded	Included	Included	Included
Additional restriction	–	–	–	–	–	Surface	Constr. year	Quality of fac.
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.88	1.77	1.74	1.64	1.63	1.64	1.64	1.62
SD(y)	0.39	0.39	0.37	0.38	0.38	0.37	0.33	0.36
Observations	1,520,503	330,261	330,261	1,595,086	1,347,824	1,131,271	344,230	280,261
Adjusted R^2	0.52	0.71	0.70	0.59	0.58	0.59	0.63	0.66
Adjusted within R^2	0.12	0.08	0.26	0.31	0.32	0.30	0.25	0.27

Panel B: Alternative matching schemes

	ln(H/P)								
	(1)	(2)	(3)	(4)	(5)	(6) Global cities	(7) Global cities	(8) Other	(9) Other
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	No	Yes	No	No	No	No	No	No
Zip code×Time FE	No	Yes	No	Yes	Yes	Yes	No	Yes	No
Area×Time FE	No	No	No	No	Yes	No	Yes	No	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Alt. matching scheme	(a)	(a)	(b)	(b)	(b)	(b)	(b)	(b)	(b)
Mean(y)	1.64	1.64	1.64	1.64	1.64	1.56	1.56	1.70	1.70
SD(y)	0.34	0.34	0.37	0.37	0.37	0.35	0.35	0.37	0.37
Observations	702,185	681,070	610,120	587,643	546,004	226,571	218,177	361,058	327,827
Adjusted R^2	0.23	0.56	0.30	0.61	0.71	0.61	0.69	0.60	0.72
Adjusted within R^2	0.21	0.24	0.28	0.31	0.31	0.36	0.36	0.29	0.28

Panel C: Local housing market characteristics

	ln(H/P)							
	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low	(7) High	(8) Low
Flat covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip code×Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat	Match. flat
Conditioning variable	(i)	(i)	(ii)	(ii)	(iii)	(iii)	(iv)	(iv)
Mean(y)	1.72	1.60	1.61	1.64	1.56	1.69	1.64	1.61
SD(y)	0.37	0.38	0.38	0.38	0.35	0.39	0.39	0.34
Observations	424,479	1,170,368	248,212	1,336,225	611,727	968,413	1,254,126	339,776
Adjusted R^2	0.54	0.60	0.62	0.59	0.58	0.58	0.59	0.58
Adjusted within R^2	0.27	0.33	0.35	0.31	0.33	0.31	0.31	0.34

Table 8: 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 zip code, district, and state-quarter level. The rent-to-price ratio is obtained via the baseline 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 5. Suspect duplicate property listings as well as geographic area-quarters with fewer than 30 observations are removed from the sample. Refer to Appendix Table A.1 for variable definitions.

	Obs.	Mean	SD	Min	P25	Median	P75	Max
Zip code-level aggregation								
SD(H/P)	12,530	1.667	0.672	0.270	1.187	1.534	1.999	5.795
SD(Filtered H/P)	12,530	1.418	0.590	0.379	1.002	1.277	1.673	5.350
District-level aggregation								
SD(H/P)	6,063	1.929	0.622	0.404	1.488	1.856	2.286	5.140
SD(Filtered H/P)	6,063	1.392	0.468	0.445	1.053	1.327	1.634	4.810
State-level aggregation								
SD(H/P)	664	2.177	0.496	0.874	1.817	2.130	2.478	3.999
SD(Filtered H/P)	664	1.410	0.342	0.613	1.189	1.366	1.615	3.131

Table 9: 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. Panel A provides basic statistics for these quantities. Panel B focuses on their time-series variation at cohort-, district-, and state- and national level. The average standard deviation and interquartile range across geographic areas are reported. These averages are computed either giving equal weight to the geographic areas or by weighting them by the number of listed flats. Refer to Appendix Table A.1 for variable definitions.

Panel A: Basic statistics

	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

Panel B: Time-series variation

Level of aggregation:	Cohort	District	State	National
Equally weighted				
SD (r)	12.293	10.068	5.890	1.283
SD (r^*)	12.370	10.136	5.927	1.324
SD (r^e)	4.258	3.271	1.673	0.561
SD (Δh)	12.318	10.099	5.972	1.469
SD (H^q/P)	0.175	0.158	0.128	0.087
P75-P25 (r)	15.938	12.966	7.806	1.756
P75-P25 (r^*)	16.066	13.064	7.892	1.849
P75-P25 (r^e)	5.470	4.189	2.276	0.800
P75-P25 (Δh)	15.978	13.050	7.976	2.490
P75-P25 (H^q/P)	0.245	0.223	0.192	0.155
Weighted by no. properties				
SD (r)	9.211	5.447	2.432	1.283
SD (r^*)	9.261	5.482	2.462	1.324
SD (Δh)	9.243	5.495	2.554	1.469
SD (r^e)	3.300	1.913	0.924	0.561
SD (H^q/P)	0.156	0.126	0.095	0.087
P75-P25 (r)	12.045	7.258	3.225	1.756
P75-P25 (r^*)	12.139	7.329	3.310	1.849
P75-P25 (r^e)	4.210	2.567	1.366	0.800
P75-P25 (Δh)	12.134	7.426	3.679	2.490
P75-P25 (H^q/P)	0.235	0.207	0.167	0.155

Table 10: Predictive regressions

This table reports estimates from predictive regressions for the 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 $k = 1,\ 4,\ 12$. 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 11: Predictive regressions including time fixed effects
This table reports estimates from predictive regressions for the housing premium and rent growth on the log-transformed rent-to-price ratio, including time fixed effects. 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 $k = 1,\ 4,\ 12$. All specifications include cohort and time 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) $k=12$
$\ln(H^q/P)$	0.502***	0.735***	0.879***	-0.050***	-0.072***	-0.086***
	(30.40)	(18.94)	(20.37)	(-16.10)	(-12.20)	(-11.42)
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Time 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	0.25	0.42	0.59	0.04	0.13	0.35

Appendix for "Housing Yields"

A Regional factors

The regional economic and social environment arguably feeds in households' expectations about discount rates and rent growth. Such expectations shaping local rent-to-price ratios 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. Hence, in this section, we study the relation between rent-to-price ratios and single district-level time-varying factors. We divide such factors in two groups: 1) demographic and economic fundamentals, and 2) housing market characteristics.

A.1 Demographic and economic fundamentals

In Appendix Table A.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 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—possibly declining—industries exhibit a more pervasive presence of displaced workers (Case and Mayer, 1996), which is in line with the previous finding on unemployment.[32] 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.

A.2 Housing market characteristics

Frictions 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, which may then transmit to valuation ratios. In other words, 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 trading frictions, 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 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 friction (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 Appendix Table A.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 median number of days a rental listing stays online.[33] The third measure is the 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. Put differently, housing valuations are increasing in the liquidity of the local housing market, which in turn tends to be higher in growing urban areas.

[32] 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.

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

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 geographic 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. Such a rate loads insignificantly.

B Costs

For robustness, we adjust our baseline measure of gross yields for costs possibly borne by property owners. Because we do not directly observe such costs, we follow several approaches to proxy for them. In the spirit of Eichholtz et al. (2021), we focus on three sources of costs: nontax costs (including expenses for maintenance), vacancy costs, and taxes. By making different assumptions, we obtain four sets of net yield estimates.

(a) To compute the first measure of net yields, we assume that nontax costs are at 30% of the annual property-specific listed rent as Eichholtz et al. (2021) do. Similar to Demers and Eisfeldt (2022), the vacancy loss is assumed to be given by the product of the local vacancy rate and the annual property-specific listed rent. The vacancy rate is measured at the district level, based on the 2011 German census, which is then assumed to be constant in the previous and subsequent years. Property taxes are computed as the product of the implicit recurrent property tax rate estimated for Germany by Barrios, Denis, Ivaškaitė-Tamošiūnė, Reut, and Vázquez Torres (2019) and the listed property-specific sale price.

(b) Relative to the first measure, the second adjusts both vacancy costs and property taxes. The vacancy costs are computed from a time-varying measure of the local vacancy rate. Starting from the 2011 census estimate of the vacancy rate, we link its variation in the surrounding years to the growth rate of property prices (e.g., an increase in residential prices relative to 2011 is assumed to lead to a lower vacancy rate). Then, the property tax component is adjusted for the district-level property assessment rate B.

(c) The third measure deviates from the second by assuming that nontax costs are equal to 1.5% of the listed property-specific sale price, as suggested by Barrios et al. (2019).

(d) The fourth measure deviates from the second by linking the sources of costs to average district-level rent and sale prices, rather than to property-specific ones.

In each case, we impose that property-level net yields cannot be lower than 0. Panel A of Appendix Table A.8 displays summary statistics for the four measures of net yields, which are on average 30% lower than gross ones.

Panel B re-estimates the same specification as in column 6 of Table 5 using the net yield as the dependent variable. No matter how we measure net yields, the adjusted R^2 lines up with the baseline result. Admittedly, the cost specifications (a)-(c) amount to implicitly assuming that costs are perfectly correlated with underlying property prices, muting any additional sources of idiosyncratic risk that could drive costs. If we were to observe costs, it is thus theoretically possible that a "diversification effect" would reduce dispersion in net rental income relative to the gross one. However, Chambers et al. (2021) empirically show that costs drive up volatility of rental income. Interestingly, even indexing costs to district- rather than property-level prices, as we do in case (d), does not produce an increase in the explanatory power of geographic fixed effects.

C Rent controls

In this section, we analyze the consequences for rent-to-price ratio heterogeneity of a specific and prominent form of regulation of the housing market widespread around Europe and present also in Germany, namely rent controls (Kholodilin, 2020). Germany has long had rent controls in place. First-generation rent control laws, which simply set rents at a certain level, were introduced already in 1922; in 1971, policy-makers moved to second-generation laws, which link allowed new rents and rent increases to a reference local rent; a nationwide cap on rent increases of 30% (over 3 years) for existing contracts was passed in 1982 and subsequently reduced to 20% in 1993 (Kholodilin, 2017). During our sample period, the federal government passed two major reforms: on rent updating in 2013 (the *capping limit*) and on rent setting in 2015 (the *rental brake*). These reforms allowed local state governments to introduce more stringent controls for specific municipalities characterized by a "tight" rental market (Kholodilin, 2017). The capping limit restricts rent increases for existing contracts at 15% over 3 years for the designated municipalities. The rental brake constraints new rents in the designated municipalities to be at most 10% higher than a reference rent from the area, except for new or modernized properties.

Also in this case, zip code-quarter fixed effects in Table 5 could absorb any increase in dispersion induced by these reforms as long as they are equally binding across the properties in a given municipality (or zip code). Suppose instead that rent controls bind only for certain properties and, at same time, that market participants expect the controls to be lifted in the future—the new municipality-level rent controls always come with an expiry date and their renewal is subject to varying degrees of uncertainty—, 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. Although nationwide rent controls were active in Germany throughout our sample period, they have been relatively loose until the 2010s, whereas cross-sectional heterogeneity in housing yields is substantial even in the early years of our sample (see Appendix Figure A.1), suggesting that the recent reforms are hardly its main driver.

To verify this conjecture, we implement a staggered difference-in-difference design on housing yields dispersion across municipalities around the time at which the new rent controls started to be enforced.[34] We estimate specifications of the following form:

$$\mathrm{SD}(\ln(H/P))_{m,t} = \alpha \cdot \textit{Rental brake}_{m,t} + \beta \cdot \textit{Capping limit}_{m,t} + \tau_m + \tau_t + \epsilon_{m,t}, \qquad \text{(A.1)}$$

where the dependent variable, $\mathrm{SD}(\ln(H/P))_{m,t}$ is the standard deviation of rent-to-price ratios in municipality m at quarter t.[35] To compute the indicators $\textit{Rental brake}_{m,t}$ and $\textit{Capping limit}_{m,t}$, we manually collected from state-level laws the list of municipalities touched by the new rent controls (for an overview, see Kholodilin, 2017). $\textit{Rental brake}_{m,t}$ is an indicator variable equal to 1 in quarters in which a given municipality is subject to the rental brake, and 0 otherwise. Starting from 2015, the rental brake has been introduced at different points in time by 13 out of 16 states. The $\textit{Capping limit}_{m,t}$ indicator is defined in a similar way. Starting from 2013, the capping limit has been adopted in a staggered fashion by 14 states. Because the house listings in our data relate to new contracts, our focus is on the rental brake reform of 2015. Yet, we control for the capping limit on existing contracts introduced in 2013 because of its potential effects on investors' expectations.[36] We then account for time-invariant differences across municipalities and for changes in macroeconomic conditions via municipality (τ_m) and time (τ_t) fixed effects. We cluster standard errors at the district level.

In Appendix Table A.9, we report coefficient estimates from equation (A.1). In column 1, we compute standard deviations of rent-to-price ratios for all properties. In column 2, we focus on properties that are neither new nor recently modernized, i.e., those that are subject to the rental brake in the designated municipalities. In column 3, we look at spillover effects of the rental brake on the dispersion of yields of new or recently modernized properties. To ameliorate overlap between treated and control municipalities, in columns 4 to 6 we re-estimate the same specifications but restricting the control group to non-designated municipalities belonging to districts comprising at least one designated municipality. If anything, the evidence is that the rental brake appears to have reduced—rather than increased—dispersion in housing yields.[37]

[34] Mense, Michelsen, and Kholodilin (2019) and Kholodilin, Mense, and Michelsen (2019) use a similar setting to examine the consequences of the rental brake for land values and new rents, respectively. Breidenbach, Eilers, and Fries (2022) perform an evaluation of the reform by means of an event-study approach.

[35] To compute these standard deviations, we remove municipality-quarters with less than 10 matched properties. We then trim the standard deviations at the 99.5% level.

[36] We do not study the *rental freeze* introduced in Berlin in 2020 because it is outside our sample period (Hahn, Kholodilin, Waltl, and Fongoni, 2023).

[37] In unreported analysis, we find that this result is robust to using the CSDID estimator developed by Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020), which helps mitigate the inference issues linked to staggered difference-in-differences designs such as ours.

D Other figures and tables

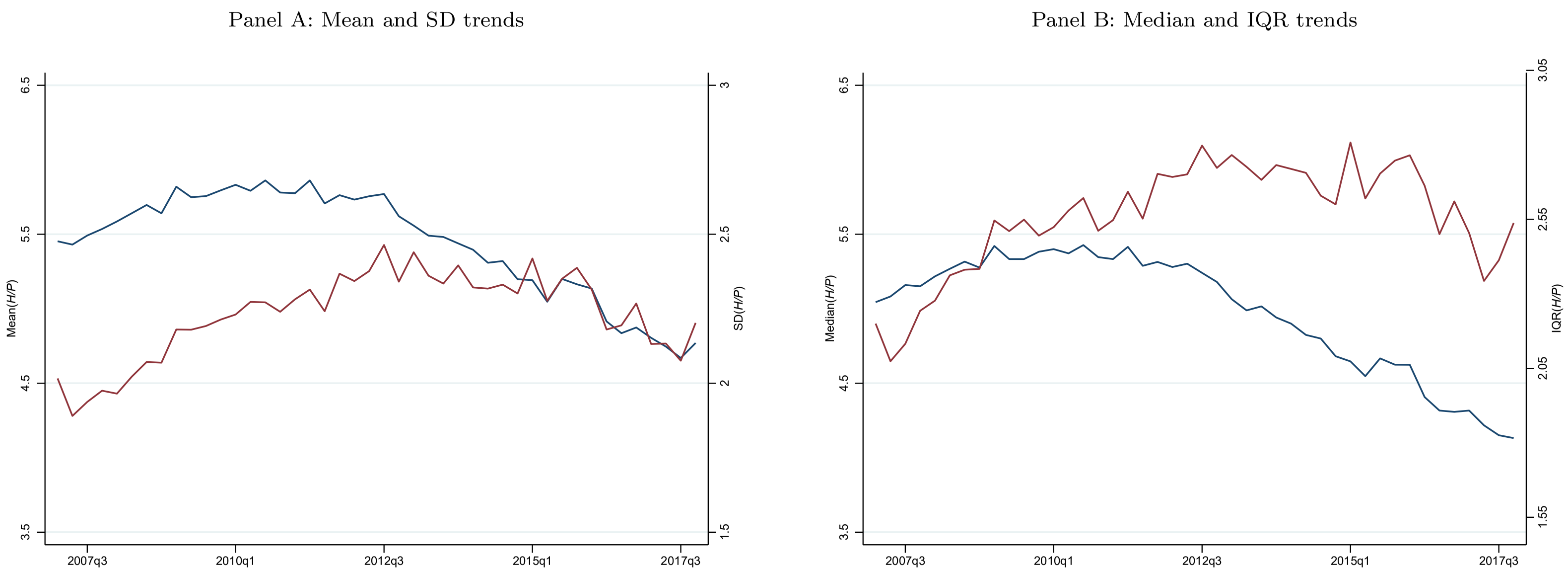


Figure A.1: Level and dispersion of rent-to-price ratios over time
This figure shows the evolution of the level and dispersion of rent-to-price ratios (obtained via the baseline matching) between 2007 and 2017. Panel A reports the mean and the standard deviation dynamics (blue and red line, respectively). 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 zip code level within other major cities
This figure visualizes the median rent-to-price ratio (obtained via the baseline matching) at the five-digit zip code level across selected German major cities, pooling all periods between 2007 and 2017. Grey-colored zip code 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.

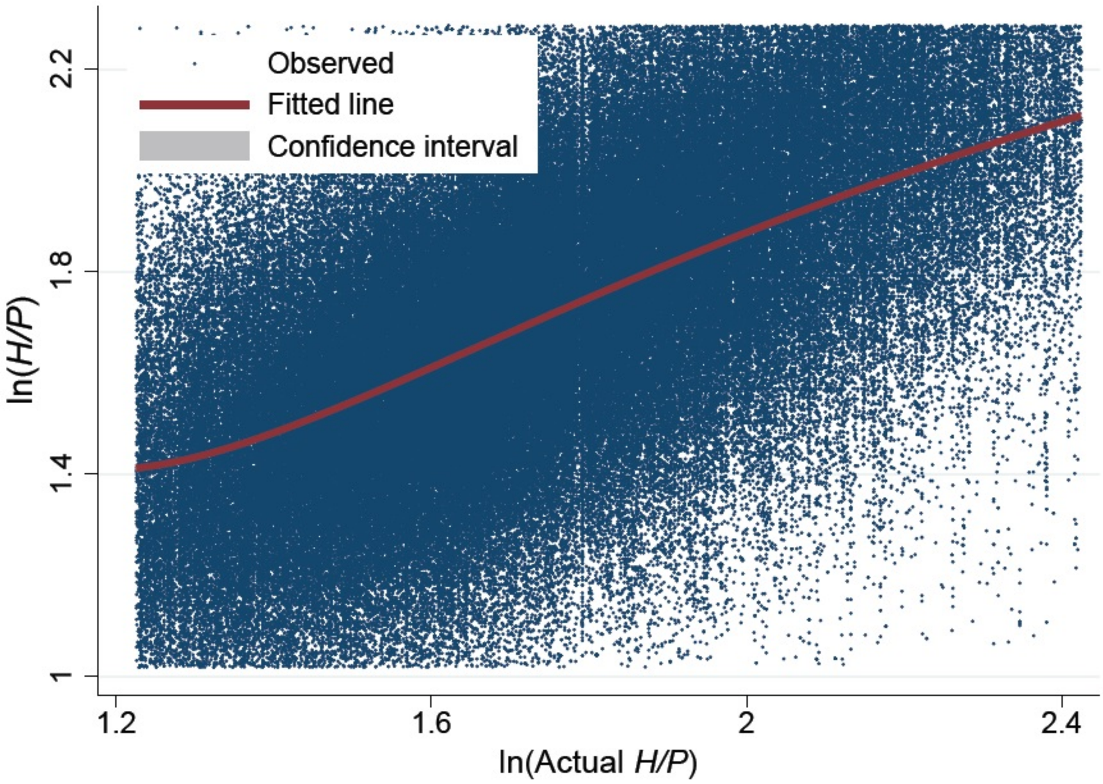


Figure A.3: Validation of rent-to-price ratios obtained via the baseline matching
This figure shows a scatter plot of rent-to-price ratios obtained via the baseline matching against their actual counterparts (all in natural logarithm). Actual rent-to-price ratios refer to properties on sale for which 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).

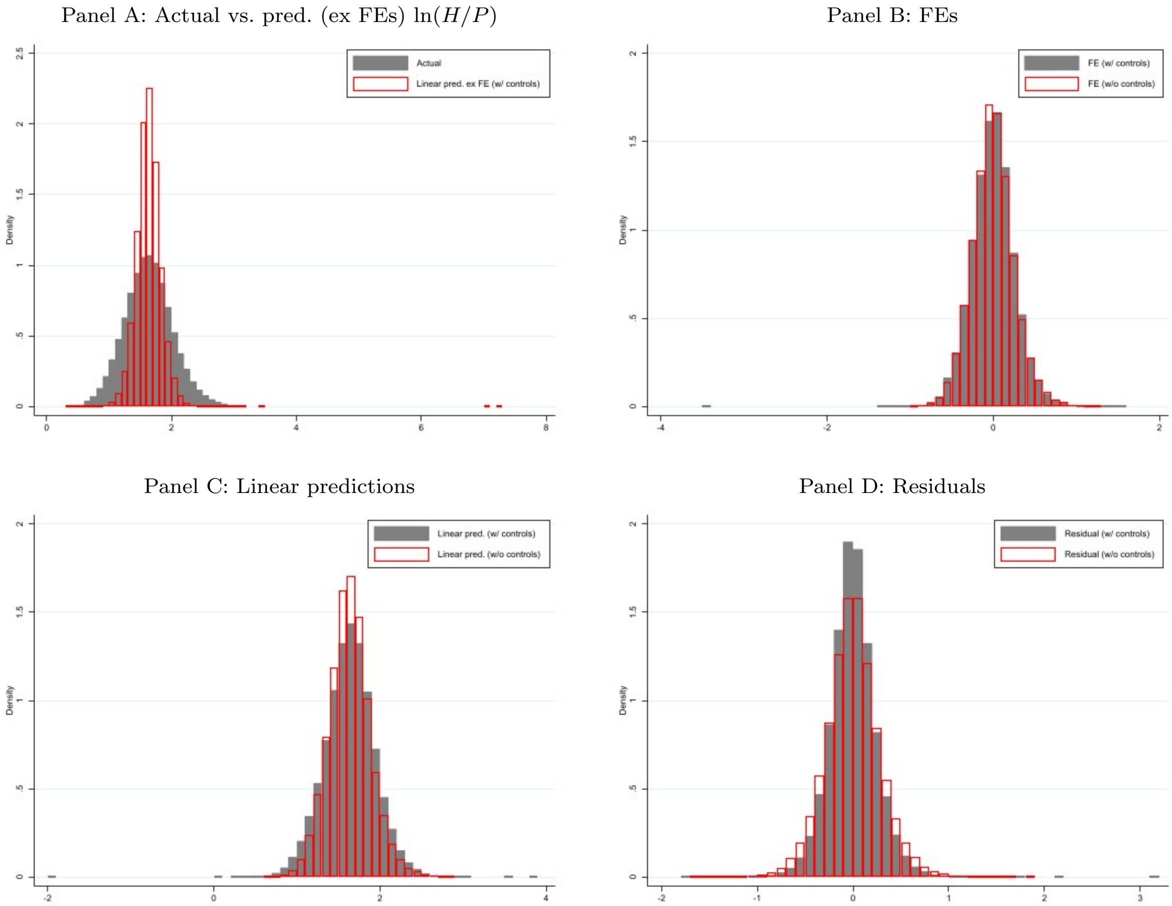


Figure A.4: The role of outliers
This figure shows densities for quantities obtained from the baseline specification in column 6 of Table 5 against those from a similar specification not including property-specific control variables. Panel A compares actual log-transformed rent-to-price ratios against fitted values (excluding estimated zip code-quarter fixed effects) from the baseline specification. Panels B, C, and D compare, respectively, estimated fixed effects, fitted values (including estimated fixed effects), and residuals from the two specifications.

Panel A: FEs $\ln(H/P)$ Panel B: Residuals

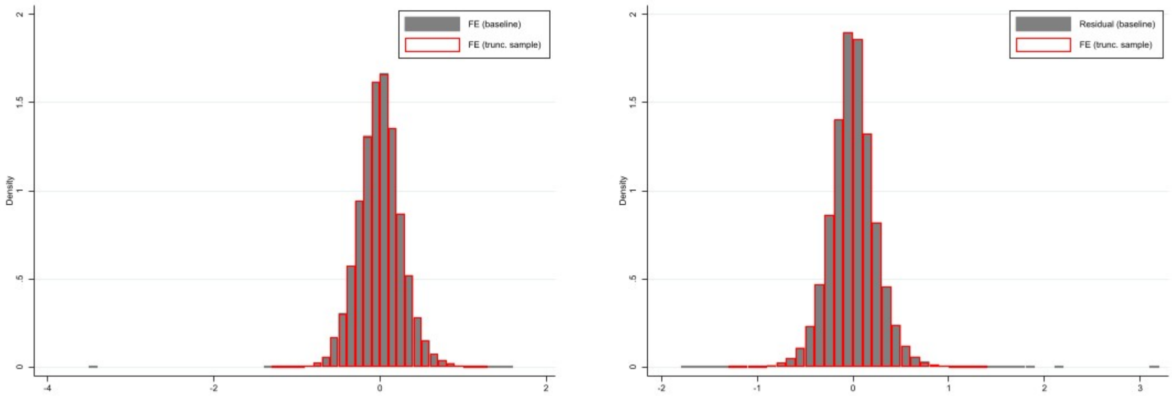
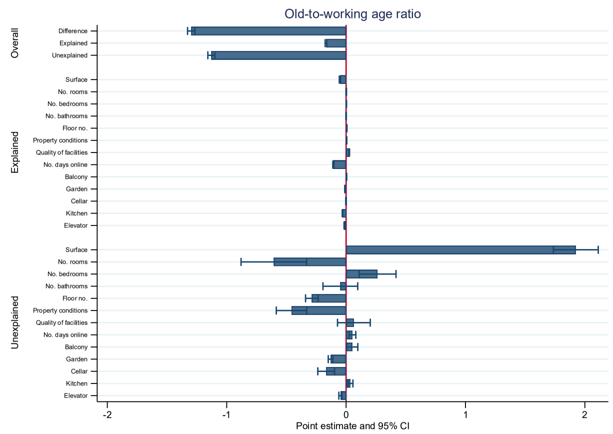
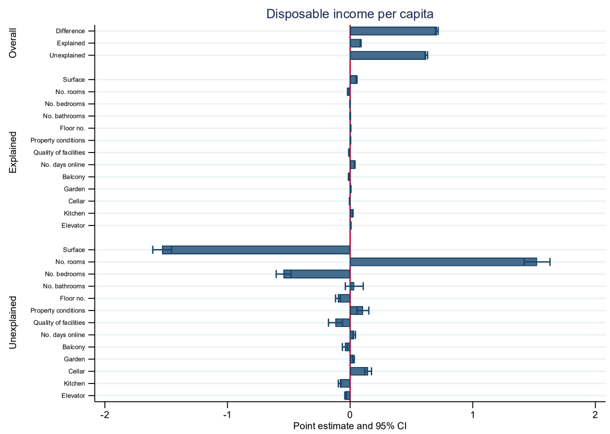


Figure A.5: Truncating the regression sample
This figure shows densities for quantities obtained from the baseline specification in column 6 of Table 5 against those from the same specification estimated over a sample that excludes observations linked to outliers. Such observations are identified as those for which either the estimated fixed effect or the residual are more than five standard deviations away from their mean. Panels A and B compare, respectively, estimated fixed effects and residuals from the two specifications.

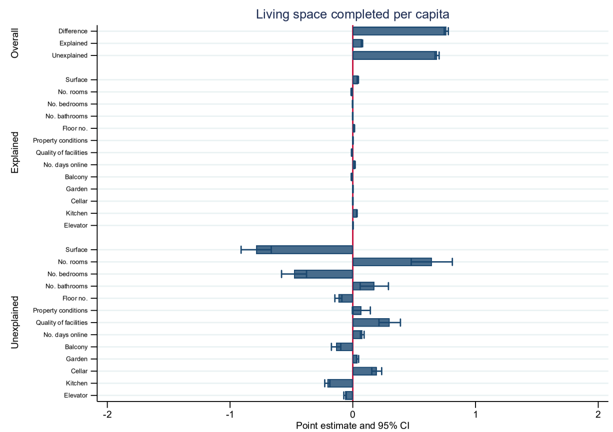
Panel A: Old-to-working age ratio



Panel B: Disposable income per capita



Panel C: Living space completed per capita



Panel D: No. businesses

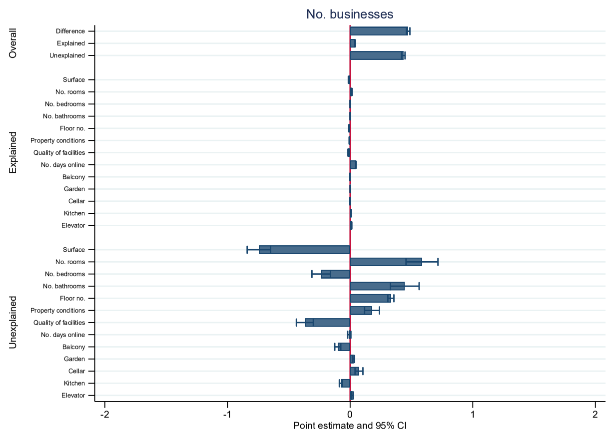


Figure A.6: Blinder-Oaxaca decomposition of the difference in rent-to-price ratios across regions
Each panel in this figure shows the output of a Blinder-Oaxaca decomposition of the difference in rent-to-price ratios (obtained via the baseline matching) between properties located in districts belonging 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.

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 exclusive of expenses for utilities (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 for utilities in EUR. Available only for flats for rent.
Annual inclusive rent	RWI-GEO-RED	Annual rent inclusive of expenses for utilities 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 (four 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.
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.

(Continued)

Table A.1: – *Continued*

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.
Median online listing time	RWI-GEO-RED	Median number of days online of the rental 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 B	Federal Statistical Office	Property tax multiplier (type B) in a given district-year.
Global city	Globalization and World Cities Research Network	Indicator equal to one if a district has a rating between “Alpha” and “Gamma”.

Table A.2: Rent-to-price ratio, property characteristics, and macroeconomic conditions
This table re-estimates the specification from column 4 of Table 4, augmenting it with interactions with measures of macroeconomic conditions in Germany. Column 1 interacts each of the matching covariates (surface (squared), no. of rooms, no. of bedrooms, no. of bathrooms, floor number) with the yield-to-maturity on the 10-year Bund. Column 2 adds similar interactions with GDP per capita growth in Germany. Column 3 adds similar interactions with housing stock per capita growth in Germany. 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)
Flat covariates	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes
Matching covariates×10Y YTM (Bund)	Yes	Yes	Yes
Matching covariates×GDP per capita growth	No	Yes	Yes
Matching covariates×Housing stock per capita growth	No	No	Yes
Time FE	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat
Mean(y)	1.64	1.64	1.64
SD(y)	0.38	0.38	0.38
Observations	1,613,889	1,613,889	1,613,889
Adjusted R^2	0.30	0.30	0.30

Table A.3: Local variation of rent-to-price ratios with sample truncation
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 the baseline matching procedure of flats for rent to flats for sale. Column 1 reports the same estimates in column 6 of Table 5. Columns 2-3 re-estimate the same specification over samples excluding observations that generate outliers in estimated zip code-quarter fixed effects or residuals. In column 2, the sample is truncated at ten standard deviations from the mean of the estimated fixed effect or of the residual. In column 3, the threshold is set at five standard deviations. Standard errors are clustered by district. Refer to Appendix Table A.1 for variable definitions.

	$\ln(H/P)$		
	(1)	(2)	(3)
Flat covariates	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes
Zip code×Time FE	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat
Tail truncation	None	10 SD	5 SD
Mean(y)	1.64	1.64	1.64
SD(y)	0.38	0.38	0.38
Observations	1,595,086	1,595,083	1,594,684
Adjusted R^2	0.59	0.59	0.59
Adjusted within R^2	0.31	0.31	0.31

Table A.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 the baseline matching procedure of flats for rent to flats for sale. Each column augments the specification of column 4 of Table 4 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 (%)	0.020***						
	(7.90)						
ln(Population)		-0.031***					
		(-3.64)					
Disposable income per capita			-0.020***				
			(-3.25)				
GDP per capita				-0.003***			
				(-4.38)			
Unemployment rate					0.013**		
					(2.45)		
Manufacturing industry share						0.019***	
						(4.58)	
ln(No. businesses)							-0.043***
							(-4.76)
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)	1.64	1.64	1.64	1.64	1.64	1.64	1.64
SD(y)	0.38	0.38	0.38	0.38	0.38	0.38	0.38
Observations	1,613,889	1,613,889	1,603,537	1,600,773	1,613,545	1,595,897	1,613,871
Adjusted R^2	0.34	0.31	0.32	0.32	0.31	0.31	0.31

Table A.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 the baseline matching procedure of flats for rent to flats for sale. Each column augments the specification of column 4 of Table 4 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***						
	(-6.35)						
Median online listing time		0.008***					
		(6.15)					
ln(No. posted flats)			-0.022***				
			(-4.20)				
Housing stock per capita				0.008**			
				(2.48)			
Living space completed per capita					-0.486***		
					(-5.24)		
Land price						-0.000***	
						(-8.87)	
Property assessment rate B							-0.000
							(-0.61)
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)	1.64	1.64	1.64	1.64	1.64	1.65	1.64
SD(y)	0.38	0.38	0.38	0.38	0.38	0.38	0.38
Observations	1,613,889	1,613,889	1,613,889	1,603,511	1,599,193	1,483,995	1,613,123
Adjusted R^2	0.33	0.32	0.31	0.31	0.32	0.33	0.30

Table A.6: 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 the baseline 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*
	(6.54)	(1.81)
Disposable income per capita	-0.032***	-0.002**
	(-9.86)	(-2.04)
Living space completed per capita	-0.414***	-0.049***
	(-8.47)	(-2.63)
ln(No. businesses)	0.004	0.033***
	(0.42)	(8.70)
Time FE	Yes	Yes
Unit of observation	District-time	District-time
Mean(y)	1.66	0.31
SD(y)	0.27	0.11
Observations	14,837	13,870
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.7: Re-estimating the main specifications with non-transformed rent-to-price ratios
This table re-estimates the specifications from column 4 of Table 4, column 1 of Table 6, and column 6 of Table 5, using non-transformed property-level rent-to-price ratio as the dependent 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.

	H/P		
	(1)	(2)	(3)
Surface	-0.112***	-0.122***	-0.121***
	(-16.42)	(-28.80)	(-31.54)
Surface squared	0.001***	0.001***	0.001***
	(21.46)	(27.55)	(28.85)
No. rooms	0.684***	0.483***	0.176***
	(14.55)	(11.59)	(5.80)
No. bedrooms	-0.123***	0.045	0.067***
	(-3.19)	(1.49)	(3.46)
No. bathrooms	-0.238***	-0.308***	-0.317***
	(-4.75)	(-7.12)	(-7.26)
Floor no.	0.047**	0.043**	0.048***
	(2.49)	(2.53)	(4.24)
Expenses	-0.001**	-0.000	0.000
	(-2.30)	(-0.40)	(0.67)
Heating expenses	-0.001**	-0.002**	-0.000
	(-2.39)	(-2.45)	(-1.17)
Heating included	0.034	0.001	0.076***
	(0.56)	(0.04)	(5.49)
Deposit	-0.000***	0.000***	0.001***
	(-4.81)	(5.09)	(13.79)
Housing benefits	0.001**	0.001***	0.001***
	(2.37)	(4.04)	(4.91)
Holiday property	-0.015***	-0.018***	-0.016***
	(-6.43)	(-8.61)	(-7.24)
Rented	0.356***	0.289***	0.328***
	(5.34)	(4.88)	(6.27)
Old-to-working age ratio		0.103***	
		(6.58)	
Disposable income per capita		-0.112***	
		(-6.71)	
Living space completed per capita		-1.660***	
		(-4.72)	
ln(No. businesses)		-0.175***	
		(-2.98)	
Missing indicators	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes
Time FE	Yes	Yes	No
Zip code×Time FE	No	No	Yes
Unit of observation	Match. flat	Match. flat	Match. flat
Mean(y)	5.53	5.53	5.53
SD(y)	2.26	2.26	2.26
Observations	1,613,889	1,588,823	1,595,086
Adjusted R^2	0.25	0.32	0.55
Adjusted within R^2	0.24	0.31	0.26

Table A.8: The role of costs in the variation of rent-to-price ratios

This table reports tests on net rent-to-price ratios for a German sample between 2007 and 2017. The log-transformed gross rent-to-price ratio is obtained via the baseline matching procedure of flats for rent to flats for sale. The net rent-to-price ratio is then computed following the four different cost specifications ((a)-(d)) described in Appendix B. Panel A shows summary statistics on gross and net rent-to-price ratios. Panel B reports coefficients of determination from regressions for rent-to-price ratios. Each column re-estimates the specification of column 6 of Table 5 using a different measure of the net rent-to-price ratio as the dependent variable. Standard errors are clustered by district. Refer to Appendix Table A.1 for variable definitions.

Panel A: Basic statistics

	Obs.	Mean	SD	Min	P25	Median	P75	Max
H/P	1,613,889	5.530	2.262	1.816	3.969	5.094	6.557	17.829
Net H/P (a)	1,610,326	3.507	1.447	0.912	2.506	3.241	4.187	11.984
Net H/P (b)	1,605,145	3.478	1.445	0.830	2.479	3.215	4.158	12.013
Net H/P (c)	1,605,145	3.636	2.119	0.000	2.172	3.245	4.624	15.831
Net H/P (d)	1,605,145	3.462	1.539	0.000	2.408	3.178	4.171	15.291

Panel B: Regression analysis

	ln(Net H/P)			
	(1)	(2)	(3)	(4)
Flat covariates	Yes	Yes	Yes	Yes
Missing indicators	Yes	Yes	Yes	Yes
Covariate distances	Yes	Yes	Yes	Yes
Zip code×Time FE	Yes	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat
Measure of costs	(a)	(b)	(c)	(d)
Mean(y)	1.18	1.17	1.12	1.12
SD(y)	0.39	0.39	0.64	0.64
Observations	1,591,924	1,587,511	1,587,296	1,587,296
Adjusted R^2	0.58	0.58	0.57	0.57
Adjusted within R^2	0.31	0.31	0.31	0.31

Table A.9: Dispersion of rent-to-price ratios and rent controls
This table reports estimates from regressions for the dispersion in rent-to-price ratios on measures of stringency of local rent controls for a German sample between 2007 and 2017, focusing on the impact of the so-called rental brake of 2015. The dependent variable is municipality-quarter-level standard deviation of log-transformed rent-to-price ratios obtained via the baseline matching procedure of flats for rent to flats for sale (at least 10 non-missing observations are required to compute each standard deviation). Columns 1 and 4 consider all properties. Column 2 and 5 consider only non-new properties. Columns 3 and 6 consider only new properties. Columns 1 to 3 use as control group all municipality-quarters not subject to the rental brake. Columns 1 to 3 use as control group municipality-quarters not subject to the rental brake but within districts comprising at least one municipality-quarter subject to the rental brake. All specifications include municipality and 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.

	SD(ln(H/P))					
	(1)	(2)	(3)	(4)	(5)	(6)
Rental brake	-0.007	-0.007	-0.015	-0.013*	-0.012*	0.011
	(-1.09)	(-1.05)	(-1.21)	(-1.97)	(-1.67)	(0.47)
Capping limit	-0.005	-0.004	-0.010	-0.009	-0.010*	-0.006
	(-0.97)	(-0.65)	(-1.19)	(-1.57)	(-1.69)	(-0.54)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Municip.-time	Municip.-time	Municip.-time	Municip.-time	Municip.-time	Municip.-time
Properties	All	Non-new	New	All	Non-new	New
Control group	All	All	All	Treated distr.	Treated distr.	Treated distr.
Mean(y)	0.29	0.29	0.31	0.27	0.27	0.29
SD(y)	0.10	0.10	0.10	0.09	0.09	0.09
Observations	13,006	12,111	1,747	6,124	5,734	1,144
Adjusted R^2	0.35	0.35	0.40	0.43	0.42	0.42

Table A.10: Predictive regressions controlling for local conditions
This table reports estimates from predictive regressions for the 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 $k = 1,\ 4,\ 12$. 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