Housing Yields

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

June 10, 2026

Abstract

We build a granular dataset of residential property yields using rental and sale listings from a major German real estate platform. With more than 1.5 million property-level rent-to-price ratios, we document a novel *heterogeneity puzzle*. About two-fifths of yield dispersion remains unexplained after controlling for extensive observable property characteristics and postal code-level time-varying factors using fixed effects. When we absorb finer neighborhood amenities through richer fixed effects, the unexplained share remains close to one-third. This evidence points to the importance of investors' beliefs and preferences—rather than several alternative explanations we test—as sources of heterogeneity in risk premia.

JEL Classification: G12, G51, R31

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

[*]We thank the Editor, George Pennacchi, and two anonymous referees for constructive feedback and guidance. Moreover, we are indebted to 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 authors used ChatGPT-5.5 to assist with proofreading, language editing, and improving the clarity and flow of the manuscript. After utilizing this tool, the authors reviewed and edited the AI-assisted text and take full responsibility for the final content.
[†]Corresponding author. Ca' Foscari University of Venice (Department of Economics) 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 (ESOMAS Department) and Collegio Carlo Alberto, Italy. E-mail: roberto.marfe@carloalberto.org.
[§]University of Oxford (Said Business School), United Kingdom. E-mail: qizhou.xiong@said.oxford.edu.

I. 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 (Plazzi, Torous, and Valkanov (2010)).[3]

[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 rate of 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)). Note that throughout the remainder of the paper, we use the terms "housing yield", "rent-to-price ratio", "valuation ratio", and

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 postal code-level time-varying factors nor by a rich set of property-level characteristics. Such heterogeneity is economically sizable, with the 90th−10th percentile range of unexplained yields in our dataset corresponding to a variation of EUR 71,837 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' beliefs and preferences. These forces, in turn, presumably generate heterogeneity in both risk premia and expected rent growth, with some evidence pointing to a more prominent role for the former. By contrast, proxies for local agglomeration effects, housing supply rigidities, informational frictions, regulation, climate risk, and ownership structure do not appear to command sizable variation in housing yields over and above postal 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 postal codes within the same city. Time-series variation in rent-to-price ratios is largely property-specific or at the postal code level, with macroeconomic fluctuations accounting only for up to half of it.

"valuation" interchangeably.

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 postal code-level). Even after controlling for an extensive set of observable property-level characteristics and fine fixed effects at the postal 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 notably symmetric distribution of residuals, in which outliers appear to have a limited role.

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 proxies for agglomeration economies, informational frictions, rigidities in housing supply, regulation, climate risk, or ownership structure in explaining within-postal code-quarter dispersion in yields.

The 41% of variation in rent-to-price ratios that remains unexplained—or 29% when accounting for neighborhood amenities in an alternative matched sample—is noteworthy. This is even more striking given that observable property characteristics alone deliver an adjusted R^2 of only 30%, while the remaining explained variation is captured by fixed effects that subsume factors ultimately unobservable in our dataset.

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 postal 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] Although a clear benchmark for a "satisfactory" adjusted R^2 in housing yield regressions is hard to define, we argue that the explanatory power of our richest specification is remarkably low, which calls into question common wisdom about the drivers of residential property valuations.

From an asset pricing perspective, indeed, 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.

Distinguishing between these channels—risk premia vs. expected rent growth—would ideally require observing the same properties over time, which is challenging because our listings data come as repeated cross-sections. We therefore construct a pseudo-panel by aggregating properties into cohorts defined by location, number of rooms, and size category, which allows us to compute quarterly rent-to-price ratios, excess returns, and rent growth rates. Consistent with the present-value logic of housing valuation (Plazzi et al. (2010)), we show that rent-to-price ratios predict both future excess returns and rent growth, corroborating the economic content of our synthetic yield measure. Moreover, predictability is stronger for returns in cohorts with high cross-sectional dispersion in yields, while rent-growth predictability is relatively more pronounced in low-dispersion cohorts. This pattern suggests that local yield dispersion is more closely related to heterogeneity in risk premia than to differences in expected rent growth.

Finally, it is worth discussing the generalizability of our findings. The German housing

[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.

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) shows 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.

A growing asset-pricing literature studies the cross-section of real estate valuations. Sinai and Souleles (2005) analyze the determinants of residential rent-to-price ratios using MSA-level data from the US, focusing on the trade-off between rent risk for tenants and house price risk for owners. Han (2013) studies how the risk-return relation for residential properties varies across US local markets. Using US postal code-level data on housing returns, Eiling, Giambona, Lopez Aliouchkin, and Tuijp (2020) document substantial dispersion across MSAs in the sources of risk priced in housing returns, with idiosyncratic risk playing a sizable role. Giacoletti (2021) and Sagi (2021) use residential and commercial property-level data to study the idiosyncratic component of real estate prices, showing that it does not follow a random walk, contrary to a common assumption in asset-pricing models. Kantak (2025) examines the relation between local expected economic growth and price-to-rent ratios at the MSA level in the US, documenting a negative cross-sectional association and explaining it through an asset-pricing model with housing consumption commitments. Relatedly, Feng, Jaimovich, Rao, Terry, and Vincent (2023) show that local industry structure, and the manufacturing share in particular, helps explain inequality in housing valuations. Demers and Eisfeldt (2022) decompose total returns to U.S. single-family rentals into net rental yields and house price appreciation, showing that yields and capital gains are negatively correlated across cities. Then, Halket (2025) documents large within-city variation in total returns to single-family houses across U.S. neighborhoods, driven primarily by differences in net yields. We contribute to this literature by studying the distribution and drivers of property-level rent-to-price ratios using a large and granular dataset for Germany. We show that a substantial and economically meaningful share of their variation cannot be explained by local factors or by observable property characteristics. In the Online Appendix, we further discuss how our contribution relates to other strands of the literature.

II. Data and Housing Yields Construction

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, postal 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, Dohmen, Kohl, and Schularick (2025)). 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.

Nationwide data on inflation and interest rates are from Federal Reserve Economic Data (FRED) of the St. Louis Federal Reserve Bank. Valuation ratios for stocks listed on the New York Stock Exchange (NYSE) are from Kenneth French's website. Regional data, used for tests in the Online Appendix, are from the German Federal Statistical Office. Additional datasets used for specific tests, along with a detailed discussion of the corresponding results, are described in the Online Appendix.

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, we trim variables at the 0.5% and 99.5% level.

A. Matching Approach

The rent-to-price ratio of a given property is an unobservable quantity as we typically do not simultaneously observe the rental and sale price of the same property. We follow different matching schemes to obtain counterfactual sale prices for flats for rent.

Baseline Matching Scheme. For the baseline analysis, we adopt a parsimonious set of matching covariates balancing the number of attainable matches against their quality, namely: flat surface (distance minimization), number of rooms, number of bedrooms, number of bathrooms, floor category, five-digit postal 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 these 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 on those dimensions.

Alternative Matching Schemes. We then implement four alternative matching exercises.

(a) First, we carry out a propensity-score matching (PSM) approach by estimating a logit model for the rental indicator on a comprehensive set of flat-level covariates across the full sample of listed properties. We then perform 1:1 nearest-neighbor matching with replacement and a caliper of 0.1, imposing exact matching on postal code and calendar quarter.

(b) Second, we use a matching scheme that follows the baseline algorithm but allows properties to be matched across different time periods. To ensure comparability, we adjust matched prices for local house price inflation and include, among the matching criteria, the distance in listing quarters between paired properties.

(c) Third, 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 sale flat are in the same conditions allows us to better control for quality differences.

(d) Finally, we impose a more precise matching in terms of flat location. Indeed, whereas in Germany five-digit postal codes are quite 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)). We thus 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 postal 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 postal 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.[5] Nonetheless, forcing matched 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, Gaubert, Handbury, and Hurst (2024)).

The schemes (a) and (b) substantially expand the set of matched observations and mitigate sample bias, likely at the cost of match quality. In contrast, the schemes (c) and (d) 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. For these reasons, below we use these alternative schemes 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.

B. Computation of Housing Yields

In the baseline matching, we remove any match for which we do not obtain exact matching on each considered discrete variable or for which the absolute distance in terms of surface is larger than 10 sqm (in robustness tests below, we also consider a lower threshold for such a distance). 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).[6] Given each match between the flat for rent and the counterfactual flat for sale, we compute the natural logarithm of the

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

[6] Allowing for multiple matches helps us have less volatile estimates of the sale price of properties for rent. At the same time, however, we remove extremes (and rare) cases, such as those properties with more than 20 matches.

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

$$(1) \qquad \ln\left(\frac{H_{f,t}}{P_{f,t}}\right) = \ln\left(100 \times \frac{12 \times H_{r,t}}{P_{\bar{s},t}}\right),$$

where r, s, and t denote the flat for rent, the flat for sale, and the calendar quarter, respectively.[7] Following the notation of Plazzi et al. (2010), $H_{r,t}$ is the monthly rent per sqm 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 per sqm, $P_{\bar{s},t}$, 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.

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, Spaenjers, and Steiner (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 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, Korevaar, Lindenthal, and Tallec (2021), Demers and Eisfeldt (2022)).

[7] We express the rent-to-price ratio in annual percentage terms to favor its readability in the cases in which we do not consider its logarithmic transformation. In baseline regressions, however, we use this transformation, under which the multiplication by 100×12 in equation (1) boils down to a constant, leaving the other coefficient estimates unchanged.

III. Heterogeneity in Housing Yields

This section documents the extent of heterogeneity in housing yields and progressively assesses how much of it can be attributed to observable property characteristics and local geographic factors.

A. Descriptive Statistics

We start with a naive examination of the degree of heterogeneity in synthetic rent-to-price ratios from equation (1). Table 1 reports summary statistics for flat characteristics under the baseline matching approach.[8] The final sample contains 1,625,131 flats for rent matched to counterfactual flats for sale.[9] By construction, the differences between these two groups of flats are statistically indistinguishable from zero for matching covariates—exactly zero in the case of covariates on which we impose exact matching, such as the number of rooms. Still, based on t-tests allowing for district-level clustering, flats for rent display statistically significant differences from flats for sale with respect to most of the other covariates, although such differences are typically modest in magnitude. Nonetheless, below we augment rent-to-price ratio regression specifications with these observable differences to absorb possible systematic patterns in ratios arising artificially from the matching exercise.

[Insert Table 1 about here]

With this caveat in mind, Figure 1 visualizes the empirical distribution of the synthetic rent-to-price ratio together with the national and local macroeconomic conditions over the sample period. In Graph A, we examine the distribution conditional on the size

[8] Variable definitions are presented in Table A.1. Variables used for additional tests in the Online Appendix are defined in Table OA.1.

[9] 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.

category of the property, uncovering a positive relation between valuations and flat surface. Variation across categories is limited with a median roughly ranging between 6% for small flats and 4.5% for very large flats. In Graph 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. Not surprisingly, the economic success of states appears to correlate negatively with rent-to-price ratios. Graph C highlights that the median ratio moderately increases up to 2010 to then decline. This pattern is mirrored, although less markedly, by within-period heterogeneity, which exhibits an initial upward trend followed by a decrease toward the end of the sample period.[10] This is broadly consistent with the economic conditions observed both nationally and locally in Germany over 2007-2017. Over the sample, the only recession was the 2008-2009 whereas around the European debt crisis a mere slowdown took place (Graph D). Periods of stronger economic conditions are associated with sales prices growing faster than rents—implying lower housing yields—and with a concurrent compression of cross-sectional dispersion. While, in general, this pattern could mechanically arise from nominal rigidities in rents, this explanation is less compelling in our setting since we focus on new rental contracts, which should adjust more flexibly.

[Insert Figure 1 about here]

This first aggregate evidence also suggests that variation in rent-to-price ratios is to a large extent cross-sectional, relating to 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, as highlighted by Figure 2. Graph A 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

[10] Consistently, Figure OA.1 shows a hump-shaped behavior of both the level and dispersion of rent-to-price ratios.

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 postal code level in Graphs B-H, we observe that media ratios greatly vary even within some of the most thriving metropolitan areas like Hamburg or Frankfurt.[11]

[Insert Figure 2 about here]

B. A Geographical Decomposition of Time-Series Variation in Housing Yields

In the spirit of Piazzesi, Schneider, and Tuzel (2007), we then decompose time-series variation of valuation ratios into postal code-, district-, federal state-, and national level variation. For each geographic area, we compute the average of rent-to-prices ratios quarter-by-quarter so to obtain time-series, for which we estimate standard deviation, coefficient of variation, and interquartile range. Table 2 reports the average of these statistics across geographic units. Similarly to Piazzesi and Schneider (2016), we find that national variation in rent-to-price ratios only explains up to half of postal code-level variation. 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 postal code areas (around 1%) against property-level 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. The analysis below centers on

[11]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 Figure OA.2.

providing bounds to the fraction of variation in yields that can be explained by local as opposed to idiosyncratic factors.

[Insert Table 2 about here]

Our estimates of variation in rent-to-price ratios for Germany can then be compared to those from other countries. Halket (2025), in an exercise akin to our Table 2, compute postal 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 0.62% and 2.13%. Our postal 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 (2025) range between 4.52% and 8.01%. The average gross yield in our sample is 5.51%. 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.

C. A Comparison Against Equity Yields

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. 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.[12] 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.

Moreover, based on the coefficient of variation of the time-series of average nationwide rental and sale prices per sqm, the former appear stickier than the latter (0.065 vs. 0.166), even though our rents refer to newly listed prices and are thus less subject to the nominal rigidities that characterize contracts in place. In other words, though new ask rental prices are more akin to market prices than stock dividends—which are subject to managerial discretion—they exhibit similarly contained variation over time. We confirm this finding below in a pseudo-panel setting, in which we can more reliably track the time-series variation of rents and prices for similar flats. Importantly, because we observe only newly posted rents rather than the stock of ongoing rental contracts, our estimates likely represent an upper bound on the true time-series variability of rents.

[Insert Table 3 about here]

D. Observable Property Characteristics

Next, we assess the role of observable property-specific hedonic characteristics in explaining equilibrium valuation ratios. We consider all property characteristics observable

[12] 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.

to us, including those not used to obtain the matched rent-to-price ratio from equation (1):

$$\ln\left(\frac{H_{f,t}}{P_{f,t}}\right) = \boldsymbol{\gamma}\mathbf{m}_{r,\bar{s},t} + \boldsymbol{\eta}\mathbf{x}_{r,t} + \boldsymbol{\theta}\mathbf{x}_{\bar{s},t} + \boldsymbol{\zeta}\mathbf{z}_{r,\bar{s},t} + \tau_t + \epsilon_{f,t}. \tag{2}$$

$\mathbf{m}_{r,\bar{s}}$ 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), with the addition of squared surface.[13] We control for matching covariates as they constitute some key dimensions along which our matched pairs of flats differ *across* each other; at the same time, because of the matching, there is typically no variation *within* each pair along such dimensions. $\mathbf{x}_r$ 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 (community charge, 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 its median value. 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.[14] The standard errors are clustered at the district level.

Table 4 presents coefficient estimates from specification (2). Column 1 includes the covariates on which we match properties. The coefficients on flat size imply a hump-shaped relationship between surface and housing yields, indicating that the latter increase at low levels of surface but decline as surface becomes larger, after controlling for other characteristics. Similarly, the correlation of the rent-to-price ratio with the number of rooms is positive. Interestingly, bedrooms and bathrooms exhibit a negative, although not statistically significant for the latter, association with rent-to-price ratios, possibly

[13]We take the average between the surface of the rented flat and matched flats for sale. Where appropriate, we use log-transformed covariates (e.g., surface).

[14]Regressions models in the main analysis are estimated via the Stata package REGHDFE by Correia (2018).

pointing to a value-increasing 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 higher valuations. 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.[15]

The correlation patterns described so far remain overall robust once we include all remaining observable covariates in column 4, with adjusted R^2 rising by 18% to 30% (from 12% in column 3).[16] Heterogeneity in the rent-to-price ratio is unlikely to be uniquely a by-product of observable flat traits.

[Insert Table 4 about here]

E. Local vs. Idiosyncratic Factors

The analyses conducted so far suggest that cross-sectional variation in rent-to-price ratios is substantial. 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 postal 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.

[15] Matched and actual log rent-to-price ratios are strongly correlated (63.77%). Figure OA.3 shows that the two measures are similarly distributed, although matched yields are slightly lower than actual yields for the same properties (median 5.09% vs. 5.65%), consistent with some selection of lower-quality properties into the rental market. The PSM approach below helps tackle possible concerns about this issue.

[16] In Table OA.2, we augment specification (2) with interactions between hedonic matching covariates and national macroeconomic variables to allow for time-varying coefficients; the adjusted R^2 remains at 30%.

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 40% and 44%, respectively. In columns 5 and 6, we introduce postal code and postal code-quarter fixed effects, reaching an adjusted R^2 of 47% and 59%, respectively.[17]

[Insert Table 5 about here]

Figure 3 sheds light on the distribution of residuals from these regressions, starting from a specification without any fixed effect up to one featuring postal 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 (bottom) quartile yield subsample, the residuals from the model without postal code-quarter fixed effects exhibit substantially negative (positive) mean and median. Hence, the 41% unexplained variation of

[17] Table OA.3 re-estimates the baseline specifications without log-transforming the dependent or explanatory variables. The adjusted R^2 upper bounds decline, indicating that the unexplained variation is not driven by the log specification.

yields from column 6 of Table 5 stems from a remarkably symmetric distribution, in which outliers seem to play a limited role.[18]

[Insert Figure 3 about here]

Time-varying factors at the postal 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%. And the challenge is not only about identifying the possible sources of the 41% of unexplained variation, but also about pinning down the *actual* local time-varying factors behind the 29% (= 59% − 30%) of variation captured by postal code-quarter fixed effects.

To attack this problem, 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. After selecting relevant demographic, economic, and local housing market factors (Tables OA.5, OA.6, and OA.7), we assess how much of the local fixed effects they explain (Table OA.8). These district-level variables account for only about two-fifths of the 14% (= 44% − 30%) of yield variation captured by district-quarter fixed effects.

To sum up, although specifications featuring postal code-quarter fixed effects explain around three-fifths of variation in yields, the part that we can re-conduct to specific property- or regional-level factors is substantially lower.

IV. 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

[18] We confirm the limited presence and impact of outliers in Figures OA.4 and OA.5, and in Table OA.4.

(2021)). But unexplained variation in yields may not be strictly idiosyncratic (e.g., related to random liquidity shocks to landlords selling or renting the property at fire-sale 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 issues, local amenities and agglomeration economies, and housing market frictions (illiquity, supply rigidities, regulation).

A. Measurement Issues

Despite the richness of the observable property-level traits we match on or control for, several sources of measurement error may still contribute to the heterogeneity in housing yields that we document. First, listed gross yields may differ from the returns relevant to market participants because of expenses faced by tenants that are not fully reflected in listed rents, or because of costs borne by owners that drive a wedge between gross and net returns. Second, measurement error may arise from imperfect matching between rental and sale properties, either because relevant property-level dimensions such as finishes, noise, and views from the flat are omitted, or because matched properties still differ along observed characteristics.

In Table 6, we seek to assess some of these sources 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 Table OA.9, we compute measures of net yields by approximating 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 postal code-quarter fixed effects.

We then consider imperfect matching due to omitted property-level dimensions, such as layout, exposure, views, decoration, cultural heritage status, or proximity to amenities. These characteristics are unobservable to us but may be relevant for pricing, especially in the presence of search and matching frictions in the housing market (Genesove and Han (2012)). To assess their importance, column 2 of Table 6 estimates the specification with postal code-quarter fixed effects using the actual, rather than synthetic, rent-to-price ratio available for a subset of flats, probably non-owner-occupied flats for sale. Since no rental-sale matching is required in this case, this exercise avoids matching errors due to omitted property-level dimensions. The adjusted R^2 increases to 71%. However, this higher explanatory power appears to reflect sample selection rather than the absence of matching error: when we use the synthetic rent-to-price ratio for the same subsample in column 3, the adjusted R^2 remains very similar, at 70%.

Yet, sample selection may itself be associated with lower matching errors. We therefore further scrutinize the role of matching quality. We next examine whether residual differences between matched properties along observed characteristics can account for the unexplained dispersion in yields. In column 4, we augment the specification with fixed effects for each percentile of a matching quality score, defined as an equally weighted average of re-scaled distances between the rental flat and the matched sale flat(s) for covariates not used in the matching procedure; when a trait is missing for one or both properties, we assign the maximum distance. As an additional check, we also address the possibility that repeated listings inflate measured heterogeneity. In column 5, we exclude properties that, despite having distinct identifiers in the RWI-GEO-RED database, display the same pricing and characteristics within a given postal code-quarter. These "likely duplicates" may reflect properties listed more than once on the platform, or by multiple realtors. In neither specification does the adjusted R^2 exceed the baseline value of 59%.

Next, we turn to imperfect matching due to residual differences between matched properties along observed characteristics. To address this concern, we impose additional restrictions in the matching algorithm that further improve comparability across matched properties, beyond controlling for covariate distances. In column 6, we retain only matches with an absolute difference in living surface of at most 5 sqm, compared with the 10 sqm threshold used in the baseline analysis. In column 7, we apply the same restriction to usable surface.[19] In column 8, we require matched properties to have construction years no more than 10 years apart. In column 9, we require them to have the same quality of facilities, based on a four-category variable ranging from "simple" to "deluxe." The stricter restriction on living surface excludes more than 450,000 matched properties and leaves the adjusted R^2 unchanged at 59%. The adjusted R^2 rises to 69% when focusing on matched properties with negligible differences in usable surface, but this comes at the cost of losing almost 1.5M matched properties. Filtering on year of construction and quality of facilities increases the adjusted R^2 to 64% and 65%, respectively; in both cases, however, the sample shrinks to around 300,000 observations. This large reduction in sample size is mainly due to missing information on these variables for either the rental flat or the matched sale flat, rather than necessarily reflecting genuine matching errors in our algorithm. Hence, the increase in explanatory power may simply reflect stronger sample selection.

[Insert Table 6 about here]

Table 7 further explores the trade-off between sample size and match quality by implementing four alternative matching schemes. Schemes (a) and (b) relax some features

[19] According to ImmobilienScout24, usable surface is typically larger than living surface, as it also includes ancillary usable areas, such as balconies, cellars, storage rooms, or attics, while excluding circulation areas such as hallways and staircases.

of the baseline algorithm and substantially expand the sample of matches, while schemes (c) and (d) impose additional restrictions to improve comparability along flat quality and location, respectively. Scheme (a) implements a 1:1 nearest-neighbor a PSM algorithm with replacement and a caliper of 0.1, while imposing exact matching on postal code and calendar quarter. Propensity scores are estimated from a logit model for the probability that a listed property is for rent, using a comprehensive set of flat characteristics.[20] PSM increases the number of matched flats from around 1.6M in the baseline sample to more than 3.6M, thereby alleviating concerns that our results are driven by selection induced by the baseline matching procedure. In column 1 of Table 7, we estimate our preferred saturated specification for the log-transformed rent-to-price ratio—obtained via PSM—with postal code-quarter fixed effects. The adjusted R^2 is 57%, close to the baseline value of 59% in column 6 of Table 5. This similarity confirms again that the unexplained heterogeneity documented in the baseline analysis is not an artifact of the exact matching scheme. It also mitigates concerns that the modest bias in synthetic yields induced by selection of lower-quality properties into the rental market, discussed above with respect to Figure OA.3, drives our main finding.

Matching scheme (b) follows the baseline algorithm but allows rental and sale properties to be matched across different listing periods. To make prices comparable over time, we adjust sale prices using district-level house price inflation before computing synthetic rent-to-price ratios. Relaxing exact matching on the listing period expands the set of feasible matches and improves matching on other covariates, increasing the matched sample to more than 4M flats. However, the larger sample appears to come at the cost of lower explanatory power—the adjusted R^2 falls to 45%—possibly due to lower match quality.

In scheme (c), we extend the baseline algorithm by also imposing exact matching on flat condition, which reduces the sample to fewer than 0.7M matched properties. In column

[20] Table OA.10 shows that rental and sale listings differ in observables before matching, but that PSM substantially improves balance across the two samples. Figure OA.6 shows distinct propensity-score distributions, yet with substantial common support across rental and sale flats.

3, we re-estimate our usual specification obtaining an adjusted R^2 of 56%, lower than in the baseline sample, which suggests that mismatches in flat quality are not a major driver of the dispersion in housing yields.

Scheme (d), instead, refines the baseline approach by imposing exact matching on highly granular geographic areas, defined by all possible combinations of 1-km raster cells, five-digit postal codes, municipalities, and labor market regions. The resulting dataset includes around 0.6M matched properties. This approach minimizes the spatial distance between each rental flat and its matched sale flat, thereby improving comparability in terms of exposure to local amenities (e.g., restaurants) and agglomeration economies, such as the pooling of highly skilled human capital. Despite this more accurate matching along spatial dimensions, the adjusted R^2 in column 4 remains at 61%.

[Insert Table 7 about here]

Overall, the alternative matching schemes point to a clear trade-off. Less restrictive algorithms, such as schemes (a) and (b), substantially expand the sample but appear to do so at the cost of lower match quality, as reflected in unreported weaker correlations with both the baseline and actual rent-to-price ratios. By contrast, the more restrictive schemes (c) and (d) yield rent-to-price ratios that remain closely aligned with the baseline. Taken together, these findings suggest that measurement issues are unlikely to be a major source of unexplained variation in rent-to-price ratios under our baseline matching scheme. Put differently, the observed dispersion in housing yields does not appear to be an artifact of our methodology.

B. Local Amenities and Agglomeration Economies

Two complementary explanations for high residual heterogeneity in yields relate to amenities and agglomeration economies operating below the postal 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 (2023), Couture et al. (2024), 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 (d). In column 5 of Table 7, we saturate the rent-to-price ratio regression with geographic area-quarter fixed effects. Relative to the specification with postal code-quarter in Table 5, the adjusted R^2 increases by 12% 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 postal 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.

C. Housing Market Frictions

Several local housing-market frictions could contribute to unexplained heterogeneity in housing valuations. One such friction is limited information diffusion. Unlike stock markets, housing markets are decentralized, illiquid, and lack an infrastructure comparable to that of equity analysts. As a result, investors, especially private individuals, may search only locally and trade infrequently, limiting price discovery. Market segmentation with limited search (Piazzesi, Schneider, and Stroebel (2020)) and property illiquidity (Head and Lloyd-Ellis (2012)) can therefore prevent localized supply and demand shocks from being fully arbitraged away, generating cross-sectional dispersion in valuation ratios. A second set of frictions concerns rigidities in local housing supply, due for instance to zoning laws or geographical constraints. When supply is inelastic and demand is strong, even small differences in property quality may translate into large valuation wedges. Finally, local regulation may further amplify yield heterogeneity. Housing-market regulation can take several forms, including zoning, taxation, and rent controls. To the extent that these rules affect properties differentially even within the same postal code, they may contribute to the unexplained dispersion in rent-to-price ratios.

In Table OA.11, we assess the role of housing market frictions by estimating our saturated specification separately across districts with high versus low liquidity, housing supply, construction activity, and property taxation. The explanatory power of the specification remains close to the baseline across all subsamples, suggesting that these frictions do not account for a quantitatively large share of the residual dispersion in rent-to-price ratios. However, these tests should be interpreted with caution, as our proxies for local frictions are relatively coarse and may fail to capture more granular variation within districts.

Rent regulation is a relevant form of housing-market regulation in the German context. Traditionally, however, German rent controls have mainly restricted rent increases in ongoing contracts, rather than rents on newly listed units such as those in our data. This is also the case for the 2013 *capping limit*, which tightened limits on rent increases for existing contracts in designated municipalities. The 2015 *rental brake* represents instead a more relevant case for our setting, as it extended rent regulation to new contracts in tight markets, while exempting new and substantially modernized properties. In Table OA.12, we therefore assess whether the rental brake affected the dispersion of rent-to-price ratios using a difference-in-differences design based on its staggered implementation across German municipalities, while controlling for the capping limit. We find no statistically significant increase in yield dispersion, suggesting that rent controls are unlikely to be a major driver of the heterogeneity we document.[21]

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 postal code-level unexplained dispersion of rent-to-price ratios. By contrast, within-postal 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, rent-to-price ratios may exhibit substantial variation within narrow geographic areas because of heterogeneous risk premia.[22] 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),

[21] We also examine whether dispersion in rent-to-price ratios is related to local climate risk and energy efficiency. Climate-risk proxies in Table OA.13 display limited explanatory power, while dispersion in energy consumption in Table OA.14 is positively associated with yield dispersion but accounts for only a small share of the overall heterogeneity.

[22] According to the static Gordon growth model, the expected rent-to-price ratio of flat f can be decomposed as $\mathbb{E}_t(H_{f,t+1})/P_{f,t} = r_t^f + \mathbb{E}_t(r_f^e) - \mathbb{E}_t(g_{H,f})$, where $\mathbb{E}_t(H_{f,t+1})$ is expected next-period rent, $P_{f,t}$ is the current price, r^f is the risk-free rate, $\mathbb{E}_t(r_f^e)$ is the risk premium, and $\mathbb{E}_t(g_{H,f})$ is expected rent growth.

Eiling et al. (2020)). The cross-sectional nature of our data limits our ability to distinguish between idiosyncratic and systematic risk as determinants of such risk premia (and housing yields dispersion). The pseudo-panel constructed in the next section helps overcome this problem by introducing a time-series dimension: by testing whether rent-to-price ratios predict future excess returns and/or rent growth in the presence of high or low yield dispersion, we can assess whether such dispersion is more closely related to risk premia or to expected rent growth.

Moreover, idiosyncratic variation in yields relates 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 (2022)). In the presence of market frictions, unexplained local dispersion in yields may reflect heterogeneity in investors' beliefs about both 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; more generally, investor disagreement about future rent growth can generate idiosyncratic dispersion in yields.[23]

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. This speaks to the work of Ludwig, Mankart, Quintana, and Wiederholt (2024), who show that heterogeneity in household expectations about future house price growth has sizable implications for the level and dynamics of equilibrium house prices, highlighting the importance of belief dispersion for housing valuations. At the same time, spatial effects related to amenities and agglomeration economies may be at work even below the neighborhood level.

[23] As an indirect test of this channel, we examine whether institutional ownership is associated with yield dispersion, using municipality-level measures of the share of dwellings owned by property companies and other institutional investors. We find little evidence that ownership structure explains the dispersion in housing yields (Table OA.15). However, this test is limited by the granularity of the available ownership data.

V. Risk Premia vs. Expected Rent Growth

Because flat listings in our database are available as monthly vintages, akin to repeated cross-sections, we construct a pseudo-panel by aggregating cohorts of similar properties to introduce a time-series dimension for prices and rents of comparable properties. We use this pseudo-panel to study whether rent-to-price ratios predict future excess returns or rent growth, also shedding light on whether local yield dispersion is more closely related to risk premia or to expected rent growth.[24]

The rent-to-price ratio, analogous to the dividend-to-price ratio for stocks (Campbell and Shiller (1988)), serves as a gauge of market participants' expectations about future housing returns and rent growth (Plazzi et al. (2010)). In particular, we estimate predictive regressions at the cohort-level

$$
\begin{aligned}
r^e_{c,t+1\to t+k} &= \beta_k \ln\left(\frac{H^q_{c,t}}{P_{c,t}}\right) + \tau_c + \nu_{c,t+1\to t+k}, \\
\Delta h_{c,t+1\to t+k} &= \lambda_k \ln\left(\frac{H^q_{c,t}}{P_{c,t}}\right) + \tau_c + \varsigma_{c,t+1\to t+k},
\end{aligned} \tag{3}
$$

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). We account for heteroskedasticity and correlation in error terms $\nu_{c,t+1\to t+k}$and $\varsigma_{c,t+1\to t+k}$ by adjusting standard errors following Driscoll and Kraay (1998).[25]

In Table 8, we estimate the predictive specifications in (3). 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

[24] Table OA.16 reports pseudo-panel summary statistics, confirming that variation in yields and returns is mostly local and that rents vary less over time than sale prices. Figure OA.7 shows that housing performed strongly over the sample period, supporting the use of the pseudo-panel to study return and rent-growth dynamics.

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

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 returns and rent growth are consistent with the traditional present-value relationship: a higher rent-to-price ratio reflects higher expected discount rates and/or lower expected rent growth.[26]

[Insert Table 8 about here]

Table 9 shows that the predictive content of rent-to-price ratios depends on the degree of cross-sectional dispersion in housing valuations. In cohorts with higher dispersion, the rent-to-price ratio more strongly forecasts future returns, while in lower-dispersion cohorts its predictive power is relatively stronger for rent growth. This result is robust to measuring dispersion using either raw yields (Panel A) or filtered yields (Panel B). This pattern is consistent with the view that cross-sectional dispersion in rent-to-price ratios largely reflects heterogeneity in risk premia, rather than differences in expected rent growth, thereby shedding light on the heterogeneity puzzle illustrated in the main analysis.

[Insert Table 9 about here]

VI. Further Analyses

In this section, we study the behavior of cross-sectional dispersion in housing yields at different levels of geographic aggregation. Moreover, we quantify the relevance of unexplained heterogeneity in housing yields by benchmarking it against the distribution of wealth of German households.

[26] Table OA.17 augments the predictive regressions with calendar-quarter fixed effects, absorbing aggregate fluctuations in the spirit of Fama and French (2025). While R^2 increases—especially for returns—the rent-to-price ratio retains predictive power, pointing to the role of local factors driving yield variation. Table OA.18 shows that the predictive power of the rent-to-price ratio is robust to controlling for local factors, although the within R^2 rises substantially, indicating that local conditions contain additional predictive information.

A. The Role of Geographic Aggregation

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

Table 10 contains information about the distribution of these cross-sectional standard deviations. Mean dispersion of raw ratios increases by 31.25% ($= 2.125/1.619 - 1$) going from postal code- to federal state-level data. After factoring out observables and postal 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). This evidence corroborates the largely local nature of cross-sectional variation in rent-to-price ratios.

[Insert Table 10 about here]

We also provide additional evidence on the time-series and spatial structure of local yield dispersion. Table OA.19 documents that dispersion in filtered rent-to-price ratios is only weakly persistent over time at the postal code-quarter level, while Figure OA.8 shows that dispersion is positively correlated across nearby postal codes but that this correlation declines rapidly with distance. These results suggest that unexplained yield dispersion is not a fixed local characteristic, but that it nevertheless exhibits a localized spatial component.

B. 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 81.06. The 10th and 90th percentiles of the rent-to-price ratio (non log-transformed) are 3.16% and 8.29%, respectively.[27] These ratios correspond to housing valuations of EUR 169,303 (= $(81.06 \times 66)/0.0316$) and EUR 64,535 (= $(80.88 \times 66)/0.0829$) for the median flat, implying a variation in the housing wealth of its owners of EUR 104,768 (= EUR $169,303$ – EUR $64,535$).

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 postal code-quarter fixed effects—, we obtain residuals of the log-transformed percentage rent-to-price ratio: $\hat{\epsilon}_{d,t}$ and $\hat{\epsilon}_{p,t}$ respectively. We then retrieve the filtered rent-to-price ratio as $\exp\left[\mathrm{P50}_{\ln(H_{f,t}/P_{f,t})} + \hat{\epsilon}_{d,t}\right]$, where $\mathrm{P50}_{\ln(H_{f,t}/P_{f,t})}$ is the unconditional median log-transformed percentage rent-to-price ratio. We proceed analogously for $\hat{\epsilon}_{p,t}$. 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 postal 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 EUR 88,531 and EUR 71,837 if we absorb time-varying district- or postal code-level factors, respectively. Magnitudes shrink—as we would expect—but remain sizable.

[27] 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 3.34% and 8.16%.

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 spans a substantial fraction of dispersion in household wealth. The forces at play at geographically granular level may thus 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.

VII. Conclusion

Relying on sale and rental prices for flats from a major German online real estate platform, we study the distribution and drivers of rent-to-price ratios. To this end, we compute property-level rent-to-price ratios by matching rental flats to comparable flats for sale. After accounting for a wide set of property-specific characteristics, much of the variation in rent-to-price ratios remains unexplained, even when we absorb time-varying postal code-level factors with fixed effects or if we account for measurement error in a variety of ways. This residual heterogeneity appears to be related to disparities in amenities across neighborhoods and to property-specific differences in risk premia and rent growth expectations, potentially driven by heterogeneity in investors' beliefs and preferences. A predictability analysis suggests that risk premia play a more important role than expected rent growth in accounting for the dispersion of housing yields.

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, limited arbitrage due to transaction costs (e.g., informational and regulatory frictions), or to other economic mechanisms.

References

Amaral, F.; M. Dohmen; S. Kohl; and M. Schularick. "Superstar Returns? Spatial Heterogeneity in Returns to Housing." *Journal of Finance*, 80 (2025), 3057–3094.

an de Meulen, P.; M. Micheli; and S. Schaffner. "Documentation of German Real Estate Market Data." Working Paper (2014).

Bach, S., and S. Eichfelder. "Real Estate Taxation Reform: Tax Land Values, Abolish Privileges." *DIW Weekly Report*, 11 (2021), 199–205.

Beraja, M.; A. Fuster; E. Hurst; and J. Vavra. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *Quarterly Journal of Economics*, 134 (2019), 109–183.

Berger, D.; V. Guerrieri; G. Lorenzoni; and J. Vavra. "House Prices and Consumer Spending." *Review of Economic Studies*, 85 (2018), 1502–1542.

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

Campbell, J. Y., and R. J. Shiller. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies*, 1 (1988), 195–228.

Chambers, D.; C. Spaenjers; and E. Steiner. "The Rate of Return on Real Estate: Long-Run Micro-Level Evidence." *Review of Financial Studies*, 34 (2021), 3572–3607.

Correia, S. "REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing Any Number of High-Dimensional Fixed Effects." Boston College (2018).

Couture, V.; C. Gaubert; J. Handbury; and E. Hurst. "Income Growth and the Distributional Effects of Urban Spatial Sorting." *Review of Economic Studies*, 91 (2024), 858–898.

Demers, A., and A. L. Eisfeldt. "Total Returns to Single-Family Rentals." *Real Estate Economics*, 50 (2022), 7–32.

Deutsche Bundesbank. "Household Wealth and Finances in Germany: Results of the 2017 Survey." *Monthly Report March*, 57 (2019).

Diamond, R., and E. Moretti. "Where Is Standard of Living the Highest? Local Prices and the Geography of Consumption." Working Paper (2021).

Driscoll, J. C., and A. C. Kraay. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics*, 80 (1998), 549–560.

Dustmann, C.; B. Fitzenberger; and M. Zimmermann. "Housing Expenditure and Income Inequality." *Economic Journal*, 132 (2022), 1709–1736.

Eichholtz, P.; M. Korevaar; T. Lindenthal; and R. Tallec. "The Total Return and Risk to Residential Real Estate." *Review of Financial Studies*, 34 (2021), 3608–3646.

Eiling, E.; E. Giambona; R. Lopez Aliouchkin; and P. Tuijp. "Homeowners' Risk Premia: Evidence from Zip Code Housing Returns." Working Paper (2020).

Fama, E. F., and K. R. French. "House Prices and Rents." *Review of Financial Studies*, 38 (2025), 547–563.

Feng, X.; N. Jaimovich; K. Rao; S. J. Terry; and N. Vincent. "Location, Location, Location: Manufacturing and House Price Growth." *Economic Journal*, 133 (2023), 2055–2067.

Ferreira, F. V., and M. Wong. "Estimating Preferences for Neighborhood Amenities Under Imperfect Information." Working Paper (2023).

Fischer, M.; R. Füss; D. Ruf; and S. Stehle. "Betting on Homes." Working Paper (2022).

Flavin, M., and T. Yamashita. "Owner-Occupied Housing and the Composition of the Household Portfolio." *American Economic Review*, 92 (2002), 345–362.

Genesove, D., and L. Han. "Search and Matching in the Housing Market." *Journal of Urban Economics*, 72 (2012), 31–45.

Giacoletti, M. "Idiosyncratic Risk in Housing Markets." *Review of Financial Studies*, 34 (2021), 3695–3741.

Guerrieri, C., and C. Mendicino. “Wealth Effects in the Euro Area.” Working Paper (2018).

Guren, A. M.; A. McKay; E. Nakamura; and J. Steinsson. “Housing Wealth Effects: the Long View.” *Review of Economic Studies*, 88 (2021), 669–707.

Gurun, U. G.; J. Wu; S. C. Xiao; and S. W. Xiao. “Do Wall Street Landlords Undermine Renters’ Welfare?” *Review of Financial Studies*, 36 (2023), 70–121.

Halket, J. “The Cross-Section of Housing Returns.” Working Paper (2025).

Han, L. “Understanding the Puzzling Risk-Return Relationship for Housing.” *Review of Financial Studies*, 26 (2013), 877–928.

Handbury, J., and D. E. Weinstein. “Goods Prices and Availability in Cities.” *Review of Economic Studies*, 82 (2015), 258–296.

Head, A., and H. Lloyd-Ellis. “housing Liquidity, Mobility, and the Labour Market.” *Review of Economic Studies*, 79 (2012), 1559–1589.

Kaas, L.; G. Kocharkov; E. Preugschat; and N. Siassi. “Low Homeownership in Germany—A Quantitative Exploration.” *Journal of the European Economic Association*, 19 (2021), 128–164.

Kantak, P. “Consumption Commitments and Housing Dynamics.” *Management Science*, 71 (2025), 8670–8697.

Kohl, S. “Urban History Matters: Explaining the German–American Homeownership Gap.” *Housing Studies*, 31 (2016), 694–713.

Kohl, S., and J. Sørvoll. “Varieties of Social Democracy and Cooperativism: Explaining the Historical Divergence Between Housing Regimes in Nordic and German-Speaking Countries.” *Social Science History*, 45 (2021), 561–587.

Kosfeld, R., and A. Werner. “Deutsche Arbeitsmarktregionen–Neuabgrenzung Nach Den Kreisgebietsreformen 2007–2011.” *Raumforschung und Raumordnung—Spatial Research and Planning*, 70 (2012), 49–64.

Lambie-Hanson, L.; W. Li; and M. Slonkosky. “Real Estate Investors and the US Housing Recovery.” *Real Estate Economics*, 50 (2022), 1425–1461.

Leonardi, M., and E. Moretti. “The Agglomeration of Urban Amenities: Evidence from Milan Restaurants.” *American Economic Review: Insights*, 5 (2023), 141–157.

Ludwig, A.; J. Mankart; J. Quintana; and M. Wiederholt. “Heterogeneity in Expectations and House Price Dynamics.” Working Paper (2024).

Mian, A.; K. Rao; and A. Sufi. “Household Balance Sheets, Consumption, and the Economic Slump.” *Quarterly Journal of Economics*, 128 (2013), 1687–1726.

Piazzesi, M., and M. Schneider. “Housing and Macroeconomics.” *Handbook of Macroeconomics*, 2 (2016), 1547–1640.

Piazzesi, M.; M. Schneider; and J. Stroebel. “Segmented Housing Search.” *American Economic Review*, 110 (2020), 720–59.

Piazzesi, M.; M. Schneider; and S. Tuzel. “Housing, Consumption and Asset Pricing.” *Journal of Financial Economics*, 83 (2007), 531–569.

Plazzi, A.; W. Torous; and R. Valkanov. “Expected Returns and Expected Growth in Rents of Commercial Real Estate.” *Review of Financial Studies*, 23 (2010), 3469–3519.

Rosenthal, S. S., and W. C. Strange. “How Close Is Close? the Spatial Reach of Agglomeration Economies.” *Journal of Economic Perspectives*, 34 (2020), 27–49.

Sagi, J. S. “Asset-Level Risk and Return in Real Estate Investments.” *Review of Financial Studies*, 34 (2021), 3647–3694.

Savills. “Eigentümerstruktur Am Wohnungsmarkt.” SPOTLIGHT Savills Research (2019).

Sinai, T., and N. S. Souleles. “Owner-Occupied Housing as a Hedge Against Rent Risk.” *Quarterly Journal of Economics*, 120 (2005), 763–789.

Voigtländer, M. “Why Is the German Homeownership Rate So Low?” *Housing Studies*, 24 (2009), 355–372.

Vuolteenaho, T. “What Drives Firm-Level Stock Returns?” *Journal of Finance*, 57 (2002), 233–264.

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 (Graph A to C) against the evolution of macroeconomic conditions in Germany (Graph D) between 2007 and 2017. In Graph A, the conditioning variable is the flat's surface category, going from small (S) to medium (M), large (L), and very large (VL). In Graph 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 Graph C, the conditioning variable is the year. Graph 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.

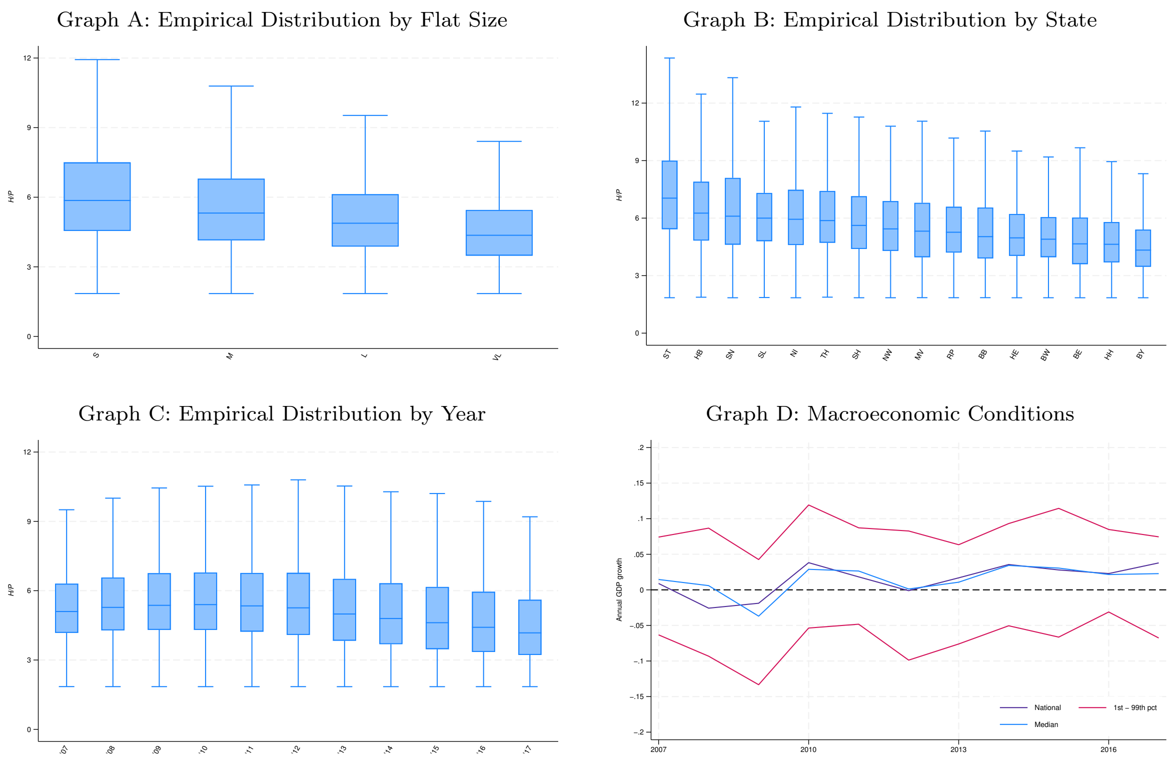

Figure 2. Geographical Dispersion in Median Rent-to-Price Ratios

This figure visualizes the median rent-to-price ratio (obtained via the baseline matching) across Germany over the period 2007–2017. Graph A reports district-level estimates nationwide, while Graphs B–H report five-digit postal code–level estimates for German global cities. Districts and postal codes with fewer than 10 observations are excluded (white colored). Global cities are those assigned a rating between “Alpha” and “Gamma” in the 2020 ranking by the Globalization and World Cities Research Network.

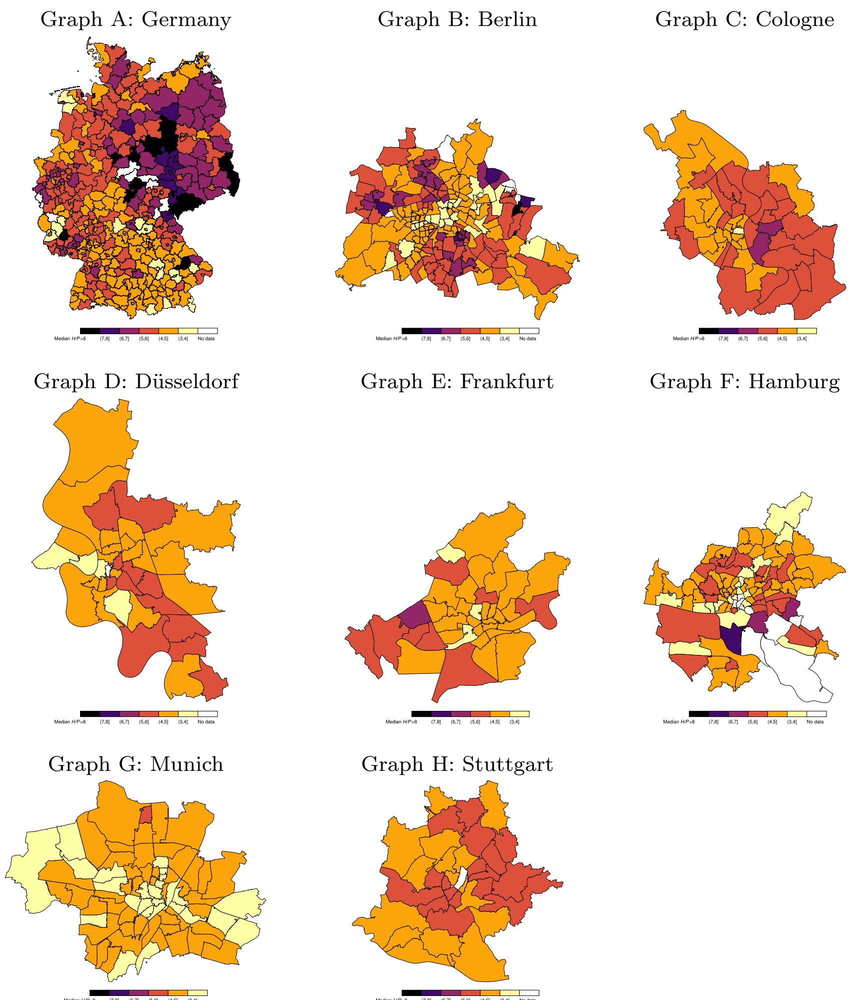

Figure 3. Distribution of Residuals from Rent-to-Price Ratio Regressions
This figure visualizes the distribution of residuals from rent-to-price ratio regressions under different fixed effects schemes by means of violin plots. The violin plot of each variable features a superimposed box plot. The residuals are first obtained from estimating equation (2) without calendar quarter fixed effects. The other specifications augment equation (2) 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) postal 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.

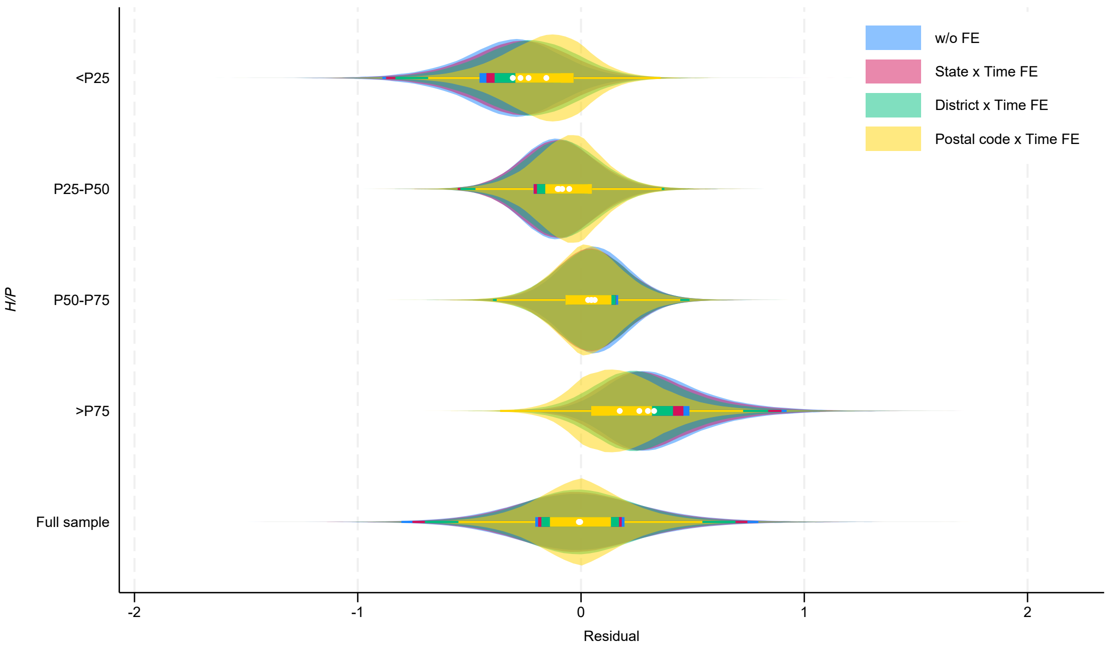

Table 1. Summary Statistics on Flats

This table shows summary statistics on property characteristics distinguishing between flats for rent and matched flats for sale between 2007 and 2017. The t-statistics for mean-comparison tests are based on standard errors clustered by district. Refer to Table A.1 for variable definitions.

	For rent			For sale (matched)			Mean-comparison test	
	Obs.	Mean	SD	Obs.	Mean	SD	Diff.	t-stat
Rental listing								
Annual rent (EUR)	1,625,131	5,916.965	2,678.813					
Annual inclusive rent (EUR)	1,538,530	7,376.999	2,787.831					
Expenses (EUR)	1,539,302	124.956	48.370					
Heating expenses (EUR)	338,398	66.399	24.522					
Heating included	1,431,683	0.694	0.461					
Deposit (EUR)	1,435,151	1,260.245	632.434					
Sale listing								
Sale price (EUR)				1,625,131	127,907.526	86,433.132		
Community charge (EUR)				1,039,854	185.498	61.848		
Rented				1,020,789	0.350	0.473		
Annual rent income (EUR)				343,019	5,391.668	3,471.617		
Holiday property				1,130,045	0.065	0.245		
Matching covariates								
Surface (sqm)	1,625,131	68.498	18.857	1,625,131	68.505	18.684	0.007	0.262
No. rooms	1,625,131	2.451	0.677	1,625,131	2.451	0.677	0.000	.
No. bedrooms	1,625,131	1.449	0.582	1,625,131	1.449	0.582	0.000	.
No. bathrooms	1,625,131	1.024	0.154	1,625,131	1.024	0.154	0.000	.
Floor no.	1,625,131	1.944	1.165	1,625,131	1.968	1.216	0.023	3.730

	For rent			For sale (matched)			Mean-comparison test	
	Obs.	Mean	SD	Obs.	Mean	SD	Diff.	t-stat
Other characteristics								
No. building floors	1,265,269	3.472	1.303	1,291,148	3.717	1.558	0.245	10.553
Usable surface (sqm)	592,967	73.090	20.650	680,945	75.848	21.014	2.759	11.018
Construction year	1,008,532	1,967.650	39.146	1,434,997	1,971.641	36.198	3.991	7.217
Year last modernization	428,778	2,009.473	5.419	464,245	2,006.766	6.578	-2.707	-9.158
Property conditions	1,369,343	2.675	0.688	1,409,938	2.617	0.790	-0.059	-5.375
Quality of facilities	884,447	2.471	0.597	903,791	2.479	0.607	0.007	1.355
Energy consumption (kWh/(sqm$\times$ y))	355,945	123.069	48.551	490,224	120.510	46.635	-2.559	-4.063
Energy rating	26,915	4.380	1.986	33,916	3.923	1.908	-0.457	-6.006
Warm water in energy exp.	876,453	0.151	0.358	888,588	0.227	0.415	0.076	15.270
Protected building	87,558	0.008	0.090	1,125,766	0.051	0.219	0.043	2.573
Balcony	1,419,668	0.794	0.404	1,501,780	0.873	0.330	0.079	22.408
Available parking	160,655	0.935	0.247	199,917	0.957	0.200	0.022	3.055
Accessible with wheelchair	16,320	0.792	0.406	24,623	0.915	0.278	0.123	4.961
Guest WC	1,116,741	0.183	0.387	1,179,567	0.214	0.407	0.031	9.223
Garden	1,017,953	0.295	0.456	1,014,797	0.336	0.469	0.041	2.988
Cellar	1,283,369	0.805	0.396	1,329,583	0.831	0.370	0.025	2.867
Kitchen	1,210,702	0.563	0.496	1,216,816	0.565	0.492	0.002	0.170
Elevator	1,096,689	0.313	0.464	1,153,651	0.448	0.495	0.135	19.212
Assisted living	518,881	0.056	0.229	492,693	0.125	0.328	0.069	19.263
Immediate availability	1,529,034	0.317	0.465	1,141,619	0.345	0.469	0.029	1.807
No. of days online	1,533,636	28.148	33.793	1,588,191	60.596	71.220	32.448	33.234
No. of hits	1,600,338	912.609	1,021.451	1,597,762	828.985	979.988	-83.623	-2.232
No. of clicks (contact button)	1,006,522	17.549	26.062	996,836	8.972	11.633	-8.577	-11.013
No. of clicks (customer profile)	295,469	7.005	7.747	299,107	3.989	4.576	-3.016	-14.432
No. of clicks (share button)	632,091	2.195	1.769	601,810	1.777	1.542	-0.418	-9.157
No. of clicks (customer URL)	797,971	6.575	7.190	925,134	6.127	7.337	-0.448	-1.881

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 postal 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, the coefficient of variation, and the interquartile range of each of these time series is computed. Finally, based on these estimates, the average of these dispersion measures across geographic areas at a given level of aggregation 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 Table A.1 for variable definitions.

Level of aggregation:	Postal code	District	State	National
Equally weighted				
SD (H/P)	1.031	0.870	0.575	0.370
CV (H/P)	0.184	0.150	0.101	0.068
P75-P25 (H/P)	1.197	1.088	0.872	0.560
Weighted by no. properties				
SD (H/P)	0.971	0.632	0.450	0.370
CV (H/P)	0.177	0.119	0.086	0.068
P75-P25 (H/P)	1.264	0.965	0.749	0.560

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 Table A.1 for variable definitions.

	P5	P25	P50	P75	P95	P75−P25	P95−P5
German housing market							
H/P	2.777	3.979	5.094	6.528	9.723	2.549	6.946
Actual H/P	3.400	4.600	5.657	7.210	11.294	2.610	7.894
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, log usable surface, year of construction, year of the last modernization, property conditions, quality of facilities, protected building status, market segment, log 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, log number of days online, and log number of clicks on different items of the listing), together with the corresponding missing value indicators. All specifications include calendar quarter fixed effects. Estimated coefficients are multiplied by 100 to favor readability. 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 Table A.1 for variable definitions.

	$\ln(H/P)$			
	1	2	3	4
ln(Surface)	38.992**	40.652**	32.283**	8.080
	(2.02)	(2.57)	(2.16)	(0.61)
ln(Surface)^2	-11.224***	-10.227***	-9.052***	-6.185***
	(-4.64)	(-5.64)	(-5.40)	(-4.54)
No. rooms	12.610***	10.997***	10.910***	10.869***
	(9.57)	(9.41)	(9.98)	(14.16)
No. bedrooms	-1.407*	-0.332	-0.129	-0.114
	(-1.74)	(-0.48)	(-0.21)	(-0.24)
No. bathrooms	-0.369	-0.176	-0.252	0.065
	(-0.25)	(-0.12)	(-0.19)	(0.06)
Floor no.	0.536	0.725	0.666	0.698*
	(0.94)	(1.41)	(1.35)	(1.84)
ln(Expenses)		-1.287	-1.368	-3.589***
		(-1.15)	(-1.26)	(-4.02)
ln(Heating expenses)		-4.312***	-4.568***	-3.047***
		(-7.25)	(-7.61)	(-5.99)
Heating included		3.572**	3.231**	1.677*
		(2.59)	(2.43)	(1.82)
ln(Deposit)		-10.133***	-9.684***	-11.156***
		(-3.18)	(-3.04)	(-4.25)
ln(Community charge)			2.733**	0.549
			(2.30)	(0.58)
Holiday property			-6.350***	-4.821***
			(-6.35)	(-5.52)
Rented			11.561***	7.894***
			(8.09)	(6.81)
Missing indicators	No	Yes	Yes	Yes
Covariate distances	No	No	No	Yes
Time FE	Yes	Yes	Yes	Yes
Unit of observation	Match. flat	Match. flat	Match. flat	Match. flat
Mean(y)	1.63	1.63	1.63	1.63
SD(y)	0.38	0.38	0.38	0.38
Observations	1,625,131	1,625,131	1,625,131	1,625,131
Adjusted R^2	0.09	0.11	0.12	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 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
Postal code FE	No	No	No	No	Yes	No
Postal 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)	1.63	1.63	1.63	1.63	1.63	1.63
SD(y)	0.38	0.38	0.38	0.38	0.38	0.38
Observations	1,625,131	1,625,131	1,625,129	1,624,137	1,624,481	1,606,359
Adjusted R^2	0.35	0.36	0.40	0.44	0.47	0.59
Adjusted within R^2	0.27	0.27	0.28	0.28	0.29	0.29

Table 6. Unexplained Variation of Rent-to-Price Ratios and Measurement Error

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 error in rent-to-price ratios. Each column builds on the specification of column 4 of Table 4 augmented with fixed effect structures (indicated below). 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 9, 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 9, additional restrictions in the matching procedure are introduced. The sample in column 6 (column 7) discards any match for which the absolute difference in terms of (usable) surface between the flat for sale and the matched flat for sale is larger than 5 sqm. The sample in column 8 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 9 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. Standard errors are clustered by district. Refer to Table A.1 for variable definitions.

	ln(Inclusive H/P)	ln(Actual H/P)	ln(H/P)						
	1	2	3	4	5	6	7	8	9
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	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Postal code×Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matching quality FE	No	No	No	Yes	No	No	No	No	No
Likely duplicates	Included	Included	Included	Included	Excluded	Included	Included	Included	Included
Additional matching	–	–	–	–	–	Surface	Usable 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	Match. flat
Mean(y)	1.88	1.77	1.74	1.63	1.63	1.63	1.63	1.64	1.61
SD(y)	0.39	0.39	0.36	0.38	0.37	0.37	0.34	0.33	0.36
Observations	1,530,719	330,696	330,696	1,606,359	1,356,087	1,138,189	123,077	348,225	283,989
Adjusted R^2	0.52	0.71	0.70	0.59	0.58	0.59	0.69	0.64	0.65
Adjusted within R^2	0.10	0.08	0.24	0.29	0.30	0.29	0.31	0.23	0.24

Table 7. Unexplained Variation of Rent-to-Price Ratios Under Alternative Matching Schemes

This table reports coefficients of determination from regressions for property-level rent-to-price ratios for a German sample between 2007 and 2017, under alternative matching schemes. The dependent variable is the log-transformed rent-to-price ratio, as obtained via four different matching procedures of flats for rent to flats for sale. Matching scheme (a) is based on a 1:1 nearest-neighbor PSM algorithm with replacement and a caliper of 0.1. To obtain the propensity scores, a logit model for a rental indicator is estimated. The covariates include traits available both for flats for rent and for sale (flat surface, number of rooms, number of bedrooms, number of bathrooms, floor, usable surface, year of construction, year of the last modernization, property conditions, quality of facilities, 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. Exact matching on postal code and calendar quarter is imposed. Matching scheme (b) follows the baseline matching but allowing for matching of properties for rent and for sale across periods. All prices are adjusted for district-level house price inflation to ensure comparability across periods. Matching scheme (c) augments the baseline matching procedure by imposing exact matching also on flat conditions. Matching scheme (d) in columns 4 to 9 imposes exact matching at a finer geographic level than the five-digit postal code, namely on the areas resulting from combinations of labor market regions, municipalities, five-digit postal 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. Each column builds on the specification of column 4 of Table 4 augmented with fixed effect structures (indicated below). Standard errors are clustered by district. Refer to Table A.1 for variable definitions.

	$\ln(H/P)$								
	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
Postal code$\times$Time FE	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No
Geo$\times$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
Matching scheme	(a)	(b)	(c)	(d)	(d)	(d)	(d)	(d)	(d)
Mean(y)	1.66	1.65	1.64	1.64	1.64	1.55	1.55	1.70	1.70
SD(y)	0.38	0.36	0.33	0.36	0.37	0.35	0.35	0.36	0.37
Observations	3,631,269	4,006,429	685,663	591,881	550,090	230,341	221,831	361,526	328,259
Adjusted R^2	0.57	0.45	0.56	0.61	0.71	0.60	0.68	0.60	0.72
Adjusted within R^2	0.23	0.30	0.21	0.29	0.29	0.33	0.33	0.27	0.26

Table 8. Predictive Regressions

This table reports estimates from predictive regressions for the housing premium and rent growth on the log-transformed quarterly 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 $r^e_{t+1\to t+k}$ (rent growth $\Delta h_{t+1\to t+k}$), with $k = 1,\ 4,\ 12$. All specifications include cohort fixed effects. Estimated coefficients are multiplied by 100 to favor readability. 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 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 $k=12$
$\ln(H^q/P)$	32.784***	43.865***	59.476***	-4.820***	-10.372***	-16.784***
	(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 9. Predictive Regressions and Local Dispersion in Rent-to-Price Ratios

This table reports estimates from predictive regressions for the housing premium and rent growth on the log-transformed quarterly rent-to-price ratio, accounting for local dispersion of rent-to-price ratios. 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 1-2 (3-4) is the 12-quarter ahead housing premium $r^e_{t+1\to t+12}$ (rent growth $\Delta h_{t+1\to t+12}$). All specifications include cohort fixed effects. Odd (even) columns restrict the estimation sample to cohorts with above-median (below-median) average cross-sectional standard deviation of rent-to-price ratios. In Panel A, such a standard deviation is based on the baseline synthetic annual rent-to-price ratios, in Panel B on filtered ones, namely the residuals from the specification in column 6 of Table 5. Estimated coefficients are multiplied by 100 to favor readability. 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 Table A.1 for variable definitions.

Panel A: Based on Local Dispersion in $\ln(H/P)$

	$r^e_{t+1\to t+12}$		$\Delta h_{t+1\to t+12}$	
	1	2	3	4
$\ln(H^q/P)$	73.426***	44.470***	-10.776***	-24.052***
	(5.97)	(2.90)	(-3.05)	(-5.61)
Cohort FE	Yes	Yes	Yes	Yes
Local SD($\ln(H/P)$)	High	Low	High	Low
Unit of observation	Cohort-time	Cohort-time	Cohort-time	Cohort-time
Mean(y)	0.27	0.29	0.04	0.06
SD(y)	0.20	0.17	0.07	0.08
Observations	7,585	7,281	7,771	7,089
Within R^2	0.21	0.09	0.04	0.15

Panel B: Based on Local Dispersion in Filtered $\ln(H/P)$

	$r^e_{t+1\to t+12}$		$\Delta h_{t+1\to t+12}$	
	1	2	3	4
$\ln(H^q/P)$	67.634***	48.748***	-13.126***	-22.010***
	(5.23)	(3.46)	(-4.00)	(-4.98)
Cohort FE	Yes	Yes	Yes	Yes
Local SD(Filtered $\ln(H/P)$)	High	Low	High	Low
Unit of observation	Cohort-time	Cohort-time	Cohort-time	Cohort-time
Mean(y)	0.27	0.29	0.04	0.05
SD(y)	0.19	0.18	0.07	0.07
Observations	8,237	6,629	8,385	6,475
Within R^2	0.19	0.10	0.06	0.13

Table 10. 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 postal 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 Table A.1 for variable definitions.

	Obs.	Mean	SD	Min	P25	Median	P75	Max
Postal code-level aggregation								
SD(H/P)	12,622	1.619	0.661	0.223	1.149	1.491	1.942	5.496
SD(Filtered H/P)	12,622	1.383	0.597	0.385	0.963	1.236	1.643	5.199
District-level aggregation								
SD(H/P)	6,063	1.878	0.606	0.481	1.456	1.805	2.213	4.982
SD(Filtered H/P)	6,063	1.360	0.470	0.400	1.017	1.296	1.611	5.262
State-level aggregation								
SD(H/P)	664	2.125	0.485	0.816	1.779	2.078	2.405	4.274
SD(Filtered H/P)	664	1.378	0.336	0.651	1.156	1.336	1.580	3.507

Appendix for
"Housing Yields"

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 II.A.
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.

(Continued)

Table A.1. – *Continued*

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.
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[28]		
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.
Δh	RWI-GEO-RED	Quarterly logarithmic rent growth rate for the cohort of properties of the pseudo-panel dataset.
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.
District-level characteristics		
Global city	Globalization and World Cities Research Network	Indicator equal to one if a district has a rating between “Alpha” and “Gamma”.

[28] Details on the construction of the pseudo-panel are provided in the Online Appendix.