Understanding Spatial House Price Dynamics in a Housing Boom*

Leo Kaas[†] Georgi Kocharkov[‡] Nicolas Syrichas[§]

Abstract

We examine the evolution of spatial house price dispersion during Germany's recent housing boom. Using a dataset of sales listings, we find that house price dispersion has significantly increased, which is driven entirely by rising price variation across postal codes. We show that both price divergence across labor market regions and widening spatial price variation within these regions are important factors for this trend. We propose and calibrate a directed search model of the housing market to understand the driving forces of rising spatial price dispersion, highlighting the role of housing supply, housing demand and frictions in the matching process between buyers and sellers. While shifts in housing supply, housing demand and search frictions all matter for overall price increases and for regional price differences, we find that variation in housing demand is the primary factor contributing to the widening spatial dispersion. Our model-based demand and supply components correlate strongly with regional fundamentals, suggesting that they capture economically meaningful variation in local housing markets.

JEL classification: D83; R21; R31

Keywords: House price dispersion; Spatial housing markets; Search frictions in housing markets;Housing boom; Germany

*We thank SAFE (project 21529) for financial support. We are grateful to Murat Alp Celik, Gueorgui Kambourov, Rachel Ngai, Nawid Siassi, Tuuli Vanhapelto, two anonymous referees, as well as audiences at EUI, Goethe University Frankfurt, University of Toronto, TU Wien, FAU/IAB Macro Seminar and the 2024 "Search in Housing Markets" Conference at Imperial College, for comments and discussions. We also thank Lars Brausewetter, Johannes Göhausen and Stephan Thomsen for sharing their data on regional supply and demand fundamentals. The views expressed in this paper represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank, the European Central Bank or the Eurosystem.

†Goethe University Frankfurt, Germany; European University Institute Florence, Italy [ORCID]. Email: leo.kaas@eui.eu

‡**Corresponding author:** Deutsche Bundesbank; European Central Bank Frankfurt, Germany [ORCID]. Email: georgi.kocharkov@bundesbank.de

§Free University of Berlin; Berlin School of Economics, Germany [ORCID]. Email: nicolas.syrichas@fu-berlin.de

1 Introduction

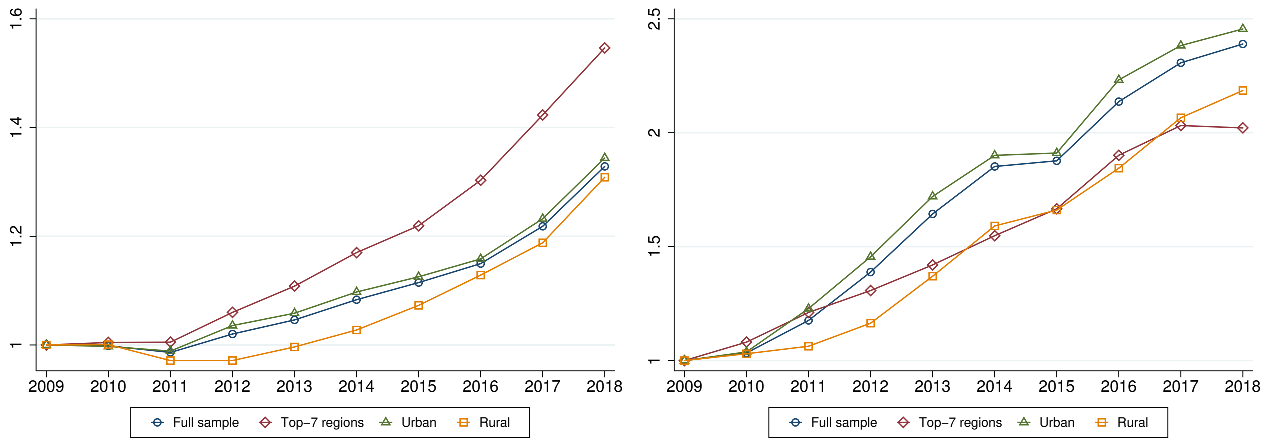
A striking feature of many housing markets is the large and often rising dispersion of house prices and rents across locations. Spatial dispersion of housing costs has several important social and economic consequences, such as widening wealth inequality between households, increasing residential segregation with spillovers on children's human capital (Fogli et al., 2023), or regional misallocation of capital and labor with detrimental effects on economic growth (Herkenhoff et al., 2018; Hsieh and Moretti, 2019). The existing literature on widening spatial price dispersion focuses on differences in house prices across metropolitan areas or municipalities (e.g. Van Nieuwerburgh and Weill, 2010; Gyourko et al., 2013), while house price dispersion at more granular levels remains largely unexplored.

This paper analyzes the trends and determinants of spatial house price dispersion during Germany's recent housing boom over the period 2009–2018. Unlike other industrialized countries, real house prices in Germany did not exhibit an upward trend in the four decades prior to 2010.[1] Since then, however, real house prices increased overall and at varying speeds in different geographic subsamples, as is visible in panel (a) of Figure 1. At the same time, panel (b) illustrates that the spatial dispersion of house prices has widened sharply, even in rural regions where average house prices went up by much less than in urban regions. Also within the relatively homogeneous group of the largest seven metropolitan areas that saw the largest overall increase of house prices, a large increase of spatial price dispersion can be observed.

After documenting the main empirical patterns of Germany's housing boom and the simultaneous rise in spatial price dispersion, we build and calibrate a simple spatial housing search model whose parameters can be identified from the price, contact-per-listing, and duration data at the postal code level. We use the calibrated model to analyze the separate roles of housing supply, housing demand, and matching frictions for the observed price trends.

In Section 2 we describe a dataset of sales listings from Germany's largest housing online platform and document the contribution of location to the observed house price trends since the year 2009. We calculate inflation- and quality-adjusted house prices and find that the cross-sectional variance has increased substantially during 2009–2018. We first document that the entire increase of this variance is accounted for by an increase of dispersion between postal codes which we use as our granular location measure (cf. Figure 1.b), whereas within postal codes there is no change of house price dispersion. Second, we dissect spatial price dispersion into between and within labor market region components. For the full sample, the within-region component accounts for over a third of the between-location variance and

[1] See Kindermann et al. (2024) for the historic house price development on the basis of different datasets. The doubling of nominal house prices during 1975–1995 shown in their paper is almost exactly offset by a doubling of the CPI during this period.



(a) Mean

(b) Variance

Figure 1: Mean and spatial dispersion of house prices

NOTES: House prices are the residuals of hedonic regressions of inflation-adjusted prices in sales listings. Panel (a) shows the mean of these residuals, panel (b) shows the variance across postal codes, where all series are normalized to unity in the year 2009. See Section 2 for further details about the data, calculation of the variables and definition of the geographic subsamples.

is responsible for about a quarter of the rise in spatial dispersion. When focusing on the more homogeneous Top-7 labor market regions, we find that rising dispersion within labor market regions accounts for about 40 percent of the overall increase. We further document an overall tightening of the housing market, as evidenced by a decline in the duration of a listing and a substantial increase in the contact-per-listing-day ratio, which point to a surge of housing demand during the observation period.

In Sections 3 and 4, we propose and calibrate a spatial housing search model that helps to understand the separate roles of demand, supply and matching frictions for rising spatial price dispersion, both between and within labor market regions. The model features homogeneous buyers and sellers whose house price valuations vary across space and over time. We further introduce time-invariant location premia that control the average market shares at the location level. While sellers choose the number of listings and the posted prices, buyers decide in which locations to search and which sellers to contact. In line with standard competitive search theory (cf. Moen, 1997), both sellers and buyers trade off prices and matching probabilities.

Importantly, our housing search model is a highly stylized, not fully structural approach to describe price setting and price variation in spatial housing markets. While abstracting from the underlying reasons for demand or supply changes, a key advantage of our model is that all structural parameters can be uniquely identified on the basis of our house listings data. Thus, the model serves the purpose of quantifying the respective contributions of demand, supply and market frictions for the observed house price dynamics.

In the equilibrium of our model, prices, listing duration and tightness in a local housing market respond to the time and space variation of buyer and seller house valuations and to a search frictions term that reflects how rents are shared between buyers and sellers in the frictional housing market. While the buyer valuation stands for the willingness to pay in certain locations, the seller valuation represents the outside option of a housing unit for sale which may reflect the outside value of renting out an existing unit.[2] These two components capture the contributions of housing demand and supply, respectively. Additionally, differential trends in house prices could reflect differences in search frictions between housing locations. Although ubiquitous in labor economics, this channel is mostly absent in the quantitative housing literature.

The model calibration uses a two-step procedure. First, we estimate matching functions on the basis of duration and contact-per-listing data. These parameters are estimated separately for each labor market region, where matching efficiency is additionally allowed to vary over time. The latter is required by our data which indicate a decrease of matching efficiency in the second half of our observation period in most labor market regions. The second step is to jointly calibrate the time- and location-specific valuations of buyers and sellers as well as the time-invariant location premia, using our data on prices, tightness, the estimates of matching functions, and the market shares. Within larger labor market regions, our model has several thousand parameters that include over 100 postal codes and 40 quarters. Nonetheless, this calibration step can be performed rather efficiently since our model is linear in nearly all parameters that are set at the second step.

In Section 5 we use the calibrated model to quantify the driving forces behind the observed house price dynamics during the period 2009–2018. Through the lens of the model, three factors generate variation in house prices over time: housing supply via the valuation of sellers, housing demand via the valuations of buyers, or search frictions in the housing market. A simple counterfactual exercise is used to quantify the respective contribution of each of these factors for the increasing trend of house prices and their dispersion in the Top-7 labor market regions and for the between- and within-region variation.

We find that the majority of the rise in house prices is driven by stronger housing demand, which accounts for 83 percent of the overall price increase and 84 percent in Top-7 labor market regions. Housing supply plays a more limited role, contributing around 20 percent to overall price growth, while search frictions have only a very small effect. Turning to spatial dispersion, we find that nearly the entire increase in house price variance is explained by

[2] Regulatory constraints and geographic barriers could impose hurdles in some premium locations driving sellers' valuations up (e.g. Saiz, 2010; Hsieh and Moretti, 2019).

demand shifts, accounting for 98 percent of the overall rise and 87 percent in the Top-7 regions. Supply effects contribute modestly, while search frictions play a minor and slightly dampening role. We further elaborate on the role of search frictions, emphasizing that the gradual convergence in search frictions across locations helps explain why search frictions did not amplify spatial price dispersion.

We also use the calibrated model to decompose the between-location variance into within- and between-region components, paralleling our data decomposition of Section 2. Similar to our findings for the Top-7 regions, the majority of within-region dispersion is attributed to demand-side changes. Nonetheless, a sizable share of between-region divergence is accounted for by housing supply, which possibly reflect the expansion of construction activity in relatively less demanded regions during this period. Changes in search frictions have little impact on within-region dispersion and even a dampening effect on the rise of between-region price differences. The latter can be explained by the regional convergence of search frictions over time.

We provide external validation of the model-generated supply and demand components by relating them to observed district-level indicators. We show that buyer valuations are positively correlated with income, credit quality, population density, and proximity to the center of the labor market region, while seller valuations respond to land prices and construction activity. This suggests that the model successfully captures economically meaningful variation.

Further, we study the geographic structure of the model-generated demand and supply components using spatial correlation techniques. Although the model itself does not feature an explicit spatial structure, we find that nearby locations tend to have more similar model-based buyer and seller valuations. This spatial correlation has grown stronger over time, particularly in the Top-7 labor market regions.

1.1 Related Literature

Spatial dispersion. Our work relates to Van Nieuwerburgh and Weill (2010) and Gyourko et al. (2013), who study reasons why house price dispersion across U.S. metropolitan areas increased over time. Van Nieuwerburgh and Weill (2010) use a dynamic spatial equilibrium model in the spirit of Rosen (1979) and Roback (1982) to show that high-ability households move into metropolitan areas with high wages and stringent regulatory housing supply. Likewise, Gyourko et al. (2013) argue that house price differentials in large metropolitan cities can be attributed to inelastic supply combined with an increasing sorting of high-income households. Our article differs from these two studies in two dimensions. First, we consider house price variation at a much more granular level. In particular, we show that house prices exhibit increasing dispersion over time, not only across labor market regions (i.e metropolitan areas) but also at the postal code level within labor market regions. Second, to use information on listing duration and contact-per-listing data, we employ a directed search matching model that accounts for frictions in local housing markets instead of the frictionless island-type model of Van Nieuwerburgh and Weill (2010). Different from other literature in urban economics (e.g. Henderson, 1974; Fujita and Thisse, 2002; Ottaviano et al., 2002; Duranton and Puga, 2023), our model does not feature an explicit geographic structure with a distance measure between locations. Nonetheless, we use the calibrated model to explore spatial correlations of the demand, supply and search frictions factors and the role of the distance to the city center for housing demand.

Our paper is related to recent empirical studies explaining differential house price trends during a housing boom. Kindermann et al. (2024) study regional disparities in house prices across German labor markets in the same ten-year period, focusing on the role of regional differences in expectation formation. Similarly, Brausewetter et al. (2024) examine district-level house price dynamics across Germany and show that regional demand fundamentals such as population density and skill composition explain much of the observed price growth, while overvaluation and speculative behavior are particularly relevant in large cities. Amaral et al. (2024) study the relationship between price and rent divergence across metropolitan areas in 15 advanced economies during a period of low-interest rates. They find that house prices have increased at a much faster pace compared to rents, both in major metropolitan areas but also on the national level. Again the focus of this paper is on house price trends at a more granular spatial level. While we do not consider rents in our main analysis, we document in Appendix D.3 that rent dispersion has also increased over time across postal codes, especially within the Top-7 labor market regions. Finally, Ahlfeldt et al. (2023) make use of the same dataset and also document rising geographic dispersion in prices and rents during Germany's housing boom next to other stylized facts. Their study, however, does not make use of the demand indicators in the data (contacts, listing duration) so that they do not aim at disentangling supply- and demand-driven price developments.

Housing market search. On the modelling side our paper relates to a literature employing directed search models to explain salient features of housing markets (Albrecht et al., 2016; Hedlund, 2016; Rekkas et al., 2022; Moen et al., 2021; Jiang et al., 2024; Garriga and Hedlund, 2020). Closest to our work is Rekkas et al. (2022) who use a directed search model with heterogeneous buyers which they estimate using listings data from the Vancouver area. Similar to us, they find that heterogeneous tastes of buyers explain much of house price dispersion, whereas search frictions matter only little for dispersion (although contribute to the price stickiness observed in their data). Our paper mainly differs in two dimensions. First, we use our model to disentangle the respective contributions of buyer and seller valuations, next to search frictions, for house price dynamics. Second, we seek to explain the factors that account for spatial dispersion between and within labor market regions.

Another closely related paper is Vanhapelto and Magnac (2024) who utilize listings and transactions data from Finland to estimate a model of segmented housing search. In their model, better liquidity in some markets is either due to higher matching efficiency or to differences in popularity among buyers (market tightness). Model-based results show that differences in market tightness contribute more to explaining differences in liquidity across markets than differences in matching technology. Our paper is different because it deals with the evolution of price dispersion across time and space. Moreover, we evaluate both supply and demand changes along with matching efficiency changes for the observed dynamics of house prices.

Our paper further relates to a literature that uses online listings data to study the role of imperfect and costly information frictions to house price variation (Ben-Shahar and Golan, 2022; Jiang et al., 2024; Guren, 2018). Our paper differs from this literature in its focus on the structural factors that explain residual variation across locations, rather than frictions that generate variation in the prices of similar houses within locations.

2 Empirical Patterns

2.1 Data

We use sales listings of residential housing units in Germany that were posted at the online platform *ImmobilienScout24* during January 2009 and December 2018.[3] The raw data are further prepared, geo-referenced and labeled by the RWI Essen within the RWI-GEO-RED dataset which can be accessed for research purposes. Next to the posted prices, the dataset contains a large number of housing characteristics, including geographical location at the square-kilometer level. It further contains information on the duration of a listing in days, the number of views that a listing received and the number of contact attempts of potential buyers.

A limitation of these data is that only listed prices are available, but not the actual transaction prices. However, comparing posted prices aggregated at the city level with the newly created German Real Estate Indices (GREIX) across cities, we find striking similarities of the levels and the evolution of these two series over time.[4] Moreover, earlier studies using both transaction and listing price data show that on average a property sells within 1.6 percent of its listed price (Guren, 2018). Nonetheless, we do observe if the same property

[3] *ImmobilienScout24* is the largest real estate listing website in Germany with a self-reported market share of over 50 percent (Georgi and Barkow, 2010).

[4] See Appendix C for further details.

has been listed multiple times within a short horizon with marginal changes. In those cases, we keep only the last listing.[5] For further details about the data, data cleaning procedures, and the number of listings across time and space, see Appendix A.

2.2 Hedonic Regressions

Since we are interested in spatial variation of house prices over time, rather than changes in the composition of housing units for sale, we control for any observable differences in the characteristics of these housing units. To this end, we estimate a standard hedonic house price regression for our sample of sales listings. We pool all observations and estimate the OLS regression

$$\log p_{ht} = \text{const} + X_{ht}'\beta_X + \varepsilon_{ht} \ , \tag{1}$$

where $\log p_{ht}$ is the (log) inflation-adjusted listed price per m^2 of housing unit h posted at time t, X_{ht} is a vector of housing characteristics of that property which includes a set of categorical variables for the number of rooms, dummy variables for guest toilet and cellar, age of the property in five-year categorical intervals, 22 categories indicating the type of the property, and quarterly dummies to take care of seasonal variation. Appendix B provides further details about the control variables.

We are interested in the residuals ε_{ht} of this hedonic regression which we aggregate at the location level in a quarterly panel. While we include controls for various characteristics of the housing unit, we deliberately do not control for any characteristics of the specific location, such as neighborhood quality, local infrastructure, or distance to the center of the city or labor market region. Therefore, we acknowledge that the price residuals, when averaged at the postal code level, reflect these location premia both in the cross section as well as their variation over time. As a robustness experiment, however, we include distance to the center of the labor market region (geodetic distance or driving times) into the hedonic regressions and show that the main descriptive results of this section remain intact (see Appendix D.1). Furthermore, while our structural model abstracts from explicit consideration of location amenities (including distance to the city center), we rationalize our calibrated demand indicators using a variety of regional demand fundamentals (see Section 5.4).

[5] Another issue is the presence of phishing or fraud listings which usually look like legitimate listings, often at below market prices to attract potential buyers. *ImmobilienScout24* has developed a sophisticated algorithm to detect and remove those listings. It is also a fee-based platform, so that the cost for listing a fake offer is high. To alleviate remaining concerns, we remove ultra-popular offers (i.e. listings with hits or contacts beyond the 99th percentile) in our data cleaning process.

2.3 Baseline Sample

Since we are interested in the spatial distribution of house prices and its changes over time, we construct a quarterly panel which builds on postal codes as our main geographical unit.[6] We restrict the sample to those postal codes that contain at least ten listing observations in all quarters of our ten-year period.

As a larger aggregate geographic unit, we use the labor market regions categorized by Kosfeld and Werner (2012). These regions, which usually combine several municipalities and districts, are characterized according to commuter links to local labor market centers. Since some rural labor market regions are not well represented, we drop all labor market regions with less than 14 postal codes.

Both restrictions mitigate the impact of regions or postal codes which are sparsely populated and contain only few listings. In the following, we refer to postal codes as *locations*, while *regions* denote the labor market regions in our classification. The final balanced panel contains 2,161 locations in 99 regions over 40 quarters. It is important to note that none of our empirical findings is sensitive to these sample restrictions.

2.4 Descriptive Statistics

Table 1 shows descriptive statistics of our baseline sample, reporting the means of selected variables, separately for five two-year periods. The first two rows illustrate the sharp rise in house prices over the ten-year horizon. The average inflation-adjusted house listed for sale in the 2009-2010 period cost around €1,451 per m^2. Ten years later the posted sales price increased by around 36 percent to €1,978 per m^2. Note that this increase cannot be attributed to changes in housing characteristics, as the hedonic house prices ε_t exhibit a similar increase as the raw prices (in log points). When restricted to the largest seven labor market regions, house prices grew by 58 percent (from €1863 to €2951 per m^2), indicating a widening of cross-regional house price dispersion which we elaborate on in the next section.[7]

The bottom four rows of Table 1 indicate a tightening of the German housing market over the same period. The average number of listings in a location per quarter decreased by 35 percent, while the average duration of a listing fell from 56 to 45 days, and the number of contacts (i.e. buyers clicking the contact button) increased by 73 percent. The last row reports the number of contacts per listing day as a flow-based measure of housing market

[6]Relative to the km^2 grid information provided in the RWI-GEO-RED data, postal codes are larger and more homogeneous in population size. Germany has about 40.9m households and 8,200 postal code locations, so that a postal code includes on average about 5,000 households.

[7]Tables 21 and 22 in Appendix D display similar patterns for the rental market. Listed rents per m^2 increased by 18 log points all over Germany and by 23 log points in the Top-7 labor market regions over the same period.

tightness. This number almost quadrupled which indicates a substantial tightening of the German housing market over this ten-year period.[8]

Table 1: Descriptive statistics

Variable	2009-10	2011-12	2013-14	2015-16	2017-18
Log price $\ln p$	7.28	7.29	7.35	7.48	7.59
Price residual ε	-0.13	-0.12	-0.07	0.03	0.17
Listings S	71	69	73	58	46
Duration in days d	56	52	44	48	45
Contacts C	169	209	280	305	292
Flow tightness $\frac{C}{dS}$	0.05	0.07	0.11	0.16	0.19
Observations	17,288	17,288	17,288	17,288	17,288

NOTES: Means of selected variables for the baseline sample of location-quarter observations. Prices are in euros and adjusted for inflation using the CPI of the federal states in Germany.

2.5 House Price Dispersion Across Space and Time

Not only has the average house price gone up during 2009-2018, but there is also a substantial widening of house price dispersion over the same period. To document this phenomenon, we go back to the level of individual listings and consider the residual posted price per m^2, denoted ε_{ht} for listing h at time t, as obtained from the hedonic regression described above. Across listings, the variance of residual prices has increased by over 50 percent, see Table 2.

To understand the spatial dimension of rising dispersion, we first decompose the variance of residual prices into within- and between-location components.[9] Suppressing the time index, the variance of residual prices is split into

$$\underbrace{\operatorname{var}\varepsilon_h}_{\text{total variance}} = \underbrace{\sum_{i\in L} s_i \operatorname{var}_i(\varepsilon_h)}_{\text{within locations}} + \underbrace{\sum_{i\in L} s_i(\bar{\varepsilon}_i - \bar{\varepsilon})^2}_{\text{between locations}} \ , \tag{2}$$

[8]Trends in the absolute number of listings S and contacts C may principally reflect changes in the market share of *ImmobilienScout24* over this ten-year period that could also vary between locations. However, to the extent that this platform is representative of the German housing market, such changes in market shares should not matter for the other two measures, namely listing duration d and flow tightness $C/(dS)$. Our identification strategy in Section 4 uses only these latter two variables. Hence, it builds on the assumption of representativeness of the platform on the buyer and seller side, regardless of potential changes in its market share (between locations or over time).

[9]We also perform the variance decompositions in (2) and (3) with the raw prices instead of the price residuals and obtain rather similar results.

where L is the set of locations (postal codes) with index i, $\bar{\varepsilon}_i = \frac{1}{n_i}\sum_{h=1}^{n_i}\varepsilon_h$ is the average residual price in location i with number of listings n_i, and $s_i = n_i/(\sum_{j\in L} n_j)$ is the listing share of location i. $\bar{\varepsilon}$ is the average residual price across all of Germany. The within-location term on the right-hand side is the listing-weighted average of the variances $\text{var}_i(\varepsilon_h) = \frac{1}{n_i}\sum_{h=1}^{n_i}(\varepsilon_h - \bar{\varepsilon}_i)^2$ over all locations i. The second term is the listing-weighted variance of location-level prices, i.e. the between-location variance. We calculate this additive decomposition separately for each year.

Table 2: Within- and between-location variance decomposition

	Total Variance			**Within Locations**			**Between Locations**		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full Sample	0.191	0.240	0.291	0.114	0.113	0.109	0.077	0.127	0.182
West Germany	0.188	0.238	0.283	0.113	0.111	0.105	0.075	0.126	0.178
East Germany	0.188	0.235	0.298	0.131	0.132	0.161	0.056	0.103	0.137
Top-7 Regions	0.184	0.201	0.230	0.114	0.100	0.090	0.070	0.101	0.140
Urban	0.190	0.230	0.283	0.119	0.108	0.100	0.071	0.122	0.183
Rural	0.181	0.208	0.266	0.110	0.111	0.112	0.071	0.097	0.153

NOTES: "Full Sample" contains the listings in all quarter-location observations in our baseline sample. "West Germany" and "East Germany" include all listings located in districts (NUTS-3) which belonged to the FRG or GDR, respectively, before the German reunification. The "Top-7 Regions" comprise the labor market regions of Berlin, Munich, Hamburg, Frankfurt am Main, Cologne, Stuttgart and Dusseldorf. "Urban" denotes all units belonging to a district indicated either as "Kreis", "Kreisfreie Stadt" or "Stadtkreis" and "Rural" all housing units located in a "Landkreis".

Table 2 reports the three terms in equation (2) separately for the years 2009, 2013 and 2018. Starting from the full sample, we see that the entire increase in variance is accounted for by the between-location component which increased steeply during 2009–2018, whereas the average within-location variance has not changed over time. In fact, while the within-location variance accounts for about 60 percent of the total variance in 2009, it merely contributes 38 percent to overall house price dispersion in 2018. Focusing on different geographic subsamples, this result is largely robust with some minor differences. In East German locations, house price dispersion has also gone up within locations, possibly reflecting rising disparities between unrenovated and modernized housing units (a housing characteristic that we cannot control in the hedonic regressions). In contrast, within urban and Top-7 locations, within-location dispersion has fallen, so that more than the entire increase of the variance is due to the between-location component.

The rising spatial dispersion of house prices is also illustrated in Figure 2 which shows the distribution of residual posted prices, averaged at the postal code level, in the four years 2009, 2012, 2015 and 2018. During 2009–2015, the mode of these distributions remains rather stable, while the rise of the average house price is driven by a widening of house prices in the upper half of the distribution. During 2015–2018, the bottom half of the distribution has also widened substantially.

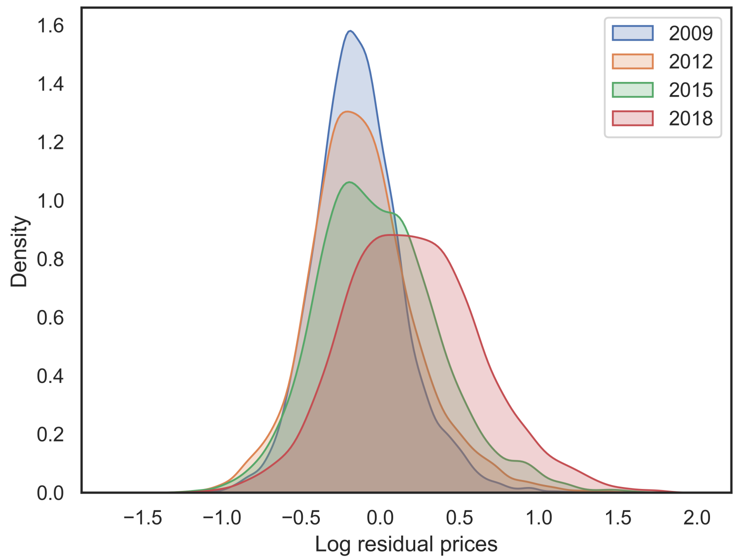


Figure 2: Distribution of residual prices across locations

NOTES: Between-location distributions of residual log prices in the years 2009 (blue), 2012 (orange), 2015 (green) and 2018 (red). The residuals are obtained from hedonic house price regressions as described in the main text and averaged in each location (postal code).

In light of the important role of location for rising house price dispersion, we are now asking to what extent these trends are driven by house price divergence between labor market regions or rising differences between locations within these regions. To do so, we decompose the between-location variance (i.e., the last term in equation (2)) into a between- and within-region component,

$$\underbrace{\sum_{i\in L} s_i(\bar{\varepsilon}_i - \bar{\varepsilon})^2}_{\text{between-location variance}} = \underbrace{\sum_{r\in R} \sigma_r \operatorname{var}_r(\bar{\varepsilon}_i)}_{\text{within regions}} + \underbrace{\sum_{r\in R} \sigma_r(\bar{\varepsilon}_r - \bar{\varepsilon})^2}_{\text{between regions}} , \tag{3}$$

where R is the set of regions, $\sigma_r = \sum_{i\in r} s_i$ is the listing weight of region r, and $\bar{\varepsilon}_r \equiv \sum_{i\in r} \frac{s_i}{\sigma_r}\bar{\varepsilon}_i$ is the mean residual price of region r. The first term is the listing-weighted average of the within-region variances $\operatorname{var}_r(\bar{\varepsilon}_i) \equiv \sum_{i\in r} \frac{s_i}{\sigma_r}(\bar{\varepsilon}_i - \bar{\varepsilon}_r)^2$, so that this term measures to what extent spatial house price differences are accounted for by differences between locations within labor market regions. The second term is the listing-weighted variance of average regional prices, i.e. the between-region variance. As before, this decomposition is calculated separately for each year. See Appendix E for derivations of the variance decompositions in equations (2) and (3).

Table 3: Within- and between-region variance decomposition

	Between Locations			**Within Regions**			**Between Regions**		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full Sample	0.077	0.127	0.182	0.033	0.050	0.055	0.045	0.078	0.127
West Germany	0.075	0.126	0.178	0.033	0.049	0.055	0.042	0.077	0.123
East Germany	0.056	0.103	0.137	0.032	0.052	0.049	0.024	0.051	0.088
Top-7 Regions	0.070	0.101	0.140	0.045	0.063	0.073	0.026	0.038	0.067
Urban	0.071	0.122	0.183	0.042	0.053	0.054	0.029	0.069	0.129
Rural	0.071	0.097	0.153	0.018	0.027	0.033	0.053	0.070	0.120

NOTES: See the notes to Table 2 for definitions of the different samples.

Table 3 shows the results of this decomposition for different years and geographic units. Two interesting patterns emerge. First, in the full sample about two thirds of the house price variance between locations in the year 2018 is accounted for by the between-region variance. Moreover, over three quarters of the rise in house price dispersion between 2009 and 2018 is driven by an increase in the variance of house prices between labor market regions, while less than a quarter of the increase is attributed to greater house price dispersion within labor market regions. Similar results are observed for the West and East German subsamples, and also if we divide the sample into rural and urban regions. On the other hand, zooming into the Top-7 subsample, we find that about 40 percent of the increase in variance is driven by diverging house prices within the labor market regions.[10] Furthermore, the within-location component accounts for the majority of overall spatial dispersion. Intuitively, labor market regions in this subsample are more comparable, so that a greater share of the variance (and its increase) is accounted for by the within-region variance (and its increase).[11]

[10] In Appendix D.1 we present the corresponding results when controls for distance to the regional center are applied in the hedonic regressions. We still observe an increase of within-region dispersion which is however a bit smaller. The explanation is that house price growth was on average faster in locations closer to the center; see Figure 14 for an illustration.

[11] In Appendix D.3 we repeat the analysis of this section for the rental market where the increase in dispersion over this ten-year period is less pronounced than in the sales market. Similarly to house prices, we find that most of the increase in variance is attributed to rising disparities across locations (postal codes),

Another way to illustrate the spatial divergence of house prices is shown in Figure 3 which plots the cumulative house price growth during 2009–2018 (i.e. the change of the average hedonic log price) on the vertical axis against the price level in the year 2009 on the horizontal axis. On average, house price growth is larger in those locations that had already a higher price level in the year 2009. This observation holds both across labor market regions (left panel) and across all postal codes (right panel) where the difference in house price growth between less expensive and more expensive locations is even larger, which points at the contribution of within-region house price divergence.

(a) Labor market regions (b) Postal codes

Figure 3: House price growth (2009–2018) across locations ranked by the price level in 2009

These empirical patterns do not, of course, settle the question of what caused rising house prices and a widening of spatial house price dispersion in the first place. In Section 5 we revisit this decomposition through the lens of our structural model that we calibrate on our data and that sheds light on the relative role of demand, supply and search frictions factors for the observed house price developments.

3 Model

We propose a simple model that can be calibrated on our data so as to analyze the driving forces behind the diverging house price trends documented in the last section. In particular, we aim to quantify the respective roles of demand, supply and search-friction shifters in house prices at the location, region and aggregate level during the ten-year horizon covered in our data. The model describes a given labor market region that is divided into locations

although rental dispersion also increases within locations. Furthermore, as for house sales, the increase in the cross-location variance is attributed to both within-region and between-region components where the latter plays a more important role.

(postal codes). In each location, potential sellers decide about entry and the posted price of the housing unit for sale. Buyers decide in which location to search and which sellers to contact at their posted prices where trade is subject to search frictions. The housing market is characterized by directed search (Moen, 1997; Wright et al., 2021), while location decisions respond to taste shocks that are common in spatial dynamic choice models (e.g. Aguirregabiria and Mira, 2010; Caliendo et al., 2019). House prices and housing market tightness endogenously depend on the time and space variation of buyer and seller housing valuations.

We deliberately keep the model parsimonious, abstracting from tenure choice, mortgage financing, differentiation of housing units by size or quality, and migration between labor market regions. While these simplifications leave out many important aspects of housing markets, they permit calibration of all key parameters on the basis of the listings data described in the previous section.

3.1 Environment

We consider a labor market region with a finite number of locations i (postal codes) over discrete time periods $t \geq 1$ (quarters). The region is populated by house buyers and sellers whose trade is subject to search frictions. While sellers solve a static problem, buyers aim to maximize discounted utility values with quarterly discount factor β. All prices, values and costs in the model are understood as quality- and inflation-adjusted prices, values and costs per square meter of a housing unit.

3.1.1 Sellers

There is a free entry of sellers whose housing unit has exogenous outside value K_{it} in location i in period t, which represents the value of an existing unit under alternative use, such as the discounted value of a lease or the monetized value of owner occupancy. Free entry requires that the endogenous value of a seller V_{it}^S equals K_{it} in all local markets (i, t). A housing unit for sale involves cost c per period, reflecting the user costs of a housing unit which is assumed to be constant across time and space.

3.1.2 Buyers

There is an exogenous inflow of new buyers into the region at time t, denoted B_t^n, so that the total number of buyers in the region, denoted B_t, is composed of unmatched buyers from the last period and the new buyers, where the stock of buyers in the first period B_1 is predetermined. Every buyer chooses in which location i to search in period t.[12] Search in location i yields utility value $V_{it}^B + \varphi_{it} + \tau_i$ where φ_{it} is an idiosyncratic (buyer-specific) taste shock which is type-I extreme value distributed with zero mean, and τ_i is a time-invariant location premium for location i that is common for all buyers and constant over time. V_{it}^B is the discounted utility value of a buyer searching in market i at time t, net of the taste shock and the location premium. If a buyer remains unmatched in market i, she decides in which location to search next period after drawing new idiosyncratic taste shocks. If a buyer is matched in period t, she pays the posted price and leaves the market with discounted utility value A_{it}. These values are exogenous to the model and represent the values that buyers attach to a (quality and size adjusted) housing unit in location i when bought at time t. In any period of search, we assume that the buyer pays a cost r_t which represents the rental cost in the region.

3.1.3 Search and Matching

Sellers post prices and buyers direct search to the sales listings, so that the housing market in a given location potentially segments into submarkets that are differentiated by posted prices and buyer-seller ratios. Both sides of the housing market trade off matching probabilities and prices, as is standard in markets with competitive search (Moen, 1997). When θ is the buyer-seller ratio (tightness) in a submarket, a seller is matched with probability $q_t(\theta)$ and a buyer is matched with probability $f_t(\theta) = q_t(\theta)/\theta$. q_t is a strictly increasing and strictly concave function, so that f_t is decreasing in tightness. We allow matching efficiency to vary over time which is why both functions are indexed by the time index t. Since all buyers and sellers searching in a given market (i, t) share the same respective values, only one submarket is active in this market which has posted price p_{it} and market tightness θ_{it}, both of which are equilibrium outcomes as described below.[13]

3.2 Value Functions and Equilibrium

The Bellman equation of sellers and buyers in market i and period t are

$$\begin{aligned} V_{it}^S &= -c + \beta V_{i,t+1}^S + q_t(\theta_{it}) \left(p_{it} - \beta V_{i,t+1}^S \right) , \\ V_{it}^B &= -r_t + \beta \bar{V}_{t+1}^B + f_t(\theta_{it}) \left(A_{it} - p_{it} - \beta \bar{V}_{t+1}^B \right) . \end{aligned}$$

[12] We rule out simultaneous search in multiple locations as we do not have enough information to discipline such a model feature.

[13] Although dispersion in residual prices exists *within* locations in our data, the within-location component exhibits no time trend, see Table 2 in Section 2. Given our interest in widening spatial house price dispersion over time, our model abstracts from this feature in the data.

A seller pays flow cost c in the current period and is matched with probability $q_t(\theta_{it})$ in which case she sells the house and hence leaves the market with continuation value p_{it}. Otherwise, she either continues to search in the same market or stops searching, yielding in both cases continuation utility $V^S_{i,t+1} = K_{i,t+1}$. A buyer pays flow cost r_t and is matched with probability $f_t(\theta_{it})$, yielding continuation utility $A_{it} - p_{it}$. Otherwise, an unmatched buyer has continuation utility

$$\bar{V}^B_{t+1} = \mathbb{E}\max_j [V^B_{j,t+1} + \varphi_{j,t+1} + \tau_j] = \ln\left[\sum_j e^{V^B_{j,t+1}+\tau_j}\right] , \tag{4}$$

where the expectation is over the realization of next period's idiosyncratic taste shocks $\varphi_{j,t+1}$.

In every local market (i,t), sellers post prices and buyers direct their search to the posted prices. Let (p,θ) denote the price-tightness combination in a potential submarket. Let Ω_{it} denote the expected buyer surplus from searching in market (i,t) which is identical for all (homogeneous) buyers in that market. Buyers must be offered at least surplus Ω_{it} to be willing to search in submarket (p,θ). A seller chooses (p,θ) to maximize the expected gain from trade,

$$\max_{p,\theta} q_t(\theta)[p - \beta V^S_{i,t+1}] \qquad \text{s.t.} \qquad f_t(\theta)[A_{it} - p - \beta \bar{V}^B_{t+1}] \geq \Omega_{it} \ .$$

The constraint says that sellers must offer at least surplus Ω_{it} to attract buyers to the submarket. Substituting the price and the matching function $f_t(\theta) = q_t(\theta)/\theta$ yields the first-order condition

$$\Omega_{it} = q'_t(\theta)[A_{it} - \beta \bar{V}^B_{t+1} - \beta V^S_{i,t+1}] \ .$$

Because the matching function q_t is strictly concave, all sellers in market (i,t) choose the same price p_{it}, so that only one submarket is active with tightness θ_{it}. Using $\Omega_{it} = f_t(\theta_{it})[A_{it} - p_{it} - \beta \bar{V}^B_{t+1}]$ and $f_t(\theta)\theta = q_t(\theta)$ gives the equilibrium price

$$p_{it} = \zeta_t(\theta_{it})\beta V^S_{i,t+1} + (1 - \zeta_t(\theta_{it}))[A_{it} - \beta \bar{V}^B_{t+1}] \ , \tag{5}$$

with matching function elasticity $\zeta_t(\theta) = q'_t(\theta)\theta/q_t(\theta) \in (0,1)$. This equation demonstrates how the posted price in market (i,t) depends on housing supply (the sellers' valuation $\beta V^S_{i,t+1}$), housing demand (the buyers' gain from trade $A_{it} - \beta \bar{V}^B_{t+1}$), and the search-frictions factor $\zeta_t(\theta_{it})$ which responds to features of the matching technology and housing market tightness in market (i,t). We build on this equation for our decomposition analysis in Section 5.

Substituting the equilibrium price into the Bellman equations gives

$$V_{it}^S = -c + \beta V_{i,t+1}^S + (q_t(\theta_{it}) - \theta_{it} q_t'(\theta_{it})) \left[A_{it} - \beta \bar{V}_{t+1}^B - \beta V_{i,t+1}^S\right] \ , \tag{6}$$

$$V_{it}^B = -r_t + \beta \bar{V}_{t+1}^B + q_t'(\theta_{it}) \left[A_{it} - \beta \bar{V}_{t+1}^B - \beta V_{i,t+1}^S\right] \ . \tag{7}$$

At the beginning of a period, all buyers B_t in a labor market region draw idiosyncratic taste shocks φ_{it} after which fraction

$$\pi_{it} = \frac{e^{V_{it}^B + \tau_i}}{\sum_j e^{V_{jt}^B + \tau_j}} \tag{8}$$

decide to search in location i. Over time, the number of buyers in the labor market region adjusts according to

$$B_{t+1} = \sum_i [1 - f_t(\theta_{it})] \pi_{it} B_t + B_{t+1}^n \ , \tag{9}$$

where B_{t+1}^n is the exogenous inflow of new buyers into the labor market region in period $t+1$ which adds to the number of unmatched buyers from the previous period.

Equilibrium Definition

Given an initial stock of buyers B_1 and buyer inflow B_t^n in periods $t \geq 2$, a *spatial competitive search equilibrium* describes, for all periods $t \geq 1$ and locations i, posted house prices p_{it}, number of buyers and sellers B_{it} and S_{it}, market tightness $\theta_{it} = B_{it}/S_{it}$, discounted values of sellers and buyers V_{it}^S, $\bar{V}_t^B$, V_{it}^B, location choices π_{it} and buyer stocks B_t satisfying equations (4)–(9), the free-entry conditions of sellers $V_{it}^S = K_{it}$, and $B_{it} = \pi_{it} B_t$.

4 Calibration

In this section, we explain how we calibrate the parameters of this model for a given labor market region with $i = 1, \ldots, N$ locations (postal codes) and $t = 1, \ldots, T$ periods (quarters).[14] We use for calibration the baseline sample described in Section 2.3 with variables aggregated at the location-quarter level. These are the residualized average hedonic price

[14]We choose a quarterly period length to smooth out very short-term volatility at the local level that might partly arise due to a low number of observations. Further, a quarter plausibly reflects the typical planned transaction time for buyers and sellers in the housing market. In the data, we assign the day of first listing to a specific quarter so that some days of an active listing may fall into the next quarter.

p_{it},[15] the number of listings S_{it} which we identify with the number of sellers,[16] average duration of a listing in days d_{it} and the number of buyer contacts C_{it}. Note that the stock of buyers, and therefore market tightness, is not observed. We explain below our identifying assumptions that allow us to back out these values and to estimate a matching function from information on listing duration d_{it} and the numbers of contacts C_{it} and listings S_{it}. To remove outliers, we winsorize these variables at the 2nd and 98th percentiles.

Three model parameters are calibrated outside the model. β is a standard discount factor at quarterly frequency that equals 0.995 to match an annual interest rate of 2%. The flow cost of a vacant housing unit, parameter c, is set to 34.7 Euros which is an estimate of the quarterly user cost per square meter.[17] For the quarterly costs of an unmatched buyer r_t, we use the average, inflation-adjusted rental rate per m^2 in the region which we take from the rental listings in the RWI-GEO-RED dataset.

The model calibration proceeds in two steps. First, we estimate a matching function, separate for each region, using the data on listing duration, number of listings and contacts. Second, we back out the buyer and seller valuations A_{it} and K_{it} that are consistent with the observed variation of prices, tightness and matching rates across time and space, and we calibrate the location premia τ_i which control the distribution of buyers across locations.

4.1 Matching Function

Stock of Buyers and Market Tightness

In the data, we measure the stock of sellers by the number of listings S_{it}, but we do not observe the stock of buyers in a given market (i, t), denoted B_{it}. Hence market tightness $\theta_{it} = B_{it}/S_{it}$ is unobserved. However, we build on the assumption that the search intensity of every active buyer is the same so that every buyer contacts a given number of listings per day which we denote by k.[18]

Given listing duration in days d_{it}, the daily matching probability of a seller is $q^d_{it} = 1/d_{it}$. When f^d_{it} is the daily matching probability of a buyer, equality of matched buyers and sellers

[15] Specifically, we take the residual of the hedonic, inflation-adjusted log prices per m^2 at the listing level ε_{ht} as defined in Section 2.1, delog and multiply them with the average price in euros, and average to the listing level to obtain p_{it}.

[16] Multiple listings of the same seller should not raise concerns since they show up as independent listings on the *ImmobilienScout24* platform. Note that we only consider listings of single residential housing units in our analysis.

[17] Balz (2024) reports monthly user costs for larger German cities whose average is 9.4 Euros. We add to this number an estimate of utility, insurance costs and property taxes of 2.17 Euros per month (https://www.mieterbund.de/service/betriebskostenspiegel.html) and multiply by three to obtain the quarterly number.

[18] Here we follow the logic of matching function estimation in labor market models where typically the stock of unemployed workers is observed, but flow measures of search intensity (e.g. the number of applications sent per day) are not observed. In our data, we only observe the flow of contacts, but not the stock of buyers.

implies

$$q_{it}^d S_{it} = f_{it}^d B_{it} \ .$$

On the other hand, a given buyer searches on average for $1/f_{it}^d$ days and contacts k listings per day, so that the number of contacts in location i and quarter t is

$$C_{it} = \frac{kB_{it}}{f_{it}^d} \ .$$

From these two relationships, we obtain that the number of buyers is

$$B_{it} = \left(\frac{q_{it}^d S_{it} C_{it}}{k} \right)^{1/2} \ .$$

Therefore, for a given parameter k, we infer B_{it} from our data on $q_{it}^d = 1/d_{it}$, C_{it} and S_{it}, so that we measure tightness

$$\theta_{it} = \frac{B_{it}}{S_{it}} = \left(\frac{q_{it}^d C_{it}}{kS_{it}} \right)^{1/2} \ . \tag{10}$$

We choose parameter k to ensure that the implied buyer matching probability is smaller than one in our data. To do so, we proceed as follows: For each observation (i,t), we calculate the implied daily buyer matching probability:

$$f_{i,t}^d = q_{i,t}^d \frac{S_{i,t}}{B_{i,t}} = k^{1/2} \left(\frac{q_{i,t}^d S_{i,t}}{C_{i,t}} \right)^{1/2} \ .$$

Note that the second term on the right-hand side is implied by the data. We set k to be the highest number such that $f_{i,t}^d \leq 1$ for all (i,t) which yields $k = 1.51$, i.e. a buyer contacts on average 1.51 listings per day. In a robustness analysis, we set k to much lower or higher values and verify that our main results are unaffected. Indeed, as will become clear below, the calibration of k impacts only our estimates of the matching function *scale* but has no impact on the estimated matching function *elasticity* which is key for our price decomposition analysis.

Parameterization and Estimation of the Daily Matching Function

We assume that the daily matching probability of a seller is Cobb-Douglas: $q^d(\theta) = q_0^d \theta^{\mu^d}$ with scale parameter $q_0^d > 0$ and matching function elasticity $\mu^d \in (0,1)$. We pool all locations and quarters in a labor market region and estimate this matching function relationship separately for each region, allowing for the possibility of time dependence of the matching function scale which takes care of any trends or cyclicality in the listing duration relationship, as well as seasonality in the housing markets as documented in previous literature (cf. Ngai and Tenreyro, 2014). Taking logs yields the estimation equation[19]

$$\ln q^d_{it} = \ln q^d_0 + \mu^d \ln \theta_{it} + g_t + \epsilon_{it} \ , \tag{11}$$

where ε_{it} is an error term and g_t is the time fixed effect. Principally, estimates of matching functions are subject to an endogeneity concern which arises because shocks to matching efficiency may be correlated with the search behavior of buyers and sellers in the market (see e.g. Borowczyk-Martins et al., 2013). Estimating equation (11) with OLS, our identification assumption is that the error term ϵ_{it} is uncorrelated with market tightness θ_{it}.

Table 4: Matching function estimation (Top-7 regions)

	Berlin	Munich	Hamburg	Frankfurt	Stuttgart	Düsseldorf	Köln
μ^d	0.63***	0.79***	0.61***	0.61***	0.65***	0.55***	0.48***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\ln q^d_0$	-2.57***	-2.57***	-2.84***	-2.69***	-2.96***	-2.93***	-3.08***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Time FE	yes	yes	yes	yes	yes	yes	yes
R^2	0.276	0.431	0.341	0.356	0.502	0.351	0.428
N	5,440	3,440	3,760	3,960	2,800	3,680	2,720

NOTES: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 shows the estimates of parameters $\ln q^d_0$ and μ^d for each of the Top-7 labor market regions. As expected, in all cases parameter μ^d is in the interval $(0, 1)$. The estimates show that a doubling of the buyer-seller ratio goes together with an increase of the seller matching probability (a decrease of sales duration) by 50–80 percent in the Top-7 regions. Figure 4 shows the distribution of the estimates of μ^d across all labor market regions in Germany which shows that the Top-7 regions fall in the upper range of this distribution, while the majority of labor market regions have estimates of μ^d below 0.5.

The regression constant also varies between regions, showing that the seller matching probability (listing duration) is about 50 log points shorter (longer) in Cologne than in Berlin or Munich in the first quarter of 2009 (the reference category for the time fixed effect) and when the buyer-seller ratio equals one (so that $\ln(\theta) = 0$). The time trends, which are shown in Appendix D, Table 25, for the Top-7 regions, also show some heterogeneity

[19] Since parameter k enters our estimated values of θ_{it} proportionately (see equation (10)) it affects the estimate of the regression constant $\ln q^d_0$ (matching function scale) but has no impact on the estimated daily matching function elasticity μ^d.

between regions, but generally remain rather flat until 2014 after which the seller matching probability has decreased, conditional on holding housing market tightness constant.

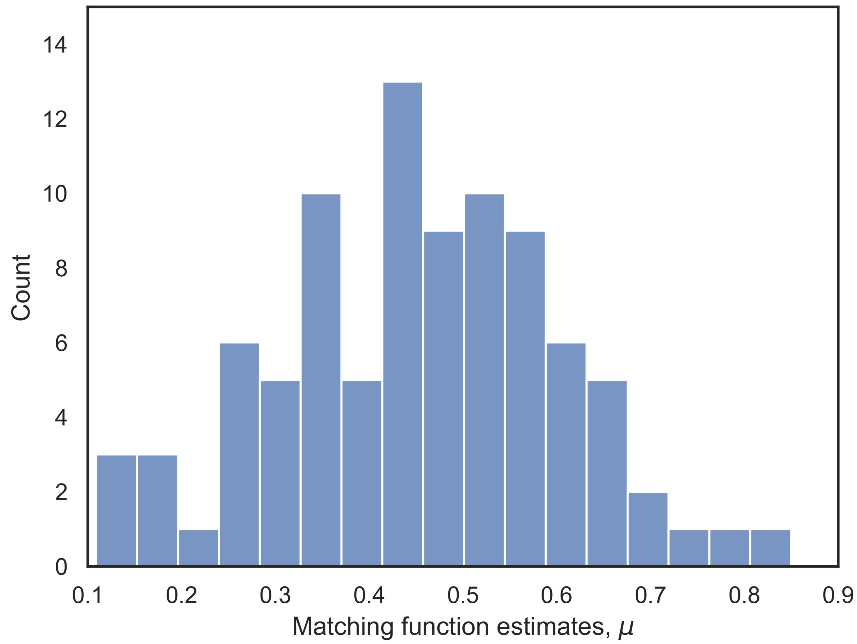


Figure 4: Distribution of estimates of μ^d across labor market regions

NOTES: The distribution includes all estimates of μ^d which are statistically different from zero.

Quarterly Matching Function

Together with the estimated daily matching function parameters, we obtain quarterly matching probabilities for buyers and sellers, i.e. the matching function relationships used in Section 3:

$$
\begin{aligned}
q_t(\theta) &= 1 - \left(1 - q^d_{0,t}\theta^{\mu^d}\right)^{90} \ , \\
f_t(\theta) &= q_t(\theta)/\theta \ ,
\end{aligned}
$$

with $q^d_{0,t} = q^d_0 e^{g_t}$ denoting the (time-varying) scale of the estimated daily seller matching function. Given this parameterization, the elasticity of the matching function, which enters the equilibrium price equation (5), is calculated as follows:

$$
\zeta_t(\theta) = \frac{90\mu^d(1 - q^d_{0,t}\theta^{\mu^d})^{89} q^d_{0,t}\theta^{\mu}}{1 - (1 - q^d_{0,t}\theta^{\mu^d})^{90}} \ . \tag{12}
$$

The elasticity is decreasing in tightness θ and in the efficiency parameter $q^d_{0,t}$ and it is increasing in the elasticity of the daily matching function μ^d. Different from the Cobb-Douglas daily matching function whose elasticity is independent of θ, our quarterly matching function elasticity not only depends on tightness θ but also on the scale parameter of the daily matching function, and hence responds to the calibrated value of parameter k (buyer contacts per day). While we verify that our main decomposition results are not sensitive to our parametrization of k, we acknowledge that the quarterly matching function elasticity and its dependence on parameters hinges on this time aggregation procedure.

4.2 Location Premia and Valuations of Buyers and Sellers

The second step of our calibration procedure is to simultaneously set the time-invariant location premia τ_i and the time-varying buyer and seller valuations of a housing unit, A_{it} and K_{it}. The latter two objects can be uniquely pinned down to exactly match the observed prices p_{it} and tightness levels θ_{it} which are both measured from our data as described before. The exact matching is achieved since the exogenous valuation parameters A_{it} and K_{it} enter linearly the Bellman equations, the pricing equation and the sellers' free-entry condition and hence are the solution of an invertible linear equation system, as we explain further below and in Appendix F. The time-invariant location premia τ_i, in turn, are set to match the average buyer market shares in all locations, to be described as follows.

Let $\hat{\pi}_{it} = \frac{\hat{B}_{it}}{\hat{B}_t}$ denote the share of buyers in market i at time t in the data where buyers in market (i, t) and the total buyer stock are measured as explained in the previous subsection. The market shares in the model π_{it} and in the data differ according to

$$\hat{\pi}_{it} = \pi_{it} e^{\eta_{it}} ,$$

where η_{it} is the log difference between the data and model market shares. We choose location premia τ_i to minimize the sum of these squared log differences, namely $\sum_{i,t} \eta^2_{it}$, subject to the requirement that the average location premium is normalized to zero, $\sum_i \tau_i = 0$.

From (4) and (8) follows

$$V^B_{it} + \tau_i = \ln \pi_{it} + \bar{V}^B_t , \tag{13}$$

so that we can write

$$\eta_{it} = \ln \hat{\pi}_{it} - \ln \pi_{it} = \ln \hat{\pi}_{it} + \bar{V}^B_t - V^B_{it} - \tau_i .$$

Minimization of $\sum_{i,t} \eta_{it}^2$ subject to the constraint

$$\sum_i \tau_i = 0 \tag{14}$$

with respect to τ_i has the first-order conditions[20]

$$\tau_i = \frac{1}{T} \sum_{t=1}^{T} \left[\ln \hat{\pi}_{it} + \bar{V}_t^B - V_{it}^B \right] - \frac{\lambda}{2T} \ . \tag{15}$$

where λ is the multiplier on the constraint.

Since our model is set up as an infinite-horizon model, we need to make an assumption about the forecasts of continuation values of buyers and sellers in the last observation period T. We do this by linearly extrapolating the values during the observation period, namely V_{it}^S and V_{it}^B for $t = 1, \ldots, T$, to the first quarter thereafter. The analytic expressions of this extrapolation procedure are:

$$V_{i,T+1}^S = \frac{2}{T(T-1)} \left\{ \sum_{t=1}^{T} V_{it}^S \left[3t - (T+2) \right] \right\} \ , \tag{16}$$

$$V_{i,T+1}^B = \frac{2}{T(T-1)} \left\{ \sum_{t=1}^{T} V_{it}^B \left[3t - (T+2) \right] \right\} . \tag{17}$$

We are now ready to explain how the unknown valuation parameters (A_{it}, K_{it}) and location premia τ_i can be calculated. From our data, we use the (quality- and inflation-adjusted, per square meter) prices p_{it}, tightness θ_{it} as calculated above, the estimated (time-varying) matching function relationships, and the buyer market shares $\hat{\pi}_{it}$ as defined above. Then the pricing equations (5), the two Bellman equations (6) and (7), the extrapolation equations (16) and (17), the optimality conditions (14) and (15), and the continuation utilities of unmatched buyers (4) constitute a system of $(3N+1)(T+1)+1$ equations in $(3N+1)(T+1)+1$ unknowns: $(A_{it})_{t=1}^T$, $(V_{it}^B, V_{it}^S)_{t=1}^{T+1}$, τ_i for $i = 1, \ldots, N$, λ, and $(\bar{V}_t^B)_{t=1}^{T+1}$. Except the $T+1$ equations (4), these are all linear equations. Their joint solution is straightforward to implement, with further details described in Appendix F. The solution gives the buyer and seller valuations A_{it}, $K_{it} = V_{it}^S$, and location premia τ_i.

By construction, the pricing equation and the Bellman equations hold with equality, so that our model matches prices and tightness in all markets (i, t) exactly. Put differently, (p_{it}, θ_{it}) pin down the demand and supply parameters (A_{it}, K_{it}) uniquely. It follows that any

[20] The minimization takes the values $\bar{V}_t^B$ as given, and hence ignores the impact of τ_i on the values of unmatched buyers, see equation (4). This approximation is innocuous when the number of locations is large so that the impact of each τ_i on $\bar{V}_t^B$ is negligible.

spatial correlation of these variables in the data will be reflected in a spatial correlation of the supply and demand parameters in the calibrated model. Likewise, the location premia τ_i are set to match average market shares across locations, and therefore also respond to the spatial correlation of the underlying data moments. In Section 5.5 we explore the spatial correlation structure of these parameters.

Given the solution of supply and demand parameters and location premia, the fraction of buyers searching in market (i,t) follows from equation (8). Setting the initial aggregate stock of buyers to the value recovered from the data (see the previous subsection), we back out the net inflow of new buyers in all periods t from the stock-flow identity (9).

5 Results

In this section, we use the calibrated model to study what factors contribute to house price dynamics in terms of the observed rise in house prices and their spatial dispersion. We distinguish between three contributing forces: *(i)* housing supply factors related to the location- and time-specific seller valuations K_{it}, *(ii)* housing demand factors represented by the buyer gains from trade $A_{it} - \beta \bar{V}^B_{r,t+1}$, and, *(iii)* search-frictions factors associated with the region- and time-specific matching function $q_{rt}(\cdot)$ as well as market tightness θ_{it}.[21] We first analyze the determinants of price changes over time and price differences across locations in subsections 5.1 and 5.2, where we separately report results for the full sample and for the more homogeneous subsample of Top-7 labor market regions which exhibited the strongest house price growth during 2009–2018. Subsequently, we discuss in subsection 5.3 the role of the search-frictions factor and in subsection 5.4 how the demand and supply factors in our model relate to external indicators of housing demand and supply. Finally, in subsection 5.5 we analyze the spatial correlation of the housing demand and supply valuations and the calibrated location premia in our model.

5.1 Decomposition of Price Changes

We use the hedonic prices p_{it}, tightness levels θ_{it}, the calibrated valuation parameters K_{it}, A_{it} and $\bar{V}^B_{rt}$, and the estimated matching functions $q_{rt}(\cdot)$ to isolate the role of factors *(i)*-*(iii)* for generating the observed house price dynamics. Building on the equilibrium pricing equation (5), we express the price in location i at time t into the following terms:

$$p_{it} = \underbrace{\zeta_{rt}(\theta_{it})}_{\text{Search Frictions}} \underbrace{K_{i,t}}_{\text{Supply}} + \underbrace{(1 - \zeta_{rt}(\theta_{it}))}_{\text{Search Frictions}} \underbrace{\left[A_{it} - \beta \bar{V}^B_{r,t+1}\right]}_{\text{Demand}}. \tag{18}$$

[21] We add the subscript r to variables which are common across markets (i,t) within a labor market region r but differ across regions.

To measure the respective contributions of the three factors, we proceed as follows. Regarding the contribution of housing supply, we fix housing demand via the buyers' gain and the search-frictions factor to their initial 2009 values, $A_{i,1} - \beta \bar{V}^{B}_{r,2}$ and $\zeta_{r,1}(\theta_{i,1})$, while allowing housing supply via the sellers' valuations $K_{i,t}$ to evolve. By doing so, we derive counterfactual prices p^{sup}_{it} through the pricing equation which reflect only the shifts in housing supply. Similarly, fixing housing supply and the search-frictions terms to their initial values, while letting housing demand evolve, we derive another set of counterfactual prices p^{dem}_{it} which reflect only shifts in housing demand. Finally, the constructed prices p^{sf}_{it} summarize only changes in the search-frictions factor $\zeta_{rt}(\theta_{it})$. Note that in all these experiments, we do not solve for a new equilibrium of the model, but rather use the pricing equation with the calibrated parameters to compute counterfactual prices.

Table 5 reports how the three factors have contributed to the overall changes of the mean and standard deviation of log prices in the data for the full sample and for the subsample of Top-7 regions. Beginning with mean price growth, housing demand in isolation produces a price change $\bar{p}^{\text{dem}}_{T} - \bar{p}_1$, while housing supply contributes $\bar{p}^{\text{sup}}_{T} - \bar{p}_1$. Finally, changes in search frictions account for $\bar{p}^{\text{sf}}_{T} - \bar{p}_1$. The numbers in parentheses show the percent of the overall price change generated by the different factors, separately for each subsample. Note that these percentages do not add to 100 since the three counterfactual scenarios build on a non-linear equation.

Table 5: Decomposition of changes of the mean and standard deviation of house prices

	Data	**Demand**	**Supply**	**Search Frictions**
Mean	$\bar{p}_T - \bar{p}_1$	$\bar{p}^{\text{dem}}_{T} - \bar{p}_1$	$\bar{p}^{\text{sup}}_{T} - \bar{p}_1$	$\bar{p}^{\text{sf}}_{T} - \bar{p}_1$
Full Sample	0.293	0.244	0.052	-0.009
	(100)	(83)	(18)	(-3)
Top-7 Regions	0.443	0.371	0.089	-0.014
	(100)	(84)	(20)	(-3)
Standard Dev.	$\text{sd}(p_T) - \text{sd}(p_1)$	$\text{sd}(p^{\text{dem}}_{T}) - \text{sd}(p_1)$	$\text{sd}(p^{\text{sup}}_{T}) - \text{sd}(p_1)$	$\text{sd}(p^{\text{sf}}_{T}) - \text{sd}(p_1)$
Full Sample	0.146	0.144	0.028	-0.006
	(100)	(98)	(19)	(-4)
Top-7 Regions	0.110	0.096	0.018	-0.002
	(100)	(87)	(16)	(-2)

NOTES: The demand, supply and search-frictions contributions to the change of the mean and standard deviation of log prices between 2009 and 2018 are derived as described in the text. Percentages of the total changes for each subsample are shown in parentheses.

It can be seen that the demand factor contributes the most to the overall price increase, explaining 83 percent of the price increase for the full sample where prices increased by 0.293 log points. For the subsample of the Top-7 labor market regions, where prices rose by 0.443 log points, the demand factor is also the dominant force. Housing supply contributes around 20 percent to the overall price increase, while search frictions play only a modest dampening role.

Turning to the second moment of house prices, the bottom panel of Table 5 displays the contribution of factors *(i)–(iii)* to the 2009–2018 change in spatial price dispersion, as measured by the standard deviation of prices across locations. We find that changes in housing demand account for nearly the entire increase in dispersion: 98 percent of the rise in the full sample and 87 percent in the Top-7 labor market regions, where supply contributes 16–19 percent while search frictions have a small and slightly negative effect. These findings suggest that differential shifts in buyer valuations have been the dominant force behind rising house price dispersion across Germany.

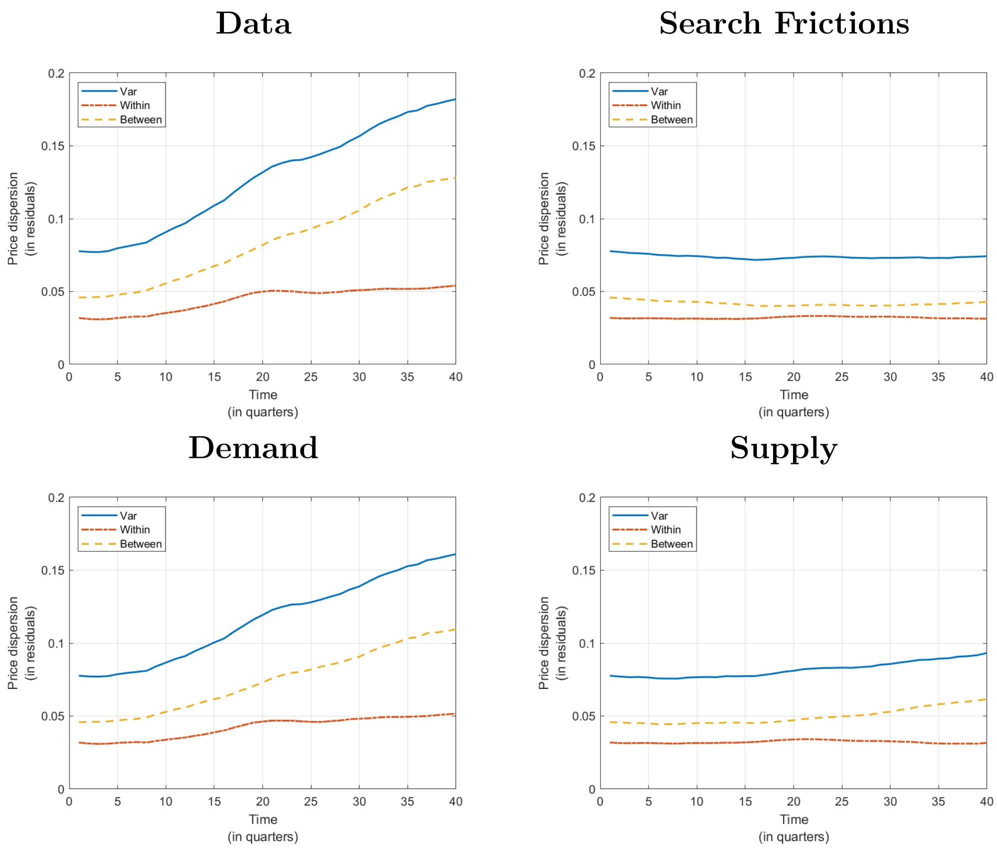


Figure 5: Variance decomposition of within- and between-region price changes

NOTES: Model-based variance decomposition of equation (3) for all the regions in 2009-2018. Within (red line) depicts the within-region dispersion, whereas Between (yellow line) refers to dispersion coming from across labor market regions. The sum of within- and between-regions dispersion equals the total variance (blue line).

We further use the counterfactual model-generated house prices to decompose the variance of house prices into within- and between-region components as in equation (3). Figure 5 reports graphically the results for each quarter for the full sample of all regions of the model-based variance decomposition. The top-left plot shows the variance decomposition using the actual prices, reiterating the results from Table 3: both within- and between-regions dispersion increase over time. However, between-regions dispersion contributes more to the overall variance increase than the within-regions dispersion. The top-right plot of Figure 5 displays the time evolution of the overall variance as well as its within- and between-regions components coming from changes in the search-frictions factor, while the two bottom panels depict the same thing but in the cases in which only housing demand or housing supply changes are at work. In line with the results of Table 5, the figure shows that rising dispersion is mostly driven by divergent trends in housing demand, which is true both within and across labor market regions. The search-frictions term does not contribute to rising overall dispersion and even contributes negatively to the within-region component. Housing supply contributes to price divergence across regions but not for rising within-region dispersion.

5.2 Within-Region Price Dispersion

While the previous subsection examines the driving forces of *price changes over time*, we can also use the pricing equation (18) to understand the determinants of *price differences in the cross section* at a given point in time. To do so, we fix two of the three components "demand", "supply" and "search frictions" at their respective means of a labor market region in a given year, and then calculate the counterfactual prices for this region and year where only the remaining component is allowed to vary across locations. For each of these counterfactual scenarios, we calculate the standard deviation of log prices which we average over all labor market regions. The results of this exercise for the three years 2009, 2013 and 2018 are reported in Table 6, separately as averages over all labor market regions and only for the Top-7 regions.

Similar to our results on price changes over time, we find that the demand factor is the most important contributor to within-region price differences, accounting for over 80 percent of the overall dispersion. The supply and search-frictions factors both play a less prominent role. Moreover, the relative importance of the supply factor for within-region dispersion has declined between 2009 and 2013. In the next two subsections we shed further light on the determinants of these three factors.

5.3 The Role of Search Frictions

The previous subsections document that the search-frictions term $\zeta_t(\theta_{it})$, i.e. the elasticity of the matching function, plays only a very small role for aggregate price trends (Table 5) and for within-region price differences (Table 6) and it even has a small negative effect on the change of dispersion (bottom panel of Table 5).

To shed light on this result, consider first how the distribution of $\zeta_t(\theta_{it})$ has changed during 2009–2018 in Figure 6. Over time, the mean and the mode of the distribution have fallen, while the distribution became a bit more concentrated; The first observation implies that the weight on the demand term in equation (18) has become larger which, however, did not contribute to average price growth. The second observation helps to understand why there is a small and negative contribution to the change of house price dispersion. Moreover, the relatively low values of ζ_t also explain why demand generally plays a more prominent role than supply for the house price developments in our model.

Table 6: Decomposition of within-region price dispersion

	Data	**Demand**	**Supply**	**Search Frictions**
	sd(p)	sd(p^{dem})	sd(p^{sup})	sd(p^{sf})
			2009	
Full Sample	0.145	0.119	0.026	0.001
	(100)	(82)	(18)	(1)
Top-7 Regions	0.199	0.160	0.040	0.001
	(100)	(80)	(20)	(1)
			2013	
Full Sample	0.185	0.158	0.026	0.001
	(100)	(86)	(14)	(0.2)
Top-7 Regions	0.246	0.210	0.036	0.001
	(100)	(85)	(15)	(0.1)
			2018	
Full Sample	0.205	0.174	0.028	0.001
	(100)	(85)	(14)	(0.1)
Top-7 Regions	0.263	0.219	0.040	0.001
	(100)	(83)	(15)	(0.1)

NOTES: The demand, supply and search-frictions contributions to the standard deviations of log prices in 2009, 2013 and 2018 within labor market regions are derived as described in the text. For each city we first average the four quarterly values that fall in the year and then take the simple mean across all labour-market regions. Percentages of the total standard deviation for each year and each subsample are shown in parentheses.

We conclude that the convergence of matching function elasticities across locations explains why search frictions did not contribute to the observed rise in price dispersion. There could be different reasons for converging matching frictions over time, such as a more similar search behavior of buyers and sellers or a more homogeneous use of standardized web platforms across locations. Further note that this result is derived from the identifying assumption that buyers are searching in only one location in a given quarter. If, in response to the house price boom, buyers increasingly search simultaneously in different locations (i.e. buyer search intensity has an extensive and intensive spatial component), the estimated matching function elasticities and their contribution to price dispersion might be different. As we do not observe individual buyer behavior in our data, we cannot explore this possibility, however.

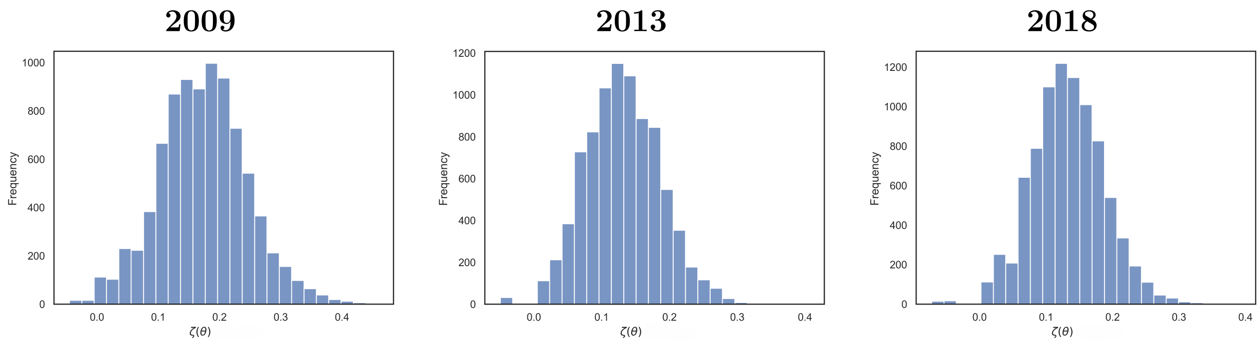


Figure 6: Distribution of the matching function elasticity $\zeta_t(\theta_{it})$ across locations

Notes: Histograms of the matching function elasticity $\zeta_t(\theta_{it})$ over all locations for years 2009, 2013 and 2018.

5.4 Indicators of Housing Demand and Housing Supply

To validate that the housing demand and supply components estimated by our model reflect economically meaningful variation, we utilize observable indicators of housing market fundamentals that vary across locations (at the district level) and over time. By demonstrating that the estimated structural components align with these indicators, we provide external validation and lend credibility to our modeling approach. We acknowledge, however, that certain factors, such as interest rates or own-to-own moves, may simultaneously influence both supply and demand (see e.g. Anenberg and Bayer, 2020; Ngai and Sheedy, 2020; Groiss and Syrichas, 2025). While our data limitations prevent us from explicitly analyzing the impact of own-to-own moves on both sides of the market, we discuss the role of different housing market fundamentals separately for supply and demand.

Table 7: Indicators of housing demand

Dependent variable:	(1) $A_{i,2009}$	(2) $A_{i,2018}$	(3) $A_{i,2018} - A_{i,2009}$
Distance to the center	-0.0032***	-0.0110***	-0.0077***
	(0.0008)	(0.0014)	(0.0009)
Age of the population	-0.0194	-0.1070***	-0.0747***
	(0.0132)	(0.0217)	(0.0136)
Domestic migration balance	-0.0008	-0.0199**	-0.0121*
	(0.0034)	(0.0069)	(0.0048)
Household income per capita	0.0455***	0.1150***	0.0400***
	(0.0069)	(0.0132)	(0.0072)
Share of subprime credit scores	-0.0495***	-0.0405***	-0.0330***
	(0.0075)	(0.0121)	(0.0075)
Population density	0.0001***	0.0002***	0.0002***
	(0.0000)	(0.0000)	(0.0000)
Share of academics	0.0254***	0.0233***	0.0123**
	(0.0037)	(0.0057)	(0.0038)
Region effects	Yes	Yes	Yes
N	2,275	2,275	2,275

NOTES: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The indicators of demand (and later, of supply) are taken from the district-level panel dataset built by Brausewetter et al. (2024). These demand-side features are: age of population, domestic migration, household income per capita, share of subprime credit scores, population density and share of academics. In addition, we include the distance to the center of the labor market region from our own geocode information averaged at the postcode level. Table 7 presents regression results examining the relationship between buyer valuations ($A_{i,t}$) and these various district-level demand indicators for the years 2009 and 2018, as well as the change between these years. The analysis includes region fixed effects to control for unobserved heterogeneity across regions.

The regression results confirm that our model-implied demand component is closely related to fundamental demand drivers, with most coefficients showing the expected signs. In both 2009 and 2018, districts with higher household income per capita exhibit significantly higher buyer valuations, whereas those with a larger share of subprime borrowers (lower credit quality) have lower buyer valuations. This suggests that wealthier areas with better credit profiles experience stronger housing demand. Population density and the share of academics are also positively associated with $A_{i,t}$ in both years, indicating that more urbanized and high-human-capital districts have higher underlying housing demand. In contrast, an older population is correlated with lower buyer valuations (a relationship that is statistically significant in 2018), consistent with younger communities contributing to stronger housing demand. Notably, the coefficient on distance to the city center is negative and grows in magnitude over time: it is relatively small in 2009 but grows fourfold by 2018. This emerging gradient implies a growing preference for centrally located properties, with demand intensifying in urban cores relative to peripheral areas, which helps explain our descriptive finding that house price growth during 2009–2018 was faster in more central locations (see Figure 14).

Examining the change in buyer valuations from 2009 to 2018 reinforces these patterns. Column (3) of Table 7 shows that districts closer to city centers, with younger populations, higher incomes, greater population density, and a more educated population experienced significantly larger increases in $A_{i,t}$ over the period. In other words, the housing boom's demand growth was concentrated in centrally located, economically strong, and highly educated locations. One notable exception in the demand regression is the domestic migration balance, which appears with an unexpected negative sign in 2009 and 2018 regressions and in the change specification. This result implies that, after controlling for other factors and regional effects, districts with higher net in-migration did not necessarily show higher buyer valuations. If anything, they saw slightly smaller demand gains. Overall, the alignment of the main fundamental indicators with $A_{i,t}$ lends credibility to our demand measure.

We perform a similar validation exercise for the model's housing supply component. Specifically, we ask whether the model-based seller valuations ($K_{i,t}$) correspond to observable supply-side fundamentals of local housing markets. The supply-side indicators we consider, also drawn from Brausewetter et al. (2024), include: average residential land prices, living area per capita (a proxy for housing stock or space availability), the number of completed new apartments per capita (a measure of recent construction activity), and the share of small apartments in the housing stock. Table 8 reports regressions of the supply component $K_{i,t}$ on these indicators and on the distance to the regional center, separately for 2009 and 2018, and the change between 2009 and 2018, again including region fixed effects.

The results point out that in both 2009 and 2018, locations with higher land prices tend to have significantly higher values of $K_{i,t}$, and likewise locations with a greater number of completed apartments per capita show higher $K_{i,t}$. These relationships suggest that our estimated supply component indeed captures variation in local building activity and land scarcity. The living area per capita is positively related to $K_{i,t}$ as well: in both 2009 and 2018, locations with more housing space per person had higher seller valuations. This initially seems counterintuitive (one might expect abundant living space to signal less tight supply), but it likely reflects that historically spacious districts were those with higher-quality housing or greater development, hence higher seller values.

Table 8: Indicators of housing supply

	(1)	(2)	(3)
Dependent variable:	$K_{i,2009}$	$K_{i,2018}$	$K_{i,2018} - K_{i,2009}$
Distance to the center	-0.0055***	-0.0162***	-0.0092***
	(0.0008)	(0.0014)	(0.0008)
Building land prices	0.0007***	0.0007***	0.0009***
	(0.0001)	(0.0001)	(0.0001)
Living area per capita	0.0256***	0.0334***	0.0074
	(0.0044)	(0.0078)	(0.0047)
Completed apartments per capita	0.1390***	0.0580**	0.0876***
	(0.0233)	(0.0191)	(0.0186)
Share of small apartments	0.0200***	0.0345***	0.0090*
	(0.0041)	(0.0062)	(0.0039)
Region effects	Yes	Yes	Yes
N	2,275	2,275	2,275

NOTES: Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

By 2018, the distance to city center effect on the supply component becomes more pronounced. In 2009, more central locations already showed a moderately higher $K_{i,t}$ (the distance coefficient was negative), and in 2018 this effect is larger and highly significant. Moreover, the distance coefficient is significantly negative in the change regression for $K_{i,t}$, indicating that central districts experienced greater increases in the seller valuations between 2009 and 2018. We also note a shift in the role of housing stock composition. In 2009, the share of small apartments in a district was already positively correlated with the supply component. By 2018, however, this relationship had strengthened considerably, with a much larger effect on $K_{i,t}$. In other words, districts with a higher proportion of small apartments—typically dense, urban areas—saw higher seller valuations in 2018. This trend likely reflects demographic changes in Germany, where the rise of smaller households has driven demand for, and hence new construction of, smaller housing units in high-demand urban markets (see Dustmann et al., 2022).

Overall, the regression evidence provides support for the economic content of our estimated structural components. The supply component $K_{i,t}$ clearly responds to local construction intensity and land market conditions, behaving as a proxy for regional housing supply constraints and development. Likewise, the demand component $A_{i,t}$ is firmly grounded in demographic and economic fundamentals of housing demand. Taken together, these findings indicate that our model-implied demand and supply factors correspond closely to real-world fundamentals, bolstering confidence in the model's structural interpretation of Germany's housing boom.

5.5 Spatial Correlation in Housing Demand and Supply

We next examine the spatial dimension of the model's housing demand and supply parameters A_{it} and K_{it} as well as the time-invariant location premia τ_i governing buyer market shares across locations. In particular, we investigate the extent to which differences in these parameters are correlated with the geographic distance. We perform the analysis using only the Top-7 regions which are more homogeneous in terms of population density and in the geographic size and distance of postal code locations. The results confirm that geographic proximity plays a crucial role: locations that are close to each other tend to have similar demand and supply fundamentals, whereas those farther apart exhibit more divergent values.

Table 9 provides quantitative evidence of this spatial dependence. We regress the absolute difference in the model-implied demand or supply parameters between any two locations on the geographic distance separating them (considering location pairs within the same labor market region for comparability). The estimated coefficients on distance are all positive and statistically significant for both the buyer (columns (1)-(3)) and seller (columns (4)-(6)) valuation components, confirming that more distant location pairs tend to have larger differences in their housing market fundamentals. In column (7), we regress the absolute differences in the permanent location premia between any two locations in a region on the separating geographic distance. We find a strong association between distance and the differences in location premia which points out that the model-based location-specific amenities capture the fact that nearby locations tend to have similar amenities partly due to their geographic proximity.

Table 9: Spatial correlation in model-based parameters, Top-7 regions

Dependent variable:	(1) Difference in buyer valuations 2009	(2) 2013	(3) 2018	(4) Difference in seller valuations 2009	(5) 2013	(6) 2018	(7) Difference in location-specific amenities
Distance	0.0584***	0.1080***	0.2070***	0.0622***	0.1100***	0.2110***	0.1410***
	(0.00275)	(0.00385)	(0.00520)	(0.00277)	(0.00385)	(0.00521)	(0.00228)
Constant	0.3960***	0.4760***	0.5290***	0.3920***	0.4730***	0.5220***	0.1350***
	(0.00803)	(0.01120)	(0.01520)	(0.00809)	(0.01120)	(0.01520)	(0.00666)
N	60,282	60,282	60,282	60,282	60,282	60,282	60,282

NOTES: The outcome is the absolute difference between any two locations within a labor market region for buyer valuations (cols 1–3), seller valuations (cols 4–6), and location-specific amenity valuations (col 7). "Distance" refers to log distance (km). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The strength of spatial correlation also increases over time. Between 2009 and 2018, the coefficients on distance roughly triple for both buyer and seller valuations, indicating that spatial differences in demand and supply fundamentals became much more pronounced during the housing boom.

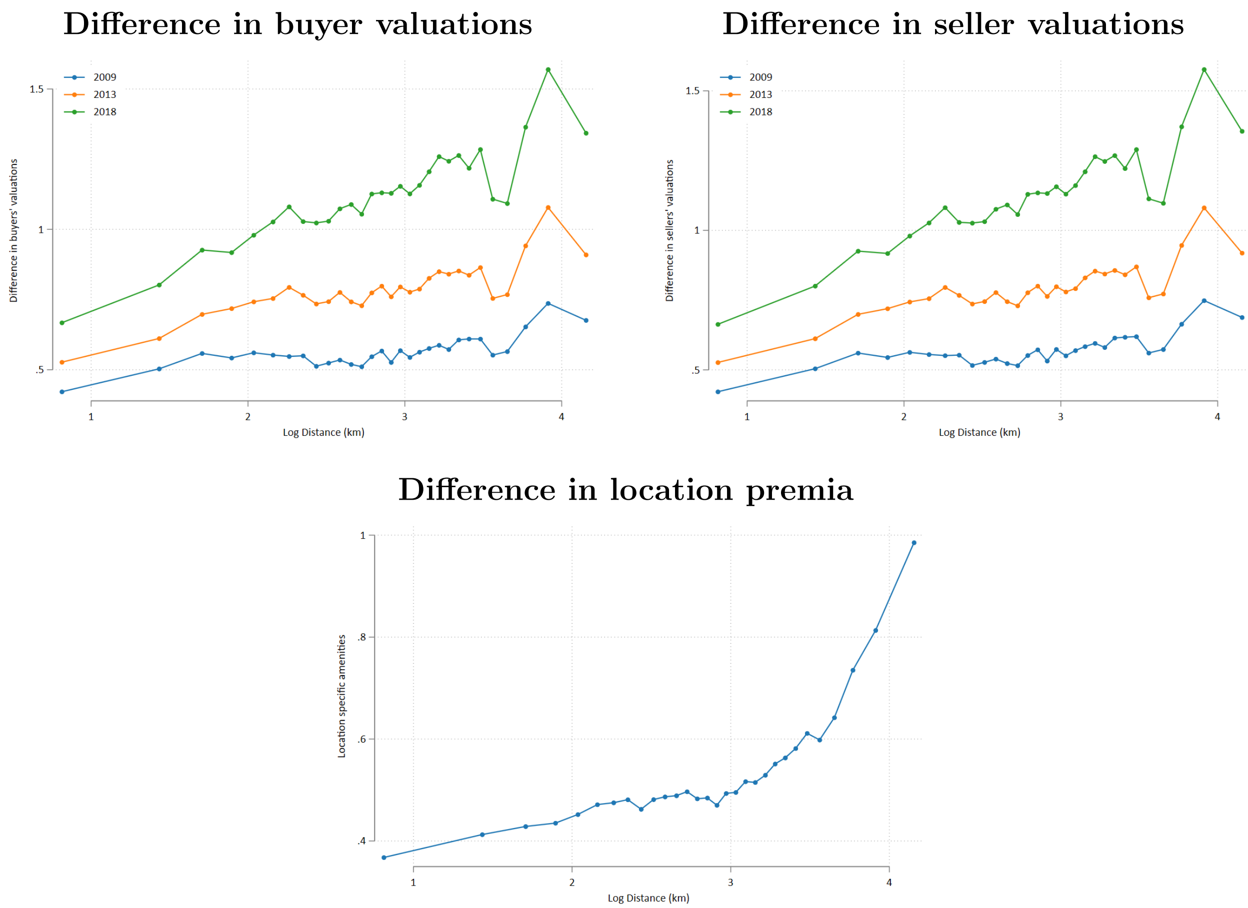


Figure 7: Difference in model-based parameters and geographic distance (Top-7 regions)

NOTES: The x-axis variable is the distance between any two locations within a Top-7 labor market region. The y-axis variables are the absolute differences between any two locations within a labor market region for the buyer valuations, seller valuations and location premia. The figures are binned scatterplots grouped by the x-axis variable.

Figure 7 visualizes these patterns. It plots the average absolute difference in buyer valuations, seller valuations, and the location premia τ_i between pairs of locations against the distance separating those locations (again using data from the Top-7 labor market regions). The relationship is upward-sloping for all three measures, corroborating the finding that greater distance is associated with larger disparities in housing market fundamentals according to the structural model. Notably, the buyer and seller valuation spatial differences increase over time, especially for locations further apart.

5.6 Within-location co-movement of demand and supply valuations

In addition to studying spatial correlation across locations, we also examine whether buyer valuations $A_{i,t}$ and seller valuations $K_{i,t}$ tend to co-move over time within the same location i. For each postal code i, we compute the correlation coefficient $\text{Corr}_t(A_{i,t}, K_{i,t})$ using the panel of quarterly valuations between 2009 and 2018. Figure 8 reports the distribution of these correlations across all postal codes.

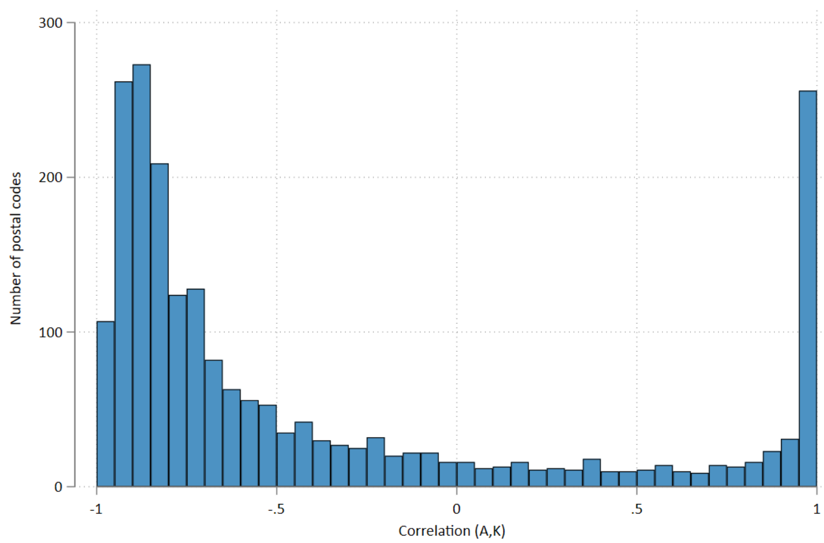


Figure 8: Correlation of A and K over time within postal codes

NOTES: For each postal code i, we compute $\rho_i = \text{Corr}_t(A_{i,t}, K_{i,t})$ across quarters $t \in$ [2009Q1, 2018Q4]. The figure shows the distribution of $\{\rho_i\}$ across all postal codes in the balanced panel.

The results reveal a wide dispersion. While a non-trivial fraction of locations exhibit strong positive correlations, a substantial mass of locations shows negative correlations, where demand and supply valuations move in opposite directions over time. This heterogeneity suggests that local housing markets are subject to diverse forces; In some locations, stronger demand is matched by rising supply valuations possibly reflecting higher opportunity costs of holding housing units while in other locations demand shocks are offset by more muted or even declining seller valuations. Importantly, this heterogeneity aligns with the patterns described in Section 5.5, where spatial correlations across locations have become stronger over time, but within-location dynamics remain heterogeneous.

6 Conclusions

This paper investigates the growing geographic dispersion of house prices in Germany over the course of the housing boom from 2009 to 2018. Using a comprehensive dataset of sales listings, we document a substantial increase in price heterogeneity across locations. Most of this increase stems from rising dispersion between labor market regions, although price differences within regions, particularly within Germany's Top-7 metropolitan areas, have also widened notably.

To better understand the underlying drivers, we propose and calibrate a stylized model of the housing market based on directed search. This framework allows us to decompose observed prices into distinct demand and supply components, and to quantify their respective contributions to rising price dispersion. We find that differential changes in housing demand across postal codes within the Top-7 regions are the main contributors to the increase in house prices and their dispersion in these regions between 2009 and 2018. Housing demand is also the main factor behind the overall rise in between-location price dispersion, while housing supply plays a secondary contribution for between-region price divergence. Search frictions matter for the average price increase and for spatial differences but do not contribute to price divergence over time.

A central innovation of our approach lies in linking model-implied valuations to observed district-level fundamentals. Estimated demand valuations correlate strongly with household income, credit quality, population density, and proximity to city centers. Supply valuations, in turn, align with land prices, completions, and housing stock characteristics.

We also show that the geographic structure of demand and supply has become more pronounced. Spatial correlation patterns reveal that both components vary systematically with geographic distance and that this spatial structuring intensified between 2009 and 2018. These results underscore the growing importance of agglomeration effects, amenities, and spatial frictions in shaping housing market fundamentals.

While our model identifies demand as the dominant source of increasing price dispersion, it remains agnostic about the deeper forces behind these shifts. Several potential drivers, including monetary policy transmission (Gorea et al., 2023), local shocks such as refugee inflows (Hennig, 2021), and the spatial sorting of high-skilled labor into productive urban firms (Dauth et al., 2022), likely contribute to these dynamics. Future research could leverage richer data or alternative modeling strategies to assess the relative importance of these mechanisms.

Statements and Declarations

Funding

The authors received financial support from SAFE (project 21529).

Competing interests

The authors declare that they have no competing financial or non-financial interests.

Data availability

The real estate listings data used in this study are available from the Research Data Centre (FDZ Ruhr) of the RWI – Leibniz Institute for Economic Research, subject to access restrictions.

References

Aguirregabiria, Victor and Pedro Mira (2010), “Dynamic discrete choice structural models: A survey.” *Journal of Econometrics*, 156, 38–67.

Ahlfeldt, Gabriel M, Stephan Heblich, and Tobias Seidel (2023), “Micro-geographic property price and rent indices.” *Regional Science and Urban Economics*, 98, 103836.

Albrecht, James, Pieter A Gautier, and Susan Vroman (2016), “Directed search in the housing market.” *Review of Economic Dynamics*, 19, 218–231.

Amaral, Francisco, Martin Dohmen, Sebastian Kohl, and Moritz Schularick (2024), “Interest rates and the spatial polarization of housing markets.” *American Economic Review: Insights*, 6, 89–104.

Amaral, Francisco, Martin Dohmen, Moritz Schularick, and Jonas Zdrzalek (2023), “German real estate index (GREIX).” ECONtribute Discussion Paper No. 231.

Anenberg, Elliot and Patrick Bayer (2020), “Endogenous sources of volatility in housing markets: The joint buyer–seller problem.” *International Economic Review*, 61, 1195–1228.

Balz, Felix Florian (2024), “User cost of housing analysis of the german real estate market.” *Acta VSFS*, 18, 41–67.

Ben-Shahar, Danny and Roni Golan (2022), “Price dispersion and time-on-market in the housing market.” *Journal of Housing Economics*, 58, 101875.

Borowczyk-Martins, Daniel, Grégory Jolivet, and Fabien Postel-Vinay (2013), “Accounting for endogeneity in matching function estimation.” *Review of Economic Dynamics*, 16, 440–451.

Brausewetter, Lars, Stephan L. Thomsen, and Johannes Trunzer (2024), “Regional supply and demand fundamentals in the German housing price boom.” *German Economic Review*, 25, 1–36.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro (2019), “Trade and labor market dynamics: General equilibrium analysis of the China trade shock.” *Econometrica*, 87, 741–835.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum (2022), “Matching in cities.” *Journal of the European Economic Association*, 20, 1478–1521.

Duranton, Gilles and Diego Puga (2023), “Urban growth and its aggregate implications.” *Econometrica*, 91, 2219–2259.

Dustmann, Christian, Bernd Fitzenberger, and Markus Zimmermann (2022), “Housing expenditure and income inequality.” *The Economic Journal*, 132, 1709–1736.

Fogli, Alessandra, Veronica Guerrieri, Mark Ponder, and Marta Prato (2023), “The end of the American dream? Inequality and segregation in US cities.” Mimeo.

Fujita, Masahisa and Jacques-François Thisse (2002), “Agglomeration and market interaction.” *Available at SSRN 315966.*

Garriga, Carlos and Aaron Hedlund (2020), "Mortgage debt, consumption, and illiquid housing markets in the Great Recession." *American Economic Review*, 110, 1603–1634.

Gehr, Katja and Michael P Pflüger (2023), "The worth of cities in germany." Technical report, IZA Discussion Papers.

Georgi, Sabine and Peter Barkow (2010), "Wohnimmobilien-Indizes: Vergleich Deutschland – Grossbritannien." ZIA Projektbericht, Zentraler Immobilienausschuss, Berlin.

Gorea, Denis, Oleksiy Kryvtsov, and Marianna Kudlyak (2023), "House price responses to monetary policy surprises: Evidence from the U.S. listings data." CEPR Discussion Paper No. 17595.

Groiss, Martin and Nicholas Syrichas (2025), "Monetary policy, property prices and rents: Evidence from local housing markets." *manuscript*.

Guren, Adam M (2018), "House price momentum and strategic complementarity." *Journal of Political Economy*, 126, 1172–1218.

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

Hedlund, Aaron (2016), "Illiquidity and its discontents: Trading delays and foreclosures in the housing market." *Journal of Monetary Economics*, 83, 1–13.

Henderson, J Vernon (1974), "The sizes and types of cities." *The American Economic Review*, 64, 640–656.

Hennig, Jakob (2021), "Neighborhood quality and opposition to immigration: Evidence from German refugee shelters." *Journal of Development Economics*, 150, 102604.

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott (2018), "Tarnishing the golden and empire states: Land-use restrictions and the US economic slowdown." *Journal of Monetary Economics*, 93, 89–109.

Hsieh, Chang-Tai and Enrico Moretti (2019), "Housing constraints and spatial misallocation." *American Economic Journal: Macroeconomics*, 11, 1–39.

Jiang, Erica Xuewei, Nadia Kotova, and Anthony L Zhang (2024), "Liquidity in residential real estate markets." Mimeo.

Kindermann, Fabian, Julia Le Blanc, Monika Piazzesi, and Martin Schneider (2024), "Learning about housing cost: Theory and evidence from the German house price boom." Mimeo.

Kosfeld, Reinhold and Alexander Werner (2012), "Deutsche Arbeitsmarktregionen – Neuabgrenzung nach den Kreisgebietsreformen 2007–2011." *Raumforschung und Raumordnung*, 70, 49–64.

Moen, Espen R (1997), "Competitive search equilibrium." *Journal of Political Economy*, 105, 385–411.

Moen, Espen R, Plamen T Nenov, and Florian Sniekers (2021), "Buying first or selling first in housing markets." *Journal of the European Economic Association*, 19, 38–81.

Ngai, L Rachel and Kevin D Sheedy (2020), "The decision to move house and aggregate housing-market dynamics." *Journal of the European Economic Association*, 18, 2487–2531.

Ngai, L Rachel and Silvana Tenreyro (2014), "Hot and cold seasons in the housing market." *American Economic Review*, 104, 3991–4026.

Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-François Thisse (2002), "Agglomeration and trade revisited." *International economic review*, 409–435.

Rekkas, Marie, Randall Wright, and Yu Zhu (2022), "How well does search theory explain housing prices?" Mimeo.

Roback, Jennifer (1982), "Wages, rents, and the quality of life." *Journal of Political Economy*, 90, 1257–1278.

Rosen, Sherwin (1979), "Wage-based indexes of urban quality of life." In *Current issues in urban economics* (P Mieszkowski and M Straszheim, eds.), 74–104, Johns Hopkins Univ. Press.

Saiz, Albert (2010), "The geographic determinants of housing supply." *Quarterly Journal of Economics*, 125, 1253–1296.

Van Nieuwerburgh, Stijn and Pierre-Olivier Weill (2010), "Why has house price dispersion gone up?" *Review of Economic Studies*, 77, 1567–1606.

Vanhapelto, Tuuli and Thierry Magnac (2024), "Housing search and liquidity in spatial equilibrium." Mimeo.

Wright, Randall, Philipp Kircher, Benoît Julien, and Veronica Guerrieri (2021), "Directed search and competitive search equilibrium: A guided tour." *Journal of Economic Literature*, 59, 90–148.

Appendix

A Data Description

Background. Immobilienscout24 is the largest online real estate listing platform in Germany, catering to real estate providers, owners, tenants, and buyers. Operating in three countries - Germany, Austria, and Spain - the platform and its mobile app collectively attract approximately 20 million visitors per month. As of the end of 2019, Immobilienscout24 boasted around 450 thousand active listings, underscoring its prominent position in the real estate market.

The online portal can be accessed at https://www.immobilienscout24.de. Upon entering the German-language website, users are presented with the interface illustrated in Figure 9. The platform prompts users to select their country, specify the location for their search (city, address, or postal code), indicate the transaction type (buy or rent) and define the property type (house, flat or other types).

Additionally, the platform offers a range of filtering options, allowing users to refine their search by specifying property characteristics beyond geographical constraints. Users have the flexibility to set price ranges by providing a lower bound, an upper bound, or both. Furthermore, there is an option to specify the desired number of rooms.

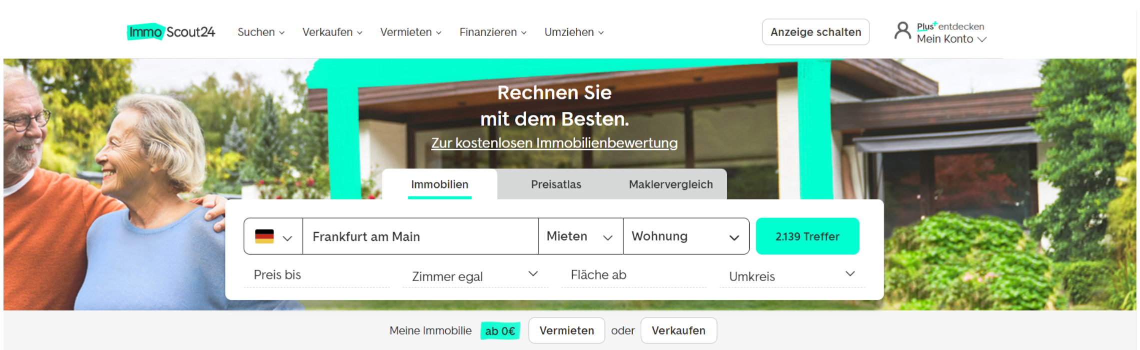


Figure 9: Immobilienscout24 web portal

Dataset. Our analysis relies on version 5.1 of the RWI-GEO-RED dataset, curated by the Research Data Centre (*Forschungsdatenzentrum* or FDZ Ruhr) at the Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen), covering the period from January 2007 to July 2021. The dataset comprises listings of residential properties on the Immobilien-

Scout24 website across Germany, categorized into four classes: house sales, flat sales, house rents and flat rents.[22]

In addition to listed prices and rents, buyer contacts and the duration of the listings, the dataset incorporates user-contributed information that influences the valuation and location of each listing. Users provide details about their listings through a guided online questionnaire, subsequently transforming their input into an advertisement on the ImmobilienScout24 website. While essential information such as location, price (rent), and space of the listed property is mandatory, the remaining questionnaire fields are optional. There are a total of 76 distinct entries available for users to provide information, categorized into eight groups by RWI Essen.

Locations. ImmobilienScout24 does not provide the address of the offered real estate. Instead, they geo-code addresses when available according to their own Mercator.[23] In turn, the RWI Essen converts the projected locations into the European standard ETRS89-LAEA based on INSPIRE which is a grid of 1-km^2 raster cells covering the whole territory of Germany. Subsequently, the grids are then assigned to broader administrative regions, in particular postal codes, municipalities, districts or local labor market regions. This is done based on the 2015 geographical shapefiles provided by the Federal Agency for Cartography and Geodesy.

To compare the geographic house price/rent dispersion across time, we pool the housing units together in terms of postal codes. We choose postal codes rather than 1-km^2 cells because the former are sufficiently large to contain enough housing units but also small enough to exhibit spatial heterogeneity within city boundaries. The highest level of geographical aggregation we use is the labor market regions categorized by Kosfeld and Werner (2012). Labor market regions combine one or more districts and are characterized according to the commuter links to local labor centres.

Basic cleaning: We allow for a two-year burn-in period at the beginning and end of the sample. This allows us to properly identify new listings and to also exclude the possibility of active listings at the end of the sample. To this end, we include in our dataset all listings that appear on the Immobilienscout24 platform between January 1, 2009 and December 31, 2018. Then we erase multiple entries that correspond to the same property within a short window.[24] In particular, we only keep the last price and we drop all previous listings for the

[22]ImmobilienScout24 claims a market share of approximately 50 percent of all advertised real estate objects in Germany (Georgi and Barkow, 2010).

[23]In the initial years covered by the dataset, it was not mandatory for users to provide the address of the real estate. They could show only urban districts or municipalities for public use. Only for the most recent years, it is obligatory to provide the property address in the offer.

[24]According to the RWI-GEO-RED data manual duplicate entries occur for several reasons: *"First, since we obtain spells that have not been concluded at the time of data delivery, these will also occur in the next*

same item if it was posted more than once within a six-month period. We treat spells with starting dates at least six months from each other as different postings.[25] Second, we drop properties with missing mandatory information such as the geo-coded location, number of rooms, size or the age of the property. We also drop properties classified as "castles" or properties built before the year 1900. Finally, we remove all postings listed for less than a day.

Censoring. We exclude all postings with unreasonable price/rent entries. These entries include ultra-luxurious properties that form a market of their own and are likely to contaminate our analysis. We drop all units with a sale price of more than €6,000,000 or a rental price that exceeds €6000 per month. On the other hand, under-market value properties might be indicative of fraudulent listings or an attempt of the sellers to manipulate in their favor the Immobilienscout24 listing algorithm. This can happen only in the case the potential buyers list the property by price/rent in ascending order.[26] We remove all listings with a sale price of less than €10000 and a rental price of less than €130.

Moreover, we censor the price of a property per m^2. House and flats for sales are censored between €150 and €20000 per m^2 and rental units between €2.5 and €25 per m^2. The living area is restricted between 25 and 400 m^2 for flats and between 45 and 800 m^2 for houses. On top, we omit flats with more than 8 rooms and houses with more than 15 rooms. Finally, we drop all properties where the number of contacts or the number of clicks is beyond the 99-th percentile. Lastly, we drop listings with a duration longer than the 99-th percentile separately for sale and rental houses and flats.

Finally, we restrict the dataset to postal codes that contain at least 10 postings within a quarter and labor market regions that contain at least 14 postal codes. We run this procedure separately for the rental and sales market.

Inflation adjustments. The house prices and rents in our dataset are in nominal terms. We compute the inflation-adjusted prices and rents by deflating the nominal values with the

delivery which continues from the time of the previous delivery. Moreover, users can make small changes to the advertisement in order to attract more people. In the data, we only observe the status of the advertisement at the time of data delivery. Hence, the same advertisement might appear twice but with slightly different features in the data when a change was made after the delivery date. Fourth, users can temporarily set an object as inactive. This may be reasonable when a prospective buyer has committed to buying an object, but the deal has not yet been finalized. While inactive, objects will not be included in queries of potential buyers and will thus not be included in the dataset. However, if the potential buyer withdraws their offer to buy, the user might decide to activate the advertisement again. Lastly, users might decide to use an old advertisement as a template for a new ad, e.g. when renting two similar flats in the same house with only a short period in between."

[25] RWI Essen has developed an automatized procedure to identify multiple entries at the same time.

[26] An example of this is properties listed with very low rent but then much higher than normal utilities.

respective state-specific consumer price index at the monthly level obtained from the Federal Statistical Office.

Location adjustments. The vast majority of geo-code coordinates and their respective administrative match are consistent but some challenges remain. First, some administrative districts have been merged or changed over time. To address this problem, we obtain from https://www.geodaten-deutschland.de a 2019 file that contains up-to-date geo-referenced administrative information.

Several districts have changed names or were merged into a different district in 2011. Table 10 shows the mapping from these changed 2011 districts to their 2015 versions.

Table 10: Changes of districts, 2011-2015

2011 District	2011 District Number	2015 District	2015 District Number
SK Aachen and LK Aachen	5313, 5354	Städteregion Aachen	5334
Nordvorpommern	13107	Vorpommern-Rügen	13073
Südvorpommern	13108	Vorpommern-Greifswald	13075
Bremerhaven	4021	Bremerhaven, Stadt	4012
Rostock	13101	Rostock	13003
Mittleres Mecklenburg	13104	Landkreis Rostock	13072
Mecklenburgische Seenplatte	13103	Mecklenburgische Seenplatte	13071
Nordwestmecklenburg	13106	Nordwestmecklenburg	13074
Schwerin	13102	Schwerin	13004
Südwestmecklenburg	13105	Ludwigslust-Parchim	13076

Finally, we drop listings without information regarding the postal code (0.2 percent of all listings). For the remaining listings, we matched the postal code and the municipality of the RWI Essen dataset with the https://www.geodaten-deutschland.de updated dataset.[27] Around 98 percent of the listings match perfectly in both dimensions. All the unmatched entries are dropped.

Listings over time and space. Table 11 shows the numbers of listings of our baseline dataset across the years for each of the four property classes. Figure 10 presents the number of listings across districts in Germany.

[27] One might expect that the postal code areas are coherent and disjoint. However, this is not the case. There are postal code areas where one area lies entirely inside another area (e.g. 53879 in Euskirchen is enclosed by 53881). There are even cases where an area contains more than one other area.

Table 11: Number of listings over time, 2009-2018

	House sales	Flat sales	House rents	Flat rents
2009	337,837	291,180	27,830	457,210
2010	324,249	281,170	27,651	487,036
2011	306,922	285,790	26,081	469,685
2012	298,577	306,469	28,720	456,756
2013	290,145	328,521	30,958	481,183
2014	286,395	361,004	31,350	626,024
2015	271,651	294,087	22,256	531,284
2016	211,567	224,835	16,397	421,648
2017	207,644	202,533	15,717	370,768
2018	189,633	187,725	15,087	356,733
N	2,724,620	2,763,314	242,047	4,658,327

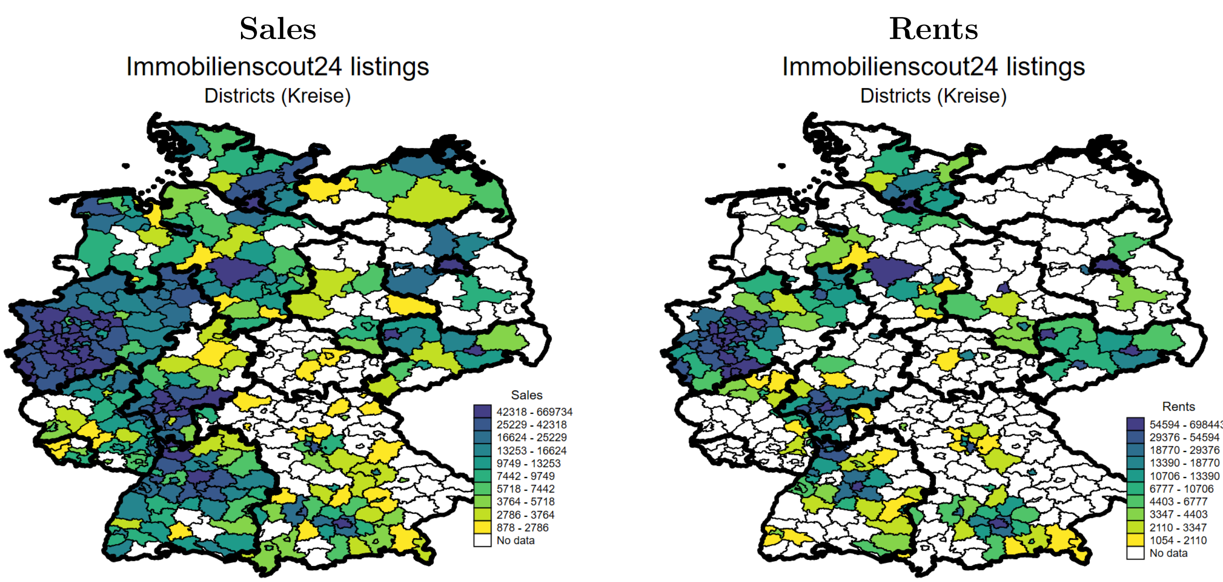


Figure 10: Listings across space, 2009-2018

B Hedonic Regressions

In the hedonic regression (1), we control for observed variable characteristics that determine the quality of the listed property. We have at our disposal a set of 76 variables many of which contain missing entries or are sparsely filled.

We divide the nominal listed price by the monthly CPI and then by the property area to create inflation-adjusted property price per m^2, which is the dependent variable. Second, we use the following explanatory variables:

- **Number of rooms.** In Germany, the number of rooms excludes kitchens, baths, or corridors. In several cases, the number of rooms is not a natural number, which is not necessarily due to a faulty entry. In Germany, there is the concept of half rooms. Following the DIN 283 norm, a half room is defined as a room with a size between 6 and 10 m^2. While this definition is outdated, it is still frequently in use. In these cases, we round up to the nearest natural integer. Then we created 14 separate dummies (excluding properties with 1 room).

- **Age of the property.** We deduct the year the listing was posted from the year it was built. Then we create 5-year age dummies. On several occasions the seller lists the price before the property is constructed. We include these entries in the first age category.

- **Type of property.** We control for 22 detailed types of property: Not specified house, Single-family house (detached), Two-family house, Semi-detached house, Terraced house, Terraced house (middle unit), Terraced house (end unit), Bungalow, Farmhouse, Mansion, Block of flats, Other property for living, Special property, Attic flat, Flat, Raised ground floor flat, Maisonette, Penthouse, Souterrain, Flat with terrace, Other Flat, and Not specified flat.

- **Cellar.** A dummy variable which indicates that the property has a cellar.

- **Guest toilet.** A dummy variable which indicates that the property has a guest toilet.

- **Quarterly dummies.** A set of dummies indicating the quarter the ad was listed.

C Comparison with Transaction Prices

A general concern related to listings data is the lack of transaction prices and information about whether or not a listing resulted in an actual sale or rent. If final transaction prices differ systematically from listing prices, the findings of our paper could potentially be biased. To deal with this issue, we compare our dataset with a transaction-based dataset from an alternative source. We find that levels and trends of these prices, aggregated at the city level, are broadly comparable.

German Real Estate Index (GREIX). A recent study by Amaral et al. (2023) compiles and disseminates quarterly transaction-level real estate data for 18 cities and their neighborhoods in Germany. The German Real Estate Index (GREIX) is based on this work. The raw micro data are collected from historical notarial archives and are then processed and aggregated at the city level across market segments (flats, single-family houses and multi-family houses).[28]

We compare the transaction-based data from the project with our listings data. Specifically, for every city in their data, we retrieve the average nominal price per square meter from inflation-unadjusted data, separately for flats and single-family houses. We exclude multi-family houses from our comparison due to the challenges in reconciling this market segment in GREIX with the multi-family units in our data.

In this exercise, we use our raw Immobilienscout24 listings data and apply the same cleaning procedure as Amaral et al. (2023). The goal is to make the two datasets comparable and limit any discrepancy that might arise due to the fact that our cleaning process is more elaborate and restrictive.

Flats. For this comparison, we use the raw data which contains all sale listings for flats in Immobilienscout24. Following the documentation of Amaral et al. (2023), we first remove the listings containing missing prices or living area for each year. Properties already listed on the market but with construction date three years or longer in the future are excluded. Additionally, we winsorize the data at the 1st and 99th percentiles of purchase price and living area in order to remove outliers. We also remove duplicate entries using flats IDs, keeping only the last listed record with identical price and features within a close time frame. Lastly, any repeated entries for the same property within a short period that show price discrepancies are also removed.

Single-family houses. We use the raw data which contains all sale listings for houses in Immobilienscout24. Then, we restrict the data to the following house types: single-family house (detached), single-family house, and semi-detached house. We also use listings with

[28] For more information about the data and access, see https://greix.de.

missing entries into the house type variable (9 percent of all listings) as it is likely that the vast majority of these entries may be single-family houses.[29] Further, we impose the same restrictions as in the case of flats.

Figures 11 and 12 show the time series of the average prices per squared meter of flats and single-family houses, both for the listings data and for the transaction data for all cities covered by GREIX. While there are some deviations, the levels and trends are rather similar.

[29]We also replicate our analysis excluding missing entries and find that the results appear almost identical.

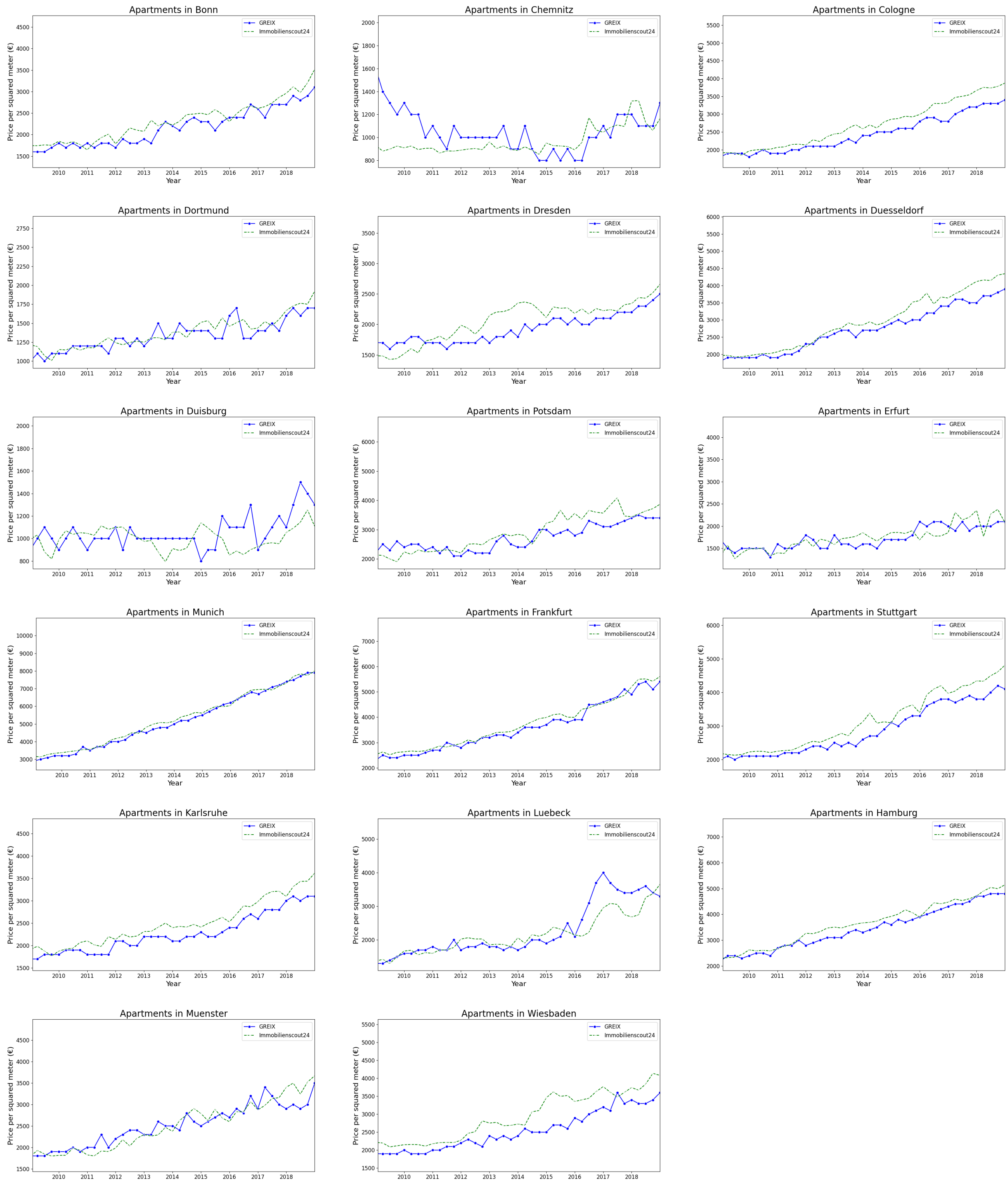


Figure 11: Flats sales prices - transactions vs listings

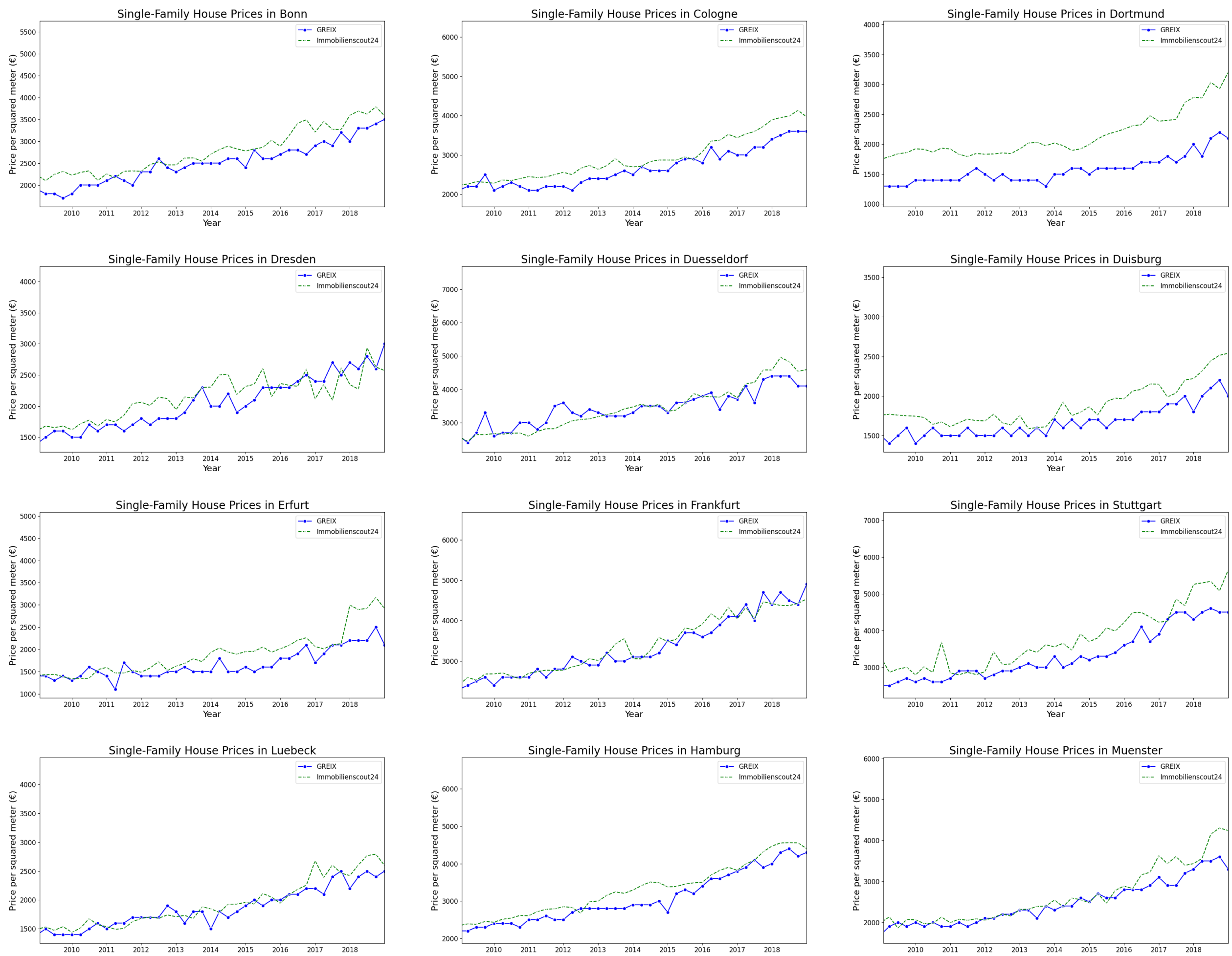


Figure 12: Single-family house sales prices - transactions vs listings

D Additional Results

D.1 Controlling for Distance to City Center

D.1.1 Methodology

The first step is to determine the economic center of each of the labor market regions in our dataset. Many labor market regions comprise multiple districts (*Kreise*). Following the approach of Kosfeld and Werner (2012), we select the *Kreis* that best represents the center of the labor market region. For example:

- The **Berlin** labor market region includes *Berlin, Potsdam, Barnim, and Dahme-Spreewald*, but it is represented by *Berlin* as its central point.
- The **Mainz** labor market region consists of *Wiesbaden, Mainz, Alzey-Worms, and Mainz-Bingen*, and it is referred to as *Mainz* based on largest city.

We use an automated Python script to determine the center of each labor market region. Following Duranton and Puga (2023) for the U.S. and Gehr and Pflüger (2023) for Germany, we determine the city center as the location returned by a Google Maps search of the city's name augmented with the term "Stadtzentrum" (city center). We then extract its latitude and longitude coordinates. In a few cases where Google Maps search does not produce credible coordinates to the actual city center, we adjusted the coordinates manually.

In the next step, we utilize the GRID dataset, which provides geolocations of all square-km grid cells in Germany. These grids are referenced using the INSPIRE geospatial framework, which employs coordinates in EPSG:3035 (ETRS89-LAEA). We transform these coordinates into latitude and longitude in EPSG:4326 (WGS84).

Following Gehr and Pflüger (2023), we compute the haversine distance to measure the distance between the midpoint of each grid cell and the identified center of the district region. We complement the analysis with an alternative distance measure, namely the driving distance between each housing grid and the center of the city using OpenStreetMap. For confidentiality reasons, RWI censors grid cells with an insufficient number of resident households. Consequently, approximately an extra 8 percent of housing units for sale are dropped from the following analysis.

Figure 13 shows histograms of both distance measures across all housing units in our data. We include a cubic function of each of the two distance measures in the hedonic regressions of Section 2 and report the results of our main variance decompositions (see Tables 2 and 3) in the next subsections.

(a) Distance in km (b) Driving time in minutes

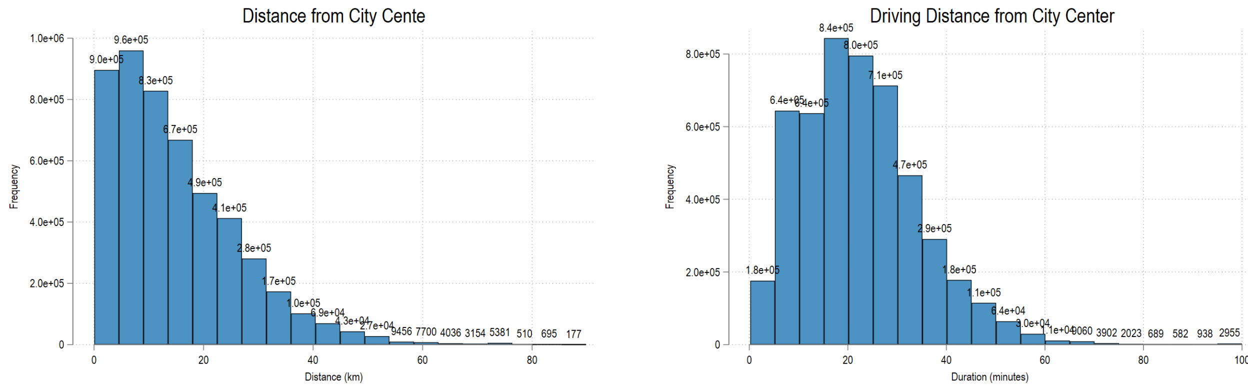


Figure 13: Histograms of two measures of distance to city center

D.1.2 Controlling for Distance in Kilometers

Tables 12 (between/within location) and 13 (between/within region) show the variance decompositions of hedonic prices when distance to city center in *km* is controlled for. Table 12 shows that our first main result is that the increase in price dispersion is almost entirely explained by the divergence of prices between locations. Controlling for distance to city center hardly changes these conclusions. Without distance controls, the between-location variance goes up by about ten percentage points between 2009 and 2018 for the full sample, while the increase is nine percentage points when the distance controls are applied. When restricting the sample to the Top-7 regions, we obtain a similar result, though controlling for distance to city center has a bigger effect: The between-location variance increases by only five percentage points, while it goes up by seven percentage points when we consider the unconditional hedonic prices in Table 2.

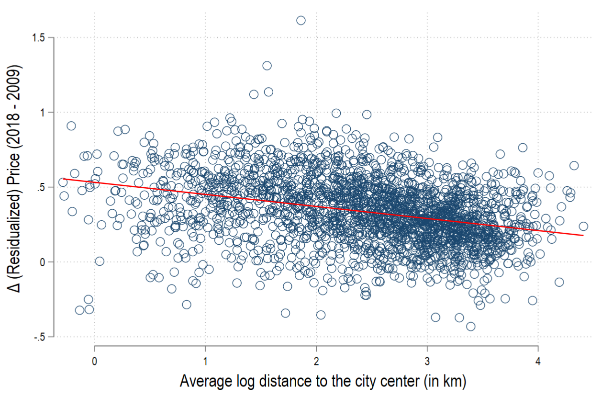
Looking at the between- vs. within-region decomposition of Table 13, we find that the contribution of the between-region component is hardly affected, which is of course little surprising. In line with the above results, the increase of the within-region variance is more muted when distance controls are applied: The within-region variance increases by 1.3 percentage points for the full sample (2.2 percentage points without controls) and by 1.7 percentage points for the Top-7 sample with distance controls (2.9 percentage points for the full sample). We conclude that controlling for distance reduces the rise of between-location variance, because less expensive locations in 2009 located further away from the center have seen a weaker price increase compared to locations closer to the center. We confirm this claim in Figure 14 which shows that house price growth during 2009–2018 is lower in those postal codes which are located further away from the center of the labor market region, both for the full sample (panel a) and for the Top-7 subsample (panel c).

Table 12: Within- and between-location variance decomposition (with distance in km controls)

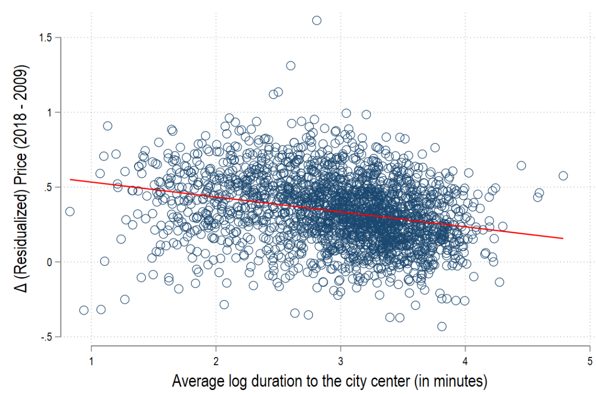
	Total Variance			Within Locations			Between Locations		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full Sample	0.180	0.222	0.267	0.104	0.103	0.100	0.076	0.119	0.167
West Germany	0.175	0.217	0.257	0.103	0.106	0.096	0.072	0.116	0.161
East Germany	0.194	0.237	0.282	0.126	0.126	0.155	0.067	0.111	0.128
Top-7 Regions	0.160	0.171	0.190	0.104	0.092	0.081	0.056	0.080	0.109
Urban	0.182	0.230	0.274	0.106	0.104	0.098	0.076	0.127	0.175
Rural	0.177	0.199	0.251	0.100	0.101	0.104	0.077	0.098	0.147

Notes: See the notes to Table 2 for definitions of the different samples.

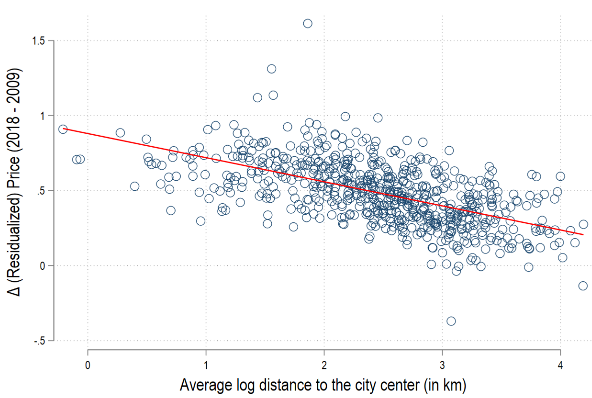
(a) Full sample (distance in km)



(b) Full sample (driving time)



(c) Top-7 regions (distance in km)



(d) Top-7 regions (driving time)

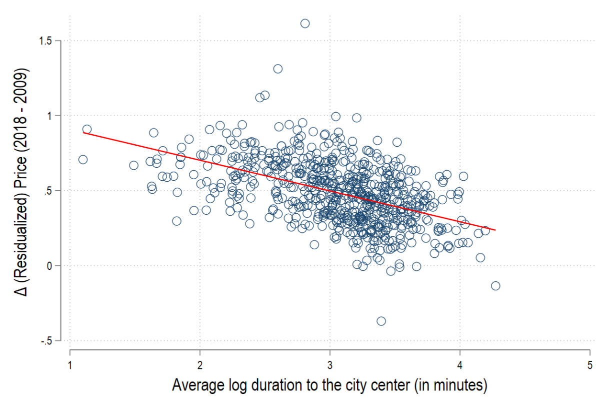


Figure 14: House price growth at the postal code level against distance to the labor market region

Table 13: Within- and between-region variance decomposition (with distance in km controls)

	Between Locations			Within Regions			Between Regions		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full Sample	0.076	0.119	0.167	0.030	0.042	0.043	0.046	0.077	0.124
West Germany	0.072	0.116	0.161	0.030	0.042	0.043	0.042	0.074	0.118
East Germany	0.067	0.111	0.128	0.038	0.051	0.042	0.029	0.059	0.086
Top-7 Regions	0.056	0.080	0.109	0.033	0.046	0.050	0.023	0.034	0.058
Urban	0.076	0.127	0.175	0.033	0.047	0.046	0.042	0.080	0.129
Rural	0.077	0.098	0.147	0.018	0.023	0.026	0.059	0.075	0.122

NOTES: See the notes to Table 2 for definitions of the different samples.

D.1.3 Controlling for Distance in Driving Time

Tables 14 and 15 show the variance decompositions (between and within location, and between and within region) for hedonic prices when distance to city center in driving time (minutes) is controlled for. As above, we find that controlling for distance to city center (here in driving time) has little impact on the main findings regarding the increase of the between-location variance. With distance controls, the between-location variance goes up by nine percentage points between 2009 and 2018 for the full sample (compared to ten percentage points when no distance controls are applied). For the sample of the Top-7 regions, the between-location variance increases by over nine percentage points (seven percentage points without distance controls).

Table 14: Within- and between-location variance decomposition (with driving time controls)

	Total Variance			**Within Locations**			**Between Locations**		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full Sample	0.184	0.227	0.274	0.107	0.106	0.102	0.077	0.121	0.172
West Germany	0.179	0.224	0.265	0.106	0.104	0.098	0.073	0.119	0.167
East Germany	0.194	0.240	0.288	0.128	0.128	0.156	0.066	0.112	0.131
Top-7 Regions	0.166	0.180	0.203	0.106	0.094	0.083	0.060	0.086	0.120
Urban	0.207	0.252	0.312	0.110	0.100	0.090	0.098	0.152	0.221
Rural	0.180	0.207	0.258	0.102	0.103	0.105	0.078	0.104	0.153

NOTES: See the notes to Table 2 for definitions of the different samples.

Regarding the contribution of the between- and within-region variance in Table 15, the results are rather similar to those where distance in *km* is controlled for. In line with the above results, the increase of the within-region variance is more muted when the distance controls are applied: The within-region variance increases by 1.4 percentage points for the full sample (2.2 percentage points without controls) and by 2.0 percentage points for the Top-7 sample with controls (2.9 percentage points for the full sample). We conclude again that controlling for distance reduces the rise of between-location variance. As above, this is because locations further away from the center have seen a weaker price increase compared to locations closer to the center; see Figure 14 which shows that house price growth during 2009–2018 is lower in those postal codes which have longer driving times to the center of the labor market region, both for the full sample (panel b) and for the Top-7 subsample (panel d).

Table 15: Within- and between-region variance decomposition (with driving time controls)

	Between Locations			**Within Regions**			**Between Regions**		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full Sample	0.077	0.121	0.172	0.029	0.042	0.043	0.047	0.079	0.129
West Germany	0.073	0.119	0.167	0.029	0.042	0.044	0.044	0.078	0.123
East Germany	0.066	0.112	0.131	0.036	0.051	0.042	0.029	0.061	0.089
Top-7 Regions	0.060	0.086	0.120	0.035	0.048	0.055	0.025	0.038	0.066
Urban	0.098	0.152	0.221	0.038	0.046	0.046	0.060	0.106	0.176
Rural	0.078	0.104	0.153	0.017	0.025	0.027	0.061	0.080	0.126

NOTES: See the notes to Table 2 for definitions of the different samples.

D.2 Controlling for distance to major urban centres

A potential concern with our distance variable is that many of the 99 German market regions do not contain a sizeable urban city. To ensure that our distance measures are meaningful, we recompute the distance measure using only the largest 30 German cities listed in Table 16. This restriction focuses the analysis on economically relevant urban centers.

Table 16: City Center coordinates of the largest 30 German cities by population

No.	Region	Latitude	Longitude	City	Population
1	Berlin	52.52001	13.40495	Berlin	3 596 999
2	Hamburg	53.55000	10.00000	Hamburg	1 964 021
3	München	48.13750	11.57500	München	1 505 005
4	Köln	50.93753	6.96028	Köln	1 024 408
5	Frankfurt	50.11092	8.68213	Frankfurt	749 596
6	Düsseldorf	51.22556	6.77667	Düsseldorf	616 319
7	Stuttgart	48.77585	9.18293	Stuttgart	613 111
8	Leipzig	51.33970	12.37307	Leipzig	608 013
9	Dortmund	51.51359	7.46530	Dortmund	601 343
10	Bremen	53.07583	8.80722	Bremen	586 271
11	Essen	51.45628	7.01052	Essen	574 082
12	Dresden	51.05041	13.73726	Dresden	563 019
13	Nürnberg	49.45237	11.07680	Nürnberg	529 508
14	Hannover	52.36670	9.71670	Hannover	520 290
15	Wuppertal	51.25719	7.14961	Wuppertal	358 592
16	Bochum	51.48184	7.21624	Bochum	357 024
17	Bielefeld	52.02073	8.53206	Bielefeld	331 519
18	Bonn	50.73390	7.09980	Bonn	321 680
19	Ludwigshafen	49.48778	8.46611	Mannheim	316 256
20	Karlsruhe	49.00689	8.40365	Karlsruhe	308 197
21	Münster	51.96250	7.62556	Münster	307 071
22	Augsburg	48.37054	10.89779	Augsburg	301 105
23	Aachen	50.77535	6.08389	Aachen	263 772
24	Braunschweig	52.26417	10.52167	Braunschweig	253 527
25	Kiel	54.32333	10.13944	Kiel	251 751
26	Chemnitz	50.83472	12.92139	Chemnitz	245 150
27	Magdeburg	52.12861	11.63528	Magdeburg	244 329
28	Freiburg	47.99901	7.84210	Freiburg i. Br.	236 236
29	Halle	51.48281	11.96624	Halle	226 767
30	Mainz	49.99861	8.26944	Mainz	222 889

Population figures are taken from the corresponding Wikipedia pages for each city. Region refers to the labour market district in which the city lies.

We follow exactly the same procedure of subsections D.1.2 and D.1.3 and compute for each listing the distance in km and driving times to the nearest largest city from the Table 16.

Figure 15 plots the histograms of the two distance measures for all housing units in our sample. As expected, the distributions in Figure 15 exhibit fatter right tails than those in Figure 13. The maximum distance to the closest labour market region is roughly 200km or about a two hour car drive and is driven by the sparsely populated Nordvorpommern region north east of Berlin. However, since many of these areas are very sparsely populated, they

(a) Distance in km (b) Driving time in minutes

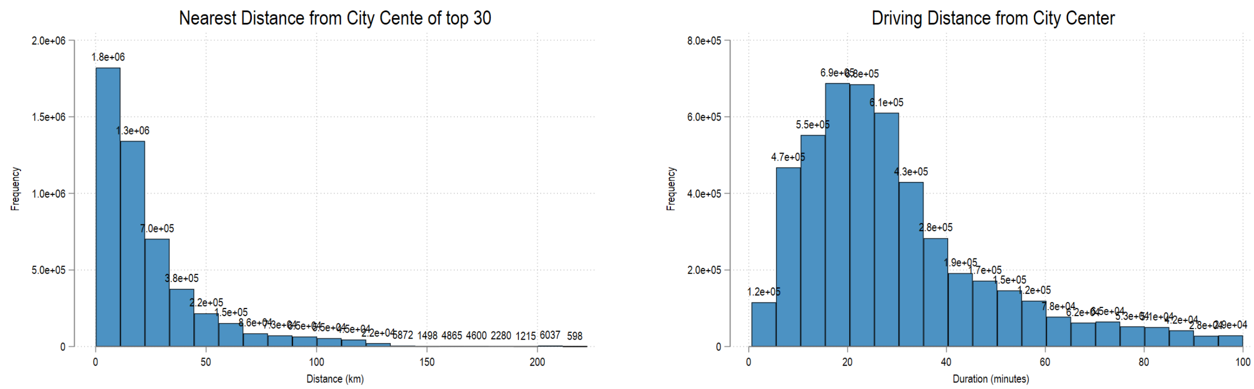


Figure 15: Histograms of two measures of distance to nearest city center of the largest 30 cities

are not expected to change our results significantly. Indeed, under this robustness check, Table 17 and Table 19 deliver the main message, albeit dispersion is more contained than in the previous exercises

Table 17: Within- and between-location variance decomposition (with distance in top 30 km controls)

	Total Variance			Within Locations			Between Locations		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full sample	0.171	0.214	0.253	0.101	0.099	0.094	0.070	0.115	0.160
West Germany	0.167	0.209	0.241	0.100	0.098	0.091	0.067	0.111	0.150
East Germany	0.150	0.182	0.198	0.123	0.121	0.145	0.026	0.061	0.054
Top-7 regions	0.156	0.169	0.187	0.100	0.089	0.078	0.056	0.080	0.109
Urban	0.175	0.226	0.263	0.104	0.101	0.096	0.072	0.125	0.168
Rural	0.151	0.166	0.195	0.095	0.090	0.085	0.056	0.077	0.111

NOTES: See the notes to Table 2 for definitions of the different samples.

Table 18: Within- and between-region variance decomposition (with distance in km 30 controls)

	Between locations			Within regions			Between regions		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full sample	0.070	0.115	0.160	0.028	0.042	0.042	0.042	0.073	0.118
West Germany	0.067	0.111	0.150	0.028	0.042	0.043	0.038	0.069	0.107
East Germany	0.026	0.061	0.054	0.025	0.037	0.029	0.001	0.024	0.025
Top-7 regions	0.056	0.080	0.109	0.032	0.045	0.049	0.023	0.035	0.060
Urban	0.072	0.125	0.168	0.032	0.047	0.046	0.040	0.079	0.122
Rural	0.056	0.077	0.111	0.016	0.024	0.025	0.040	0.053	0.085

NOTES: See the notes to Table 2 for definitions of the different samples.

Table 19: Within- and between-location variance decomposition (driving distance to nearest top-30 city)

	Total variance			Within locations			Between locations		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full sample	0.174	0.218	0.259	0.102	0.100	0.095	0.072	0.118	0.165
West Germany	0.170	0.214	0.248	0.101	0.099	0.092	0.069	0.115	0.156
East Germany	0.153	0.190	0.204	0.124	0.121	0.145	0.028	0.069	0.058
Top-7 regions	0.160	0.174	0.200	0.101	0.090	0.079	0.059	0.084	0.121
Urban	0.180	0.233	0.272	0.105	0.102	0.097	0.075	0.130	0.175
Rural	0.152	0.169	0.198	0.096	0.091	0.086	0.056	0.078	0.112

NOTES: See the notes to Table 2 for definitions of the different samples.

Table 20: Within- and between-region variance decomposition (driving distance to nearest top-30 city)

	Between locations			Within regions			Between regions		
	2009	**2013**	**2018**	**2009**	**2013**	**2018**	**2009**	**2013**	**2018**
Full sample	0.072	0.118	0.165	0.029	0.043	0.045	0.043	0.075	0.120
West Germany	0.069	0.115	0.156	0.029	0.043	0.046	0.040	0.072	0.110
East Germany	0.028	0.069	0.058	0.024	0.036	0.029	0.004	0.033	0.030
Top-7 regions	0.059	0.084	0.121	0.034	0.046	0.055	0.025	0.038	0.066
Urban	0.075	0.130	0.175	0.032	0.047	0.048	0.042	0.083	0.127
Rural	0.056	0.078	0.112	0.017	0.024	0.026	0.040	0.054	0.086

NOTES: See the notes to Table 2 for definitions of the different samples.

D.3 Rental Market

Table 21: Descriptive statistics for Germany, rents

	2009-10	2011-12	2013-14	2015-16	2017-18
Log rent ln r	1.89	1.91	1.94	1.99	2.07
Rent residual ε	-0.04	-0.03	-0.01	0.03	0.10
Listings S	74	73	87	73	56
Duration in days d	32	27	25	23	22
Contacts C	484	628	945	1,254	1,490
Flow tightness $\frac{C}{dS}$	0.30	0.47	0.66	1.30	2.02
Observations	13,520	13,520	13,520	13,520	13,520

NOTES: Means of selected variables for the baseline sample of location-quarter observations. Rents are in euros and adjusted for inflation using the CPI of the federal states in Germany.

Table 22: Descriptive statistics for Top-7 regions, rents

	2009-10	2011-12	2013-14	2015-16	2017-18
Log rent ln r	2.05	2.08	2.13	2.18	2.28
Rent residual ε	0.09	0.12	0.15	0.20	0.28
Listings S	92	86	96	71	51
Duration in days d	28	24	22	20	18.73
Contacts C	744	969	1,389	1,718	1,898
Flow tightness $\frac{C}{dS}$	0.43	0.68	0.95	1.83	2.75
Observations	5,888	5,888	5,888	5,888	5,888

NOTES: Means of selected variables for the baseline sample of location-quarter observations. Rents are in euros and adjusted for inflation using the CPI of the federal states in Germany.

D.4 Further Results for Top-7 Regions

Table 25 shows estimates of the time dummies in the matching function regression (11). Figure 17 shows time series of the variance of house prices, separate for each of the Top-7 labor market regions.

Table 23: Within- and between-location variance decomposition, rents

	Total variance			Within locations			Between locations		
	2009	2013	2018	2009	2013	2018	2009	2013	2018
Full Sample	0.088	0.093	0.106	0.030	0.034	0.035	0.058	0.059	0.071
West Germany	0.085	0.090	0.104	0.031	0.034	0.037	0.054	0.056	0.068
East Germany	0.033	0.042	0.043	0.025	0.027	0.025	0.008	0.016	0.018
Top-7 Regions	0.093	0.084	0.098	0.032	0.037	0.040	0.061	0.047	0.058
Urban	0.091	0.098	0.107	0.030	0.034	0.034	0.062	0.064	0.072
Rural	0.063	0.066	0.098	0.032	0.032	0.038	0.031	0.034	0.060

NOTES: See the notes to Table 2 for definitions of the different samples.

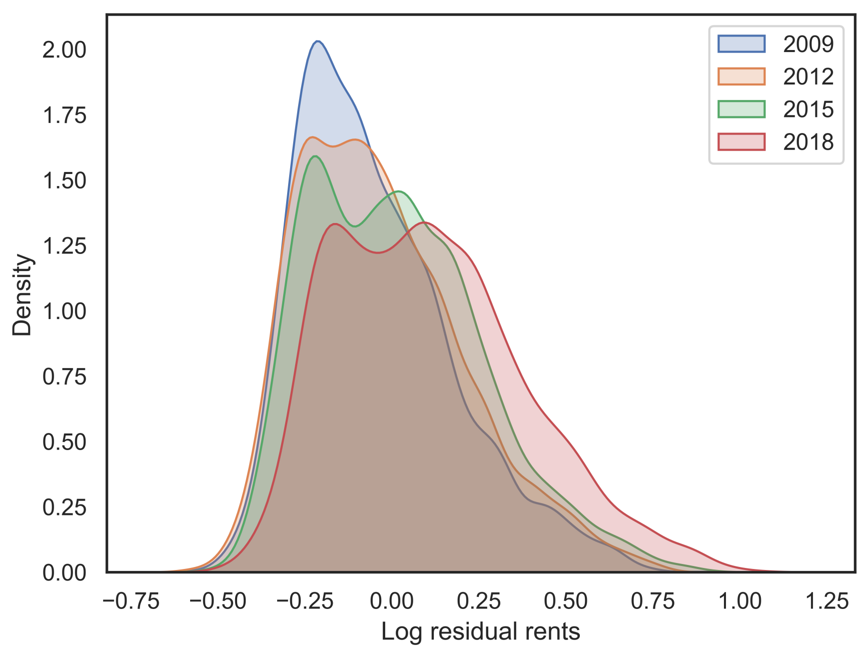


Figure 16: Distribution of residual rents across locations

NOTES: Between-location distributions of residual log rents in the years 2009 (blue), 2012 (orange), 2015 (green) and 2018 (red). The residuals are obtained from hedonic regressions of posted rents per m^2 and averaged in each location (postal code).

Table 24: Within- and between region variance decomposition, rents

	Between-location variance			Within regions			Between regions		
	2009	2013	2018	2009	2013	2018	2009	2013	2018
Full Sample	0.058	0.059	0.071	0.018	0.021	0.021	0.040	0.039	0.050
West Germany	0.054	0.056	0.068	0.019	0.022	0.024	0.035	0.034	0.044
East Germany	0.008	0.016	0.018	0.006	0.010	0.006	0.003	0.006	0.012
Top-7 Regions	0.061	0.047	0.058	0.025	0.031	0.037	0.036	0.017	0.022
Urban	0.062	0.064	0.072	0.016	0.019	0.019	0.046	0.045	0.054
Rural	0.031	0.034	0.060	0.009	0.010	0.011	0.022	0.024	0.049

NOTES: See the notes to Table 2 for definitions of the different samples.

Table 25: Estimates of time fixed effects in equation (11)

t	Berlin	Munich	Hamburg	Frankfurt	Stuttgart	Dusseldorf	Cologne
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	-0.04	-0.06	-0.06	-0.04	0.07	-0.05	-0.07
3	0.04	-0.06	-0.05	0.01	0.01	0.03	0.04
4	0.09	0.02	0.00	0.08	0.15	0.16	0.12
5	0.03	0.01	-0.06	-0.01	0.11	-0.01	-0.05
6	0.00	-0.09	-0.18	-0.12	0.07	-0.01	-0.03
7	-0.01	-0.05	-0.23	-0.17	0.05	-0.11	-0.08
8	0.01	-0.07	-0.25	-0.15	0.11	-0.13	-0.09
9	-0.00	-0.06	-0.09	-0.05	0.04	-0.11	-0.05
10	0.04	-0.00	-0.07	-0.03	0.11	-0.09	-0.04
11	-0.00	0.03	-0.04	0.07	0.19	-0.04	-0.09
12	0.05	0.01	-0.11	-0.09	0.16	-0.07	-0.11
13	-0.07	-0.09	-0.09	-0.16	0.09	-0.03	-0.15
14	-0.06	-0.06	-0.08	-0.16	0.17	-0.11	-0.12
15	-0.07	-0.11	-0.14	-0.16	0.10	-0.15	-0.16
16	-0.02	-0.12	-0.15	-0.17	0.11	-0.12	-0.16
17	0.00	-0.06	-0.14	-0.15	0.18	-0.02	-0.01
18	0.01	-0.05	-0.16	-0.13	0.14	-0.05	0.01
19	0.08	-0.09	-0.09	-0.10	0.16	-0.03	0.01
20	0.07	-0.11	-0.14	-0.13	0.09	0.08	-0.02
21	0.06	-0.13	-0.13	-0.09	0.04	0.04	-0.01
22	-0.04	-0.18	-0.12	-0.11	0.01	-0.04	0.04
23	-0.05	-0.22	-0.19	-0.22	-0.01	-0.03	0.04
24	-0.04	-0.22	-0.18	-0.33	-0.06	-0.06	0.01
25	-0.13	-0.19	-0.26	-0.32	-0.07	-0.20	-0.15
26	-0.12	-0.17	-0.24	-0.35	-0.03	-0.20	-0.19
27	-0.04	-0.17	-0.25	-0.32	-0.07	-0.13	-0.01
28	-0.06	-0.28	-0.27	-0.24	-0.02	-0.16	-0.02
29	-0.11	-0.28	-0.19	-0.31	-0.01	-0.12	-0.13
30	-0.33	-0.41	-0.38	-0.45	-0.13	-0.37	-0.30
31	-0.28	-0.45	-0.33	-0.35	-0.14	-0.32	-0.26
32	-0.38	-0.35	-0.26	-0.38	-0.12	-0.24	-0.22
33	-0.29	-0.39	-0.25	-0.38	-0.07	-0.28	-0.15
34	-0.27	-0.39	-0.21	-0.44	-0.16	-0.20	-0.15
35	-0.26	-0.44	-0.24	-0.41	-0.08	-0.12	-0.12
36	-0.24	-0.44	-0.18	-0.41	-0.11	-0.18	-0.09
37	-0.31	-0.41	-0.26	-0.43	-0.09	-0.13	-0.14
38	-0.39	-0.40	-0.20	-0.50	-0.16	-0.23	-0.16
39	-0.39	-0.42	-0.22	-0.38	-0.15	-0.16	-0.15
40	-0.40	-0.38	-0.26	-0.36	-0.11	-0.14	-0.09

NOTES: This table shows the estimated g_t (time-fixed effects for quarters 2009Q1–2018Q4) in equation (11) separately for each Top-7 labor market region.

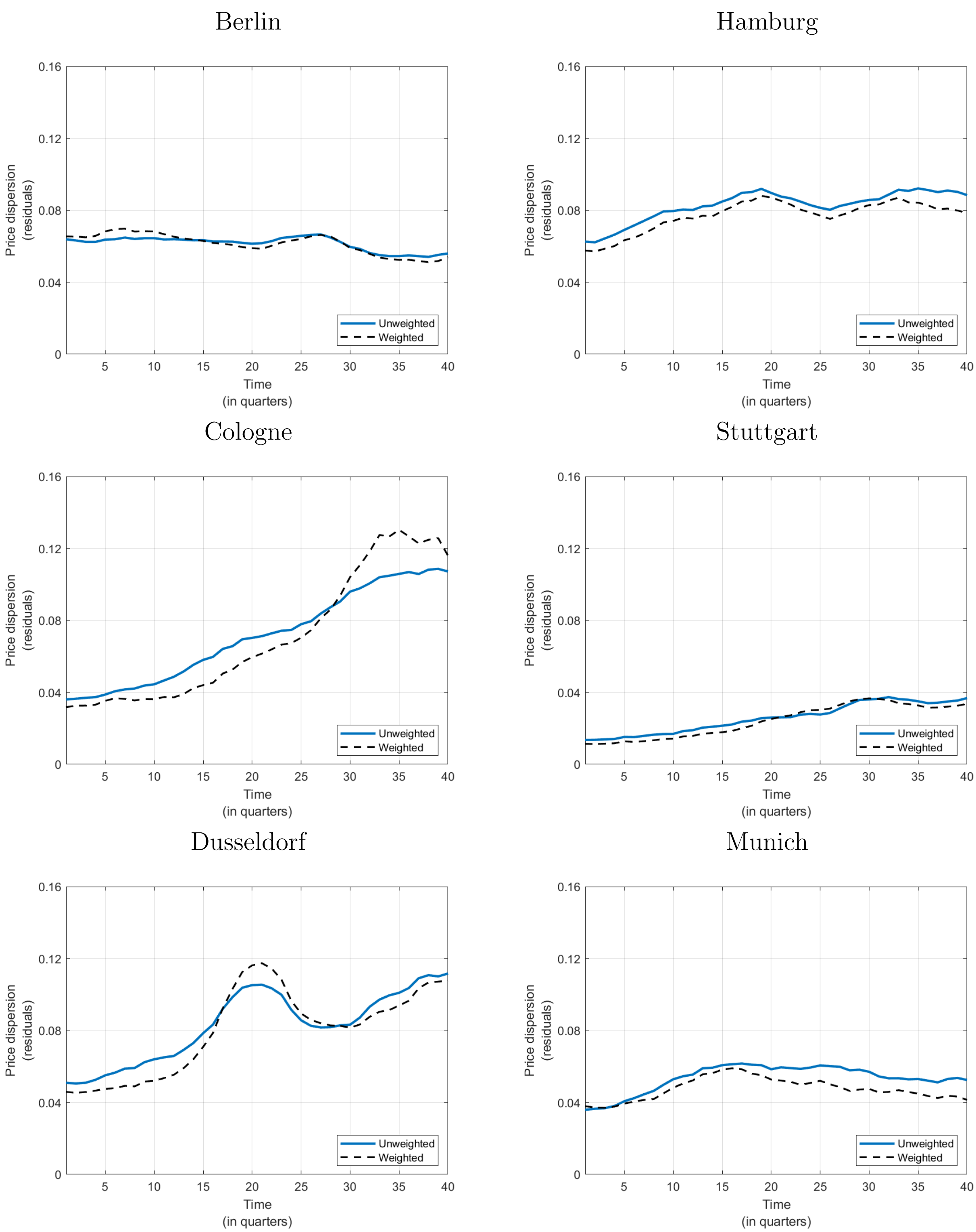


Figure 17: Price dispersion in selected regions

Notes: This figure shows the dispersion of the residualized log prices from the first quarter of 2009 to the last quarter of 2018. The blue solid lines show the unweighted dispersion and the black dashed lines the weighted dispersion based on the number of listings in each postal code.

E Variance Decomposition Derivations

Proof of Decomposition (2)

Write H for the set of listings and H_i for the set of listings in location i. Write n for the cardinality of H and n_i for the cardinality of H_i.

$$\begin{aligned}
\operatorname{var} \varepsilon_h &= \frac{1}{n} \sum_{h \in H} (\varepsilon_h - \bar{\varepsilon})^2 \\
&= \frac{1}{n} \sum_{i \in L} \sum_{h \in H_i} \left[(\varepsilon_h - \bar{\varepsilon}_i)^2 + 2(\varepsilon_h - \bar{\varepsilon}_i)(\bar{\varepsilon}_i - \bar{\varepsilon}) + (\bar{\varepsilon}_i - \bar{\varepsilon})^2 \right] \\
&= \sum_{i \in L} \frac{n_i}{n} \underbrace{\frac{1}{n_i} \sum_{h \in H_i} (\varepsilon_h - \bar{\varepsilon}_i)^2}_{=\operatorname{var}_i(\varepsilon_h)} + \frac{2}{n} \sum_{i \in L} (\bar{\varepsilon}_i - \bar{\varepsilon}) \underbrace{\sum_{h \in H_i} (\varepsilon_h - \bar{\varepsilon}_i)}_{=0} + \sum_{i \in L} \frac{n_i}{n} (\bar{\varepsilon}_i - \bar{\varepsilon})^2 \\
&= \sum_{i \in L} s_i \operatorname{var}_i(\varepsilon_h) + \sum_{i \in L} s_i (\bar{\varepsilon}_i - \bar{\varepsilon})^2 \ .
\end{aligned}$$

Proof of Decomposition (3)

$$\begin{aligned}
\sum_{i \in L} s_i (\bar{\varepsilon}_i - \bar{\varepsilon})^2 &= \sum_{r \in R} \sum_{i \in r} s_i \left[(\bar{\varepsilon}_i - \bar{\varepsilon}_r)^2 + 2(\bar{\varepsilon}_i - \bar{\varepsilon}_r)(\bar{\varepsilon}_r - \bar{\varepsilon}) + (\bar{\varepsilon}_r - \bar{\varepsilon})^2 \right] \\
&= \sum_{r \in R} \sigma_r \underbrace{\sum_{i \in r} \frac{s_i}{\sigma_r} (\bar{\varepsilon}_i - \bar{\varepsilon}_r)^2}_{=\operatorname{var}_r(\bar{\varepsilon}_i)} + 2 \sum_{r \in R} (\bar{\varepsilon}_r - \bar{\varepsilon}) \underbrace{\sum_{i \in r} s_i (\bar{\varepsilon}_i - \bar{\varepsilon}_r)}_{=0} + \sum_{r \in R} \sigma_r (\bar{\varepsilon}_r - \bar{\varepsilon})^2 \\
&= \sum_{r \in R} \sigma_r \operatorname{var}_r(\bar{\varepsilon}_i) + \sum_{r \in R} \sigma_r (\bar{\varepsilon}_r - \bar{\varepsilon})^2 \ .
\end{aligned}$$

F Estimation Strategy

Equilibrium conditions. We estimate the model separately for every labour market region. Consider a given labour market region r. For every location $i = 1, \ldots, N$ and quarter $t = 1, \ldots, T$ the model is defined by

$$\bar{V}_t^B = \ln\Big[\sum_j e^{V_{j,t}^B + \tau_j}\Big], \tag{F.1}$$

$$p_{it} = \zeta(\theta_{it})\, \beta V_{i,t+1}^S + \big[1 - \zeta(\theta_{it})\big]\big(A_{it} - \beta \bar{V}_{t+1}^B\big), \tag{F.2}$$

$$V_{it}^S = -c + \beta V_{i,t+1}^S + (q(\theta_{it}) - \theta_{it} q'(\theta_{it}))\left[A_{it} - \beta \bar{V}_{t+1}^B - \beta V_{i,t+1}^S\right], \tag{F.3}$$

$$V_{it}^B = -r_t + \beta \bar{V}_{t+1}^B + q'(\theta_{it})\big[A_{it} - \beta \bar{V}_{t+1}^B - \beta V_{i,t+1}^S\big], \tag{F.4}$$

$$V_{i,T+1}^S = \frac{2}{T(T-1)} \sum_{s=1}^{T} \big[3s - (T+2)\big] V_{i,s}^S, \tag{F.5}$$

$$V_{i,T+1}^B = \frac{2}{T(T-1)} \sum_{s=1}^{T} \big[3s - (T+2)\big] V_{i,s}^B, \tag{F.6}$$

$$\sum_i \tau_i = 0, \tag{F.7}$$

$$\tau_i = \frac{1}{T} \sum_{t=1}^{T} \Big[\ln \hat{\pi}_{it} + \bar{V}_t^B - V_{it}^B\Big] - \frac{\lambda}{2T}. \tag{F.8}$$

Note that given $(\bar{V}_{t+1}^B)_{t=1}^{T+1}$ the remaining equations are *linear* in the remaining unknowns

$$\Big\{A_{it},\ K_{it} = V_{it}^S,\ V_{it}^B,\ \tau_i,\ \lambda\ :\ i = 1, \ldots, N,\ t = 1, \ldots, T\Big\}.$$

Our numerical routine nests a linear inner loop inside a non-linear outer loop. The outer loop updates the buyer value $\bar{V}^B$ and the inner loop solves, conditional on $\bar{V}^B$, the block-diagonal linear system that delivers all location-specific unknowns.

1. Linear inner problem

Unknown vector X_i. For a given location i we order the $3T+3$ unknowns as

$$X_i \;=\; \big(A_{i1}, \ldots, A_{iT} \mid \bar{V}_1^B, \ldots, \bar{V}_{T+1}^B \mid K_{i1}, \ldots, K_{iT} \mid V_{i1}^B, \ldots, V_{i,T+1}^B \mid \tau_i\big)^{\top}.$$

Notice that only the $\bar{V}^B$ block is common across cities)

Coefficient matrix C_i. All terms that multiply an unknown are gathered in the square matrix $C_i \in \mathbb{R}^{(3T+3)\times(3T+3)}$:

Constant vector B_i. All observed variables $(p_{it}, \theta_{it}, r_t, c)$ and the current guess of $\bar{V}^B_{t+1}$ enter the right-hand side:

$$C_i X_i \;=\; B_i, \qquad B_i \in \mathbb{R}^{3T+3}.$$

Global sparse system. Stacking city blocks on the diagonal and appending one row and column for the time-invariant normalisation $\sum_i \tau_i = 0$ gives

$$C_{\text{big}} \;=\; \begin{bmatrix} C_1 & & & \\ & C_2 & & \\ & & \ddots & \\ & & & C_N \end{bmatrix}, \qquad B_{\text{big}} \;=\; \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_N \\ 0 \end{bmatrix},$$

an $(3T+3)N+1$ dimensional sparse system that we solve once per outer iteration:

$$C_{\text{big}}\, X \;=\; B_{\text{big}} \quad\Longrightarrow\quad X = C_{\text{big}}^{-1} B_{\text{big}}.$$

2. Outer fixed-point update

With the location-specific values V^B_{it} and τ_i just computed, update

$$\bar{V}^{B\,\text{new}}_t \;=\; \ln\Bigl[\sum_{i=1}^{N} \exp\bigl(V^B_{it} + \tau_i\bigr)\Bigr], \qquad t = 1, \ldots, T{+}1,$$

and move to inner step 1 and continue iterating until $\max |\bar{V}^{B\,\text{new}}_t - \bar{V}^B_t| < 10^{-9}$.