ARTICLE IN PRESS

Regional Science and Urban Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Micro-geographic property price and rent indices[☆]

Gabriel M. Ahlfeldt [a,b,c,d,*,1], Stephan Heblich [d,e,c,f,g,1], Tobias Seidel [h,c,i,1]

[a] *London School of Economics and Political Sciences (LSE), United Kingdom*
[b] *Center for Economic Policy Research, United Kingdom*
[c] *CESifo, Germany*
[d] *CEP, United Kingdom*
[e] *University of Toronto, Canada*
[f] *IFW Kiel, Germany*
[g] *NBER, United States of America*
[h] *University of Duisburg–Essen, Germany*
[i] *CRED, Switzerland*

ARTICLE INFO

Dataset link: https://doi.org/10.7807/immo:red:ahs:v1

JEL classification:
R10

Keywords:
Index
Real estate
Price
Property
Rent

ABSTRACT

We develop a programming algorithm that predicts a balanced-panel mix-adjusted house price index for arbitrary spatial units from repeated cross-sections of geocoded micro data. The algorithm combines parametric and non-parametric estimation techniques to provide a tight local fit where the underlying micro data are abundant, and reliable extrapolations where data are sparse. To illustrate the functionality, we generate a panel of German property prices and rents that is unprecedented in its spatial coverage and detail. This novel data set uncovers a battery of stylized facts that motivate further research, e.g. on the positive correlation between density and price-to-rent ratios in levels and trends, both within and between cities. Our method lends itself to the creation of comparable neighborhood-level rent indices (*Mietspiegel*) across Germany.

1. Introduction

Reliable indices that capture the market value of real estate at micro-geographic scales such as neighborhoods are important inputs into housing policy. The ability of a regulator to enforce rents that are deemed fair critically depends on the capacity to observe the market value of real estate. The German "Mietspiegel", for example, represents a core instrument to settle disputes between landlords and tenants over rent levels. Micro-geographic property price indices are also an increasingly important input into economics research. For instance, quantitative spatial models require spatially disaggregated data with full geographic coverage for the inversion of the structural fundamentals before they can be used for quantitative analysis (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015).[2] However, the gold standard in house price index construction—repeat sales indices such as the prominent Case—Shiller Home Price Index—are not suitable for micro-geographic areas because property transactions are rare events at this scale, let alone repeated transactions.

Our contribution is to develop an algorithmic approach to construct a panel of micro-geographic house price and housing rent indices that draws on spatial methods to overcome the limitations of sparse property transaction data. Because our approach is entirely point-pattern based, it is applicable to arbitrary spatial units and does not depend on administrative boundaries. The input is a conventional data set containing pooled cross sections of real estate transactions with information on prices or rents, geographic coordinates, transaction dates, and property characteristics. The output is a balanced panel of mix-adjusted purchase or rental prices for arbitrary spatial units. The algorithm automatically adjusts to spatially varying densities of observations using a combination of parametric and non-parametric

☆ We are grateful to Sandra Schaffner and the FDZ group at RWI for sharing a comprehensive property micro data set. We also thank Thomas Sandner from the Berlin Senate Department for providing access to the Berlin land registry data (*Kaufpreissammlung*). We thank Philipp Breidenbach, Max von Ehrlich, Tobias Just, Hans Koster, Thies Lindenthal, Ronan Lyons, Wolfgang Maennig, Claus Michelsen, and Nikodem Szumilo for helpful comments. The usual disclaimer applies.

* Corresponding author at: London School of Economics and Political Sciences (LSE), United Kingdom.
E-mail address: g.ahlfeldt@lse.ac.uk (G.M. Ahlfeldt).

[1] All authors contributed equally to the paper.

[2] See Redding and Rossi-Hansberg (2017) for a survey and Monte et al. (2018), Tsivanidis (2019), Heblich et al. (2020), Almagro and Domínguez-Iino (2021) for recent examples.

https://doi.org/10.1016/j.regsciurbeco.2022.103836
Received 21 September 2021; Received in revised form 17 September 2022; Accepted 19 September 2022
Available online 6 October 2022

estimation techniques. Conveniently, it allows users to manage the *bias–variance trade-off* in the program syntax. A spatial sub-market variable further allows users to specify where they would expect spatial discontinuities in an otherwise smooth surface. The result is a reliable and transparent tool that generates spatial house price and rent indices in an environment that is typically dominated by commercial data providers to whom their algorithms are the "secret sauce". In contrast, we publish our source code along with novel price and rent indices covering all of Germany at the level of local labor markets, counties, municipalities, and postcodes for a period of fifteen years.[3]

The house price and rent index we propose combines several techniques that are established in urban economics and data science. We start with the popular hedonic regression approach developed by Rosen (1974) to adjust for observable property characteristics and combine it with recent extensions of early work by Clark (1951) on price indices that treat spatial units as the nucleus of a spatial price gradient (Combes et al., 2019; Ahlfeldt et al., 2020). Similar to Clapp (2004) and Sunding and Swoboda (2010), we nest this approach into a locally weighted regression framework (LWR) that was originally suggested by Cleveland and Devlin (1988) and first adopted to studies of property price data by Meese and Wallace (1991) and McMillen (1996). More recently, the method has become a widespread tool in geographic data science under the label *Geographically Weighted Regression*. The following three features distinguish our approach from existing work. First, we apply LWRs to create real estate price indices in a panel setting; second, we use local samples when performing LWRs *and* supplement them with local spatial covariates; and third, we accommodate the existence of sub-markets. As a result, our approach generates rich and even spatially discontinuous variation in predicted prices, both in levels and trends.

Intuitively, we treat the computation of the indices for *any* spatial unit as a separate problem that we address in a *separate* iteration of the algorithmic approach. In each iteration, the algorithm considers the density of observations in the vicinity of the targeted location and flexibly defines the size of a spatial window that provides a sufficient amount of observations. Inside this spatial window, observed prices are adjusted for structural and location characteristics using conventional regression techniques. Moreover, observations are allowed to be on separate trends depending on whether or not they fall within the same user-specified sub-market as the *target* unit for which a price index is being predicted. To predict the price and rent indices right at the target location, we control for a first-order polynomial of distance from the center. We also allow for a spatial fixed effect which depends on the density of observations. Combining parametric and non-parametric specifications avoids the problem that higher-order polynomials tend to chase after outliers in the tails of a distribution. The strength of the algorithmic approach is that it loads the predictive power on non-parametric components where many observations are available (e.g. in high-density urban neighborhoods), whereas the predictive model becomes more parametric if observations are sparse (e.g. in rural regions). Importantly, the user retains control over the *bias–variance trade-off* via a set of parameters to be determined in the programming syntax. Our advice is to proceed with conservative parameter values that minimize the risk of generating implausible outliers. However, users may choose different values—resulting, for example, in smaller spatial windows—that best suit their aversion to outliers. If users were willing to formalize their objective function that trades off bias against variance, they could also delegate the identification of the critical parameter values to another algorithm. In this case, our approach would become a variant of supervised machine learning.

For transparency and to facilitate use, we publish a ready-to-use version of our algorithm in the appendix to this paper and we employ the algorithm in a practical application to Germany where we describe its functionality. Our application makes use of geocoded data from the online platform *Immoscout24* for the period 2007–2021. This time window is largely representative of the rising pool of information on prices and rents that is accessible to researchers and data scientists around the world. Beside address information, the data also hold information on basic property characteristics which we exploit following the conventions in the hedonic pricing literature. We start with an application where we aggregate the price information into official spatial units, i.e. labor market areas, counties, municipalities and postal codes. This allows us to visually assess the accuracy of our data but it also reveals that German postal codes are more coarse than they are in e.g. the UK or the U.S. To illustrate how we can capture even smaller, arbitrary spatial units, we introduce another application where we aggregate the house price information in hexagons with a $500m$ diameter. To evaluate whether our procedure accurately predicts a house price index in sparsely populated areas, we exclude information from about three quarters of all hexagons and recalculate the index for all locations in an out-of-sample prediction. A comparison between the within-sample and out-of-sample predictions shows a tight fit that underlines the validity of our procedure. Using a spatial boundary discontinuity design and the former Berlin Wall as an example, we also demonstrate how our algorithm generates spatial discontinuities in price levels and trends along the boundaries of user-defined sub-markets.

The application to the case of Germany fills a gap in the literature since house price and rent indices that cover all of Germany are not available below the county level. This implies that a lot of spatial heterogeneity within counties remains unobserved and a location's attractiveness may be confounded by commuting costs (Combes et al., 2019). By contrast, our index reports year-specific conditional means of either rents or house prices adjusted for property characteristics and location. Applying our LWR-approach with endogenously adjusting spatial windows to micro property data provides a spatial resolution that is well below the county level. This allows us to zoom into local housing markets and complement the labor market data provided by the Research Data Centre of the Federal Employment Agency in Germany at all spatial aggregation levels with a cost-of-living measure (see Ahlfeldt et al., 2020, for an application).

Another benefit of our data is that they include both house price and rent information. Especially in German cities where ownership rates are still below 60 percent, any picture of the national real estate market remains incomplete unless the rental market is taken into consideration. We directly relate to an emerging literature that analyzes the determinants of price-to-rent ratios, albeit at a much finer spatial resolution. Our micro-geographic rent and purchase price indices reveal new stylized facts that call for further analyses: price–rent ratios strongly diverged in more densely populated areas since 2010, both in levels and trends. Price-to-rent ratios tend to be higher in large cities and within cities, they tend to be higher in more central parts. The data set we share will allow researchers to further investigate the origins of this relationship that may relate to e.g. supply conditions (Glaeser et al., 2008; Hilber and Mense, 2021; Büchler et al., 2021), credit constraints (Himmelberg et al., 2005), foreign direct investment (Badarinza and Ramadorai, 2018), rent regulation (Diamond et al., 2019; Mense et al., 2019; Breidenbach et al., 2022), or gentrification (Couture and Handbury, 2020).

More generally, our work connects to various important research strands that are concerned with either the generation or use of spatial price data. The literature on house price indices is too large to be comprehensively summarized here. Instead, we refer to European Commission (2013) for an overview.[4] Recent notable developments in

[3] The public code directory can be found at https://doi.org/10.7807/immo:red:ahs:v1.

[4] House price indices are often reported with a higher frequency (e.g. quarters), which comes at the expense of less spatial granularity (see for example Francke, 2010; van Dijk et al., 2022). By contrast, we put more emphasis on spatial granularity at the expense of a lower frequency, i.e. annual time periods.

this literature are the use of matching approaches (Lopez and Hewings, 2018) to broaden samples beyond repeat sales (Bailey et al., 1963), adaptive weights smoothing to produce land value surfaces (Kolbe et al., 2015; Hill and Scholz, 2018), or machine learning approaches that capture otherwise unobservable housing characteristics (Shen and Ross, 2021). This strand of research is a manifestation of a broader trend to fit flexible functional forms to data that support out-of-sample predictions. For a discussion of prediction algorithms with a specific focus on housing, we refer the interested reader to Mullainathan and Spiess (2017) and to Athey and Imbens (2019) for a more general discussion of the use of machine learning in economics. Our contribution to this literature is to combine various recent techniques with the aim of laying out a transparent and theory-consistent methodology for the generation of micro-geographic price and rent indices in a panel setting.

On the applied side, fine-grained house price data are routinely used to evaluate housing policies such as rent control (Diamond et al., 2019; Autor et al., 2014; Sims, 2011), quantify spatial models (see Redding and Rossi-Hansberg, 2017, for a review), measure the costs of agglomeration (see Ahlfeldt and Pietrostefani, 2019, for a review), infer quality of life (Roback, 1982; Ahlfeldt et al., 2020), evaluate economic cycles (Mian and Sufi, 2014; Hoffmann and Lemieux, 2015; Charles et al., 2018), or value local public goods such as clean air (Chay and Greenstone, 2005), safety (Linden and Rockoff, 2008) or the quality of public schools (Cellini et al., 2010), just to name a few. Our contribution to this vast literature is to provide researchers with a convenient, transparent, and flexible tool for the preparation of an essential input into their research.

The rest of the paper is organized as follows. Section 2 introduces our algorithm. Section 3 provides an application to Germany. Section 4 provides new stylized facts based on the novel indices we generate. The final Section 5 concludes.

2. Procedure

This section introduces our algorithm. We lay out the locally weighted regression (LWR) approach formally (Section 2.1) and then discuss various use cases along with our preferred parameter values (Section 2.2).

2.1. Algorithm

The empirical approach outlined in this section generates a mix-adjusted property price index for an arbitrary set of *target* spatial units indexed by $j \in \mathcal{J}$. For each j, we run a locally weighted regression of the following type:

$$\begin{aligned}\ln \mathcal{P}_{i,t} = {} & \alpha_t^j + \bar{S}_i b^j + d_t^j D_i^j + e^j I(D_i^j > T^j)_i \\ & + f^j(X_i - X^j) + g^j(Y_i - Y^j) + h_t^j I(M_i \neq M^j) + \epsilon_{i,t}^j,\end{aligned}$$

where $\mathcal{P}_{i,t}$ is the purchase or rental price of a property $i \in \mathcal{I}$ transacted in year $t \in \mathcal{T}$. α_t^j captures j-specific time-fixed effects, $\bar{S}_i$ is a vector of covariates stripped off the national average (we subtract the national mean from the observed value of S_i), and b^j are the LWR-j-specific hedonic implicit prices. D_i^j is the distance from a transacted property i to the target unit j with d_t^j being the LWR j-specific gradient parameter in year t. $I(.)$ is an indicator function that returns a value of one if a condition is true and zero otherwise and T^j is a threshold distance. Hence, $e^j I(D_i^j > T^j)_i$ is a fixed effect for all transacted properties i that are outside the vicinity of the catchment area. X_i and Y_i are the coordinates of transacted properties, X^j and Y^j are the coordinates of the target unit, and f^j and g^j are parameters to be estimated. M_i and M^j describe spatial sub-markets that can be defined arbitrarily by the user. Hence, the term $h_t^j I(M_i \neq M^j)$ allows for transactions i to be on a different time-trend if they fall into a different spatial sub-market than target unit j. $\epsilon_{i,t}^j$ is the residual term.

The threshold T^j is chosen using the following rule:

$$T^j = \begin{cases} T^1, & \text{if } N^{(D_i^j \le T^1)} \ge N^T \\ T^2, & \text{if } N^{(D_i^j \le T^1)} < N^T \le N^{(D_i^j \le T^2)} \\ T^3, & \text{if } N^{(D_i^j \le T^2)} < N^T \le N^{(D_i^j \le T^3)} \\ T^4, & \text{if } N^{(D_i^j \le T^3)} < N^T, \end{cases}$$

where $N^{(D_i^j \le T^{s\in\{1,2,3,4\}})}$ gives the number of transacted units from a target unit within distance threshold $T^{s\in\{1,2,3,4\}}$ and N^T is a minimum-number-of-transactions threshold, all to be chosen by the user in the program implementation of this algorithm.

In each LWR j, all transacted properties i are weighted using the following kernel weight:

$$W_i^j = \frac{w_i^j}{\sum_i w_i^j}$$

$$w_i^j = \begin{cases} I(D_i^j \le A^1), & \text{if } N^{(D_i^j \le A^1)} \ge N^A \\ I(D_i^j \le A^2), & \text{if } N^{(D_i^j \le A^1)} < N^A \le N^{(D_i^j \le A^2)} \\ I(D_i^j \le A^3), & \text{if } N^{(D_i^j \le A^2)} < N^A \le N^{(D_i^j \le A^3)} \\ I(D_i^j \le A^4), & \text{if } N^{(D_i^j \le A^3)} < N^A, \end{cases}$$

where $\{A^1, A^2, A^3, A^4\}$ are distance thresholds and N^A is a minimum-number-of-transactions threshold, all to be defined by the user in the program implementation of this algorithm.

The LWR-j-specific estimates of time-fixed effects correspond to the conditional expectation

$$\hat{\alpha}_t^j = \mathbb{E}\left(\ln P_t^j | S = \bar{S}, D^j = 0, X = X^j, Y = Y^j, M = M^j\right),$$

which is the expected log price at location j at time t for a property with the average national characteristics. We convert this conditional expectation into a price index measured in the same units as $P_{i,t}$ (here, €/m^2) as follows:

$$\hat{\mathcal{P}}_t^j = \exp(\hat{\alpha}_t^j) \times C^j = \mathbb{E}\left(P_t^j | S = \bar{S}, D^j = 0, X = X^j, Y = Y^j, M = M^j\right),$$

where $C^j = \exp\left(\frac{1}{2}(\hat{\sigma}_\epsilon^2)^j\right)$ is an adjustment factor that depends on the variance of the error $(\sigma_\epsilon^2)^j$ of LWR j.[5] This adjustment is necessary to correct for the bias that would otherwise arise when reversing the log transformation (Duan, 1983). This way, we ensure that our index can be interpreted as the expected price of a property with average characteristics. To facilitate the computation of confidence bands, we also report standard errors

$$\hat{\sigma}_{\mathcal{P}_t^j} = \hat{\mathcal{P}}_t^j \times \hat{\sigma}_{\alpha_t^j},$$

where $\hat{\sigma}_{\alpha_t^j}$ are estimated allowing for clustering within the areas inside and outside the spatial fixed effect $(I(D_i^j > T^j)_i)$.[6]

Intuitively, the price index for a target unit is a year-specific local conditional mean that is adjusted for property characteristics (deviations from the national average), location (time-varying distance from j effects, and time-invariant spatial trends in X and Y coordinates), and a spatial fixed effect. Since $\{w_i^j, T^j\}$ are endogenously chosen by the algorithm, the precision of the index automatically increases as the density of observations increases.

The parameters $\{A^1, A^2, A^3, A^4, N^A, T^1, T^2, T^3, T^4, N^T\}$ allow the user to flexibly control the *bias–variance trade-off*. Smaller values in all parameters will generally lead to greater spatial variation which comes at the cost of an increasing sensitivity to outliers in the underlying micro-data. In choosing N^A, it is worth recalling that N^A describes the

[5] We use the square of the Root Mean Square Error (RMSE) as an estimate of the error variance.

[6] Notice that our standard errors are derived from a first-order approximation of the second moment of the function $\hat{\mathcal{P}}_t^j(\alpha_t^j) = \exp(\hat{\alpha}_t^j) \times C^j$ assuming a constant variance of the error term $\epsilon_{i,t}^j$.

number of observations that occur over multiple years, but estimates of conditional means and distance gradients are year-specific. Thus, as a rule of thumb, N^A should increase proportionately to the number of years over which an index is predicted.

2.2. Implementation

The algorithm introduced in Section 2.1 is a flexible tool that assists the user with a mix-adjusted aggregation and interpolation of spatial point-pattern transactions data. The algorithm is flexible enough to serve a range of popular use cases in urban economics research, but the parameter values must be chosen to fit the intended purpose. In the below, we provide a discussion that should guide the user in the choice of parameter values and the application of the algorithm more generally. In particular, we introduce what we view as canonical parameter values for spatial units that are popular in spatial economics research. Based on our experience with transactions data, these parameter values should be applicable to most international contexts.

2.2.1. Pre-processing

Independently of the purpose of the application, some careful pre-processing of raw transactions data is strongly recommended. There is a large literature summarized by Silver (2018) discussing why observations that are not well predicted by a hedonic model—outliers—carry a greater weight in a constant-quality house price index than they deserve. This problem is naturally aggravated in a LWR approach that returns a micro-geographic property price index because the potentially large set of transactions is partitioned into smaller subsets in any LWR $j \in J$. As a consequence, any outlier will carry an even greater weight. It is, thus, especially important to carefully inspect the raw data and eliminate outliers before applying our algorithm. There are many ways of detecting outliers and we refer the interested reader to Silver (2018) for a general discussion and to Section 3.1 for our choices in an illustrative example.

2.2.2. Indices based on the monocentric city model

Once the raw transactions data set is in good shape, the most pressing question is the potential use of the index to be generated. One typical application concerns the measurement of housing costs in local labour markets (LLMs). LLMs are reasonably self-contained areas where people live and work, roughly corresponding to metropolitan areas. Since LLMs are relatively large spatial units, there are typically enough transactions to estimate a conventional hedonic house price index. However, from a theoretical perspective, a problem arises with the interpretation of the standard non-spatial index. In the standard urban model, differences in mix-adjusted property prices reflect the fundamental value of locations, net of commuting costs (Alonso, 1964; Mills, 1967; Muth, 1969). Therefore, a naive price index does not only capture the fundamental value property occupiers derived from a location, but also a discount that arises from commuting costs. This is a problem because commutes are mechanically longer in larger cities and, hence, the fundamental attractiveness is mechanically undervalued in conventional regional price indices. Albouy and Lue (2015) demonstrate that failure to account for differences in commuting cost leads to mismeasurement of quality-of-life differences across cities.

Combes et al. (2019) popularized the idea of measuring the cost of agglomeration at the city center, where prices are not confounded by commuting in the standard urban model. To this end, they introduce distance from the city center as a variable into a hedonic regression model, along with a city-specific implicit price. Our algorithm offers a convenient way of estimating such a price index that is not only mix-adjusted for property characteristics, but also for commuting cost to the city center. All that is needed, besides the transactions raw data, is a set of *target* coordinates that approximate city centers, and a suitable choice of parameter values. Our recommended parameter values for this approach are summarized in the first two columns of Table 1.

Table 1
Parameter choices by spatial layer.

	Monocentric city model		Best local fit		
	LLMs	Counties	Municipalities	Neighborhoods	Micro-Grids
A^1	25	25	10	10	5
A^2	50	50	25	25	10
A^3	75	75	50	50	25
A^4	100	100	100	100	50
N^A	10,000[a]	10,000[a]	10,000[a]	10,000[a]	10,000[a]
T^1	100	100	10	2.5	1
T^2	100	100	15	5	2
T^3	100	100	20	10	5
T^4	100	100	50	20	10
N^T	0	0	1,000[a]	1,000[a]	1,000[a]

Notes: LLM refers to local labor markets. Parameters A^s and T^s represent threshold distances measured in kilometers.
[a]Recommended parameter for indices spanning 10 periods.

We set $N^A = 10{,}000$, which implies that the algorithm will seek to ensure that there are at least 10,000 transactions used in each LWR j. This is a sufficiently large number to minimize the leverage of outliers. Notice that the choice of N^A should not be independent from the number of periods for which an index is predicted. With 10 periods, there will be about a thousand transactions per fixed effect a_t^j to be estimated. Even if the number of transactions varies over time, there will be a sufficiently large number of transactions even in years with thin markets. The same argument does not necessarily apply if the number of periods is much greater. Therefore, we recommend to scale N^A proportionately by the number of periods the index is predicted for.

We set critical distance threshold parameters $\{A^1, A^2, A^3, A^4\}$ in the area weights W_i^j to $\{25, 50, 75, 100\}$, all measured in km. The implication is that the algorithm selects transactions within 25 km of a target location j for the respective LWR unless there are less than 10,000 observations within this area. The algorithm will expand to 50, 75, or 100 km if necessary, but in practice this is unlikely to happen as the density of transactions around city centers tends to be reasonably high. We choose 25 km as the minimum distance threshold because this appears like a reasonable approximation of an area from which the great majority of commuters will commute to the center of a LLM.

For the spatial fixed effect $e^j I(D_i^j > T^j)$, we set the distance threshold parameter values to $T^{s\in\{1,2,3,4\}} = \bar{T} = 100$. Since $\bar{T} \geq A^s \forall s \in \{1,2,3,4\}$, the spatial fixed effect drops out in all LWR. This is, of course, intentional, since we seek to measure the average fundamental value within a LLM defined by $D_i^j \leq A^1$ and not the fundamental value of the target location (a dimensionless point) itself. Since the fixed effect drops out anyways, the choice of N^T does not matter, which is why we set it to zero for convenience.

We believe that the approximation of LLMs works best with spatial units that are delineated based on the actual spatial distribution of economic activity, be it based on commuting data (e.g. Kosfeld and Werner, 2012), population densities (e.g. OECD, 2012), built-up areas (e.g. de Bellefon et al., 2021), or night lights (e.g. Dingel et al., 2021). However, there are instances where researchers must proceed with administratively defined areas due to compatibility with other administrative data sets. A popular spatial unit in this context is the county (e.g. the European NUTS3). As long as a county contains a sizeable fraction of a LLM–and there is typically no more than one dominating employment center within a county–the procedure outlined in this section can also be applied to counties.

2.2.3. Local best fit indices

As we move from larger spatial units describing LLMs to smaller spatial units such as neighborhoods, there is no longer a need for a theory-consistent adjustment by commuting cost. Typically, micro-geographic property price indices will be interpreted through the lens of an urban model and may feed into a quantitative spatial model that accounts for commuting costs (Ahlfeldt et al., 2015). Therefore, the user will be interested in the best possible prediction of the fundamental value of a target location, a point that represents an area that is sufficiently small to be considered homogeneous. The perhaps most obvious example is a neighborhood, a subdivision of a city or town that is just about walkable. Ideally, neighborhoods will be delineated such that they encompass addresses that are reasonably similar in terms of density, building structure, and socioeconomic characteristics. Naturally, when compared to the theory-based LLM index, a different parametrization will be suitable when solving the purely empirical problem of obtaining the best local fit for a neighborhood.

As we are now looking into cities and ignoring commuting costs, we can allow the LWR to become more local. Therefore, we set the critical distance threshold parameters $\{A^1, A^2, A^3, A^4\}$ in the area weights W_i^j to $\{10, 25, 50, 100\}$, as reported in the fourth column of Table 1. Hence, our algorithm will fit a hedonic model to an area within 10 km of a target location j if our minimum requirement of $N^A = 10{,}000$ observations is satisfied. If necessary, the algorithm will search over greater distances until the minimum-requirement $N^A = 10{,}000$ is satisfied. In principle, one could consider smaller thresholds, in particular for A^1. However, considering a whole country, there will be many sparsely populated areas where the observation threshold would never be reached. Importantly, the target-location-j-specific distance control ensures that the index prediction is made for the target location. In other words, we estimate the conditional mean at a zero distance $\mathbb{E}\left(\ln \mathcal{P}_{i,t} | D_i^j = 0\right)$ and not the naive average price adjusted for property characteristics within the 10-km zone. Moreover, we use much smaller values for the distance thresholds in the fixed effect $\{T^1, T^2, T^3, T^4\}$, which we set to $\{2.5, 5, 10, 20\}$. This fixed effect accounts for time-invariant effects and further increases the local fit. Of course, the observation threshold for the fixed effect $N^T = 1000$ *must* be lower than N^A since only a subset of observations can fall into a circle with a radius $T^s < A^s$. It *can* be lower, because we only identify one parameter from the spatial fixed effect, whereas we estimate $\mathcal{T}$ fixed effects α_t^j from all transactions drawn in a LWR j. A lower value could be chosen at the cost of increasing the sensitivity to the outlier leverage problem. As we show in our application in Section 3.3, the chosen parameter values result in a local fit that captures large differences in fundamental values over short distances within cities. At the same time, it also accommodates regions in which transactions are sparser. This is important when creating a spatial price index for an entire country, which will also encompass many rural areas.

For many applications, researchers may be interested in a particular urban area where the density of transactions is generally high. In this case, even smaller values can be chosen for the distance threshold parameters. With this in mind, we set $\{A^1, A^2, A^3, A^4\}$ to $\{5, 10, 25, 50\}$ and $\{T^1, T^2, T^3, T^4\}$ to $\{1, 2, 5, 10\}$ in the last column of Table 1. Now, the algorithm will seek to run a LWR which uses only observations within 5 km and account for arbitrary local fundamentals within 1 km via the spatial fixed effect if the observation thresholds $\{N^A, N^T\}$ are satisfied. With this approach, it is possible to generate a spatial index at the sub-neighborhood level, such as for the 500 × 500 m cells in Fig. 5. If the underlying transactions data are exactly geocoded (as opposed to geocoding at a higher level such as a postcode or grid cell) and the density of observation is sufficiently high, even smaller parameter values may be feasible.

Another popular, administratively defined spatial unit are municipalities. They are generally smaller than counties. Typically, there will be multiple municipalities within a local labor market, which is why they are no obvious candidates for a theory-consistent index based on the monocentric city model. Therefore, we recommend the same parameter values $A^{s\in\{1,2,3,4\}}$ as for neighborhoods in Table 1. For $T^{s\in\{1,2,3,4\}}$, we use larger values because municipalities are normally not as small as neighborhoods within cities.

2.2.4. Standard regional house price indices

In Table 1, and the discussions above, we have outlined two use cases for our algorithm: a theory-consistent index based on the monocentric city model that is suitable for larger spatial units and a local best fit index for smaller units. We have ruled out the need for theory-consistent local indices. A fourth possible case are local best fit indices for larger spatial units with no theory-consistent adjustment. Such standard regional hedonic house price indices return constant-quality average property prices by region. If spatial units are sufficiently large, such indices can conveniently be recovered from region-by-period fixed effects added to a conventional hedonic regression. This is why we do not see this application as the primary use case for our algorithm.

Still, our algorithm can help in instances where the spatial distribution of transactions departs from the spatial distribution of the economic activity. Consider a property transactions data set which contains owner-occupied single-family houses in suburbs and renter-occupied multi-family houses downtown. The former will transact a lot more frequently, but the latter host a lot more people. In a conventional hedonic regression, all observations would carry the same weight and the resulting index would consequently be driven by single-family homes even if more people lived in multi-family houses. A convenient alternative would be to use our algorithm to generate a local price index for neighborhoods and use population weights to aggregate it to a higher level, such as LLMs.

2.2.5. Rolling indices

Researchers will often be interested in using our algorithm to create a spatial house price index as an input in a closed-ended analysis. For example, they may use our index to invert a quantitative spatial model for various periods or estimate a VAR model, both of which will require a balanced panel. Beyond such specific applications, a spatial price index is interesting in its own right as a means to monitor the evolution of property prices over time. To this end, the estimation of the index will have to be repeated over time.

Extending the index calculations as data for additional years become available is challenging because it will not only result in new index values for the added periods, but also change the index values for all previously computed periods. This is a standard problem in the computation of rolling house price indices which is discussed extensively in various guides to practice (e.g. European Commission, 2013). The standard approach is to estimate a hedonic model for a rolling window of periods, for example 11 years. Over time, the size of the window remains constant, but the base year period shifts each time the hedonic model is re-estimated. For example, one could estimate the hedonic model at the beginning of 2011, including all observed transactions that occurred from 2000 to 2010. In 2012, the procedure could be repeated, this time including all transactions from 2001 to 2011. To obtain the 2011 index value, the 2010 to 2011 growth rate from the later hedonic index is applied to the last period of the earlier hedonic index.

We see no reason why this established approach should not be extended to spatial panel indices generated using our algorithm. Since our algorithm is designed to return predicted prices for a property with representative characteristics in the same units (e.g. rent in €/m^2) as the underlying micro data, one may as well just add the last new index values from the latest run to an already existing panel. In this context, we wish to highlight that we provide an alternative version of the algorithm that takes spatial weights W^j and the fixed effects distance thresholds T^j as exogenously given. This allows the user to feed threshold parameters identified in the initial application into later applications to ensure that these parameters remain constant in any update.

2.2.6. Spatial discontinuities

Fig. 1 illustrates how our algorithm restricts the sample of transactions in a j-specific LWR to a circular area (highlighted by the red circle) that surrounds a target coordinate $(X^j, Y^j$, marked by the red dot). Intuitively, properties within a distance $D_i^j < A^s$ are relatively

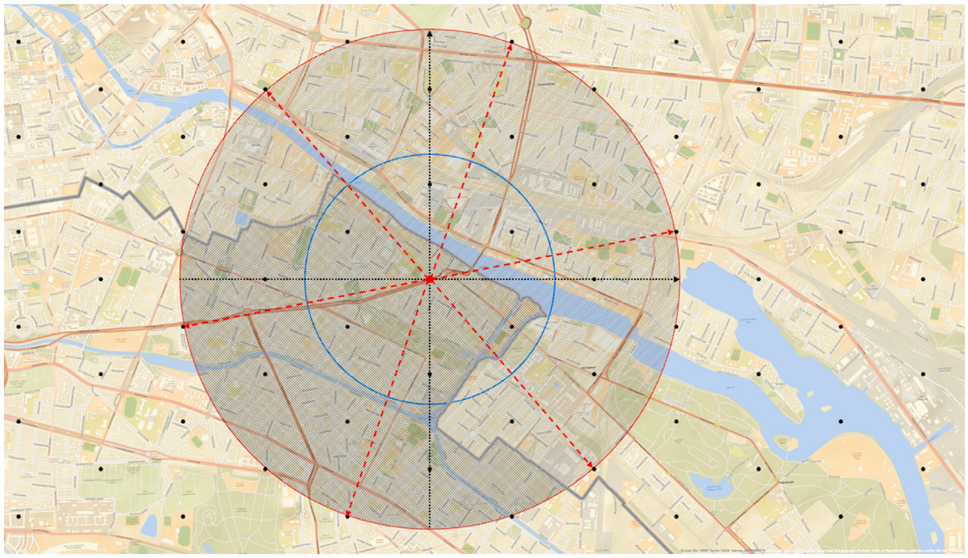


Fig. 1. Spatial trends and sub-markets. Note: The red dot marks the centroid of an arbitrary target unit j for which a price index is being predicted. Black dots are other target units. The red circle marks the area sampled in a LWR j defined by $D_i^j \leq A^s$. The blue circle marks the spatial fixed effect $D_i^j \leq T^s$. Dashed red arrows illustrate the spatial trend in D_i^j. Black dotted arrows illustrate the spatial trends in X_i and Y_i. The thick gray line is the former Berlin Wall that separates to sub-markets in LWR j (cross-hatched and hatched areas). Intuitively, LWR j predicts a price trend for the red dot allowing the hatched area to be on a different trend. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

similar in terms of locational characteristics compared to the rest of the sample (here, all of Germany). Still, location characteristics may differ *within* this area which is why we allow for year-specific distance trends (along the red dashed arrows). Since, as one moves away from the target, prices may fall or increase more or less depending on the cardinal direction, we, in addition, control for trends in geographic coordinates (along the black dotted arrows). Finally, we allow for a smaller locational fixed effect (marked by the blue circle) to account for unobserved locational characteristics that are not fully captured by the distance trends. As with any spatial interpolation approach, the rationale for distance-based weights and covariates rests on the assumption that unobserved characteristics, which determine property prices and rents, vary smoothly in space. This assumption is generally plausible because residents can travel in space: Therefore, the effects of location characteristics will vary smoothly in space even if the underlying characteristics do not.

If accessibility is restricted, however, this logic will no longer apply and the value of locations can change discontinuously in space. Restrictions to mobility may arise because there are no direct connections due to physical barriers such as rivers or mountains, but they may also arise due to political, historical, or cultural reasons (e.g. language barriers, differences in the institutional framework, or perceptions of space). Our algorithm offers a simple way to allow for such spatial discontinuities via the spatial sub-market variable M_i. As a case in point, we use the perhaps most prominent barrier to mobility in urban history: the Berlin Wall (indicated by the thick gray line in Fig. 1). While the Wall no longer exists during our sampling period, it is entirely conceivable that property markets on either side were on different price trends during our study period due to massive place-based policies in the aftermath of the German Reunification, idiosyncratic preferences that developed during the division period and, not least, because the Wall often followed actual geographic barriers such as rivers and canals that still persist. Thus, we set M_i to different values, depending on whether a transaction i falls into former West Berlin or East Berlin (or the rest of Germany). When predicting the price index for the exemplary target (the red dot), the algorithm will now take into account that transactions on the other side of the wall may be on a different time trend than those on the same side of the wall that belong to the sub-market of the target location (the cross-hatched area in which the red dot is located). In the same way, the algorithm will control for a sub-market-specific trend within the cross-hatched area when predicting the price index for any of the targets on the other side of the Wall (black dots within the hatched area). This procedure implies that our spatial price index can feature discontinuities at the Berlin Wall in levels and trends.

The user can easily adopt this procedure to their own definitions of sub-markets M_i, for example by assigning different values to transactions that are separated by rivers, mountains, or railroads.

3. Application

To illustrate the functionality of our prediction method that can be applied to spatial contexts in any country, we use rich point-pattern-based housing market data for Germany—described in Section 3.1—to derive house price and rent indices for local labor markets (Section 3.2) and postcode areas (Section 3.3). The latter proxy neighborhoods in Table 1. We relegate further applications to other spatial units to the appendix. Section 3.4 offers a validation check of our approach.

3.1. Data

We rely on highly detailed information on properties listed for rent and purchase. The data are provided by *Immoscout24* via the FDZ-Ruhr. We observe about 22.6 million objects listed for rent and about 21.2 million properties listed for purchase over the period 2007–2021. Since we only observe *asking prices*, it is important to understand their connection to what we are ultimately interested in, i.e. sales prices. The main concern is that asking prices are nothing but a strategic instrument for home sellers that set off an auction. In reply to this concern, Han and Strange (2016) point out that a nontrivial share of sales still settled with a price equal to the asking price and then introduce a search model to further rationalize why asking prices are still a commitment. Put differently, their work suggests that asking prices are a useful and relevant predictor of sales prices although we cannot rule out noise or temporal lags in the adjustment of asking prices to changing market conditions. Indeed, there is evidence suggesting that the degree to which asking prices are a good proxy of transaction prices depends on the cycle and the socio-demographics of a housing market (Miller and Sklarz, 1986; Knight et al., 1994; Genesove and Mayer, 2001; Hayunga and Pace, 2017). To further investigate this, Section 3.4.3 exploits a sample of asking prices observed for Berlin.

Reassuringly, we observe that asking prices track transaction prices during our study period, which echoes evidence from Ireland Lyons (2019) and Canada (Han and Strange, 2016).

Beside asking prices, the data set contains the usual property characteristics (e.g. price, date, floor space, etc.) and a text description which we use to extract a range of additional characteristics, e.g. information on the type of heating system. We use the readily accessible scientific use files which are georeferenced at the level of 1 km^2 grid cells in projected units of the ETRS coordinate system.[7] We refer to Schaffner (2021) for a detailed data description and initial steps to clean the data from e.g. duplicate spells. In our analysis, we discard properties with (i) a monthly rental price below 1€/m^2 or above 50€/m^2; (ii) a purchase price below 250€/m^2 or above 25,000€/m^2; and (iii) floor space below 30 m^2 or above 500 m^2. We further drop all listings where the per-m^2 price is less than 20% or more than 500% of the county median. In total, this removes about 5% of all the transactions.

To illustrate the house price index, we use shapefiles from the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie, BKG) representing jurisdictional boundaries in 2019.

3.2. Application I: Local labor markets (LLMs)

3.2.1. Context

Quantitative research where commuting decisions are not or cannot be considered explicitly usually rely on local labor markets (LLMs) that are constructed to minimize inter-regional commuting flows (see Ahlfeldt et al., 2020; Henkel et al., 2021, for recent applications in the German context). We follow the classification by Kosfeld and Werner (2012) who define 141 German LLMs. LLMs can vary greatly in size which results in sizeable variation in average commuting costs. For the interpretation of naive averages of prices or rents within LLMs, this is a problem because it is well established that households trade housing against commuting costs (Alonso, 1964). To disentangle housing from commuting costs, Combes et al. (2019) propose to compute housing costs at the center of the city, where —assuming a monocentric city structure—commuting cost are zero.

3.2.2. Parameter choices

We follow (Combes et al., 2019) and argue that a theory-consistent index that captures pure housing cost in a LLM should control for a parametric distance gradient that captures commuting costs in the spirit of the monocentric city model (see Alonso, 1964; Mills, 1967; Muth, 1969). For the spatial window, we use the following parameter values $\{A^1 = 25, A^2 = 50, A^3 = 75, A^4 = 100, N^A = 10{,}000\}$, i.e. we consider a commuting zone of 25 km from the center and only revert to larger distances if we do not meet the minimum number $N^T = 10{,}000$ observations. Since we wish to capture the price level in the entire commuting zone (albeit adjusted for commuting cost), we employ the same distance thresholds for the spatial fixed effect, knowing that in most iterations the fixed effect will be dropped: $\{T^1 = 100, T^2 = 100, T^3 = 100, T^4 = 100, N^T = 0\}$.

3.2.3. Results

We present our results for the years 2007 and 2021 in Fig. 2. Panels (a) and (b) depict prices for purchases while panels (c) and (d) are based on rental prices. The indices clearly reveal an increase in both the levels and the spatial dispersion of prices. The LLM *München* was leading the list in terms of purchase prices with 3,750€ (2007) and 12,495€ (2021) while *Prignitz* (710€, 2007) and *Lüchow-Dannenberg* (1,287€, 2021) had the lowest prices per square meter. *Berlin* developed most dynamically with an annual growth rate over the period of 9.2% while prices rose by only 1.2% per year in *Hagen*. Describing regional disparities in house prices based on the coefficient of variation, our index implies an increase in inequality by 53.4% between 2007–2021.

Turning to rental prices, *München* was the most expensive local labor market in both 2007 (12.13€) and 2021 (21.68€). *Emden* (3.68€, 2007) and *Lüchow-Dannenberg* (5.09€, 2021) were at the lower end of the ranking. Rents have grown by an annual rate of 6.5% in *Berlin* while they declined by 0.3% in the LLM *Mecklenburgische Seenplatte*. Rental price dispersion has increased by 30.0% over this period.

3.3. Application II: Postcodes

3.3.1. Context

The smallest administrative units, municipalities, provide great spatial granularity outside independent cities (*kreisfreie Städte*). However, they lack spatial detail within cities as exemplified by the extreme case Berlin, which is one municipality. A suitable spatial unit for the analysis of variation between and within cities are postcodes.[8] There are 8255 postcodes that are designed to accommodate similar populations, but they may vary substantially in terms of geographic size. Within urban areas they can be small and correspond to neighborhoods; in rural areas they can be larger than municipalities. As more data become available at finer grids, postcode-level precision will become an option to zoom into German cities (restricted access labor market data from the Institute for Employment Research are not yet available at this level). Of course, disaggregate property price and rent data at the neighborhood level are useful in their own right since they can inform hedonic regressions that are typically employed to value (dis)amenities.

3.3.2. Parameter choices

We mainly face an empirical problem of predicting an index for a relatively small area within which there will typically not be enough observations to estimate a credible conditional mean. As before, we overcome this limitation by using observations from neighboring municipalities. Specifically, we allow for the following choices for thresholds: $\{A^1 = 10, A^2 = 25, A^3 = 50, A^4 = 100, T^1 = 2.5, T^2 = 5, T^3 = 10, T^4 = 20\}$ (all in km) and we require a minimum of $\{N^A = 10{,}000, N^T = 1000\}$ transactions. These choices allow for a tight local fit in areas where the density of transactions is high while ensuring that the LWR are run on a sufficiently large sample in areas that are more sparsely populated. Note that the small value of T^1 reflects that within urban areas postcodes can be very small. The small scale fixed effects ensure that we account for large differences in prices that are typically observed within cities over relatively small distances.

3.3.3. Results

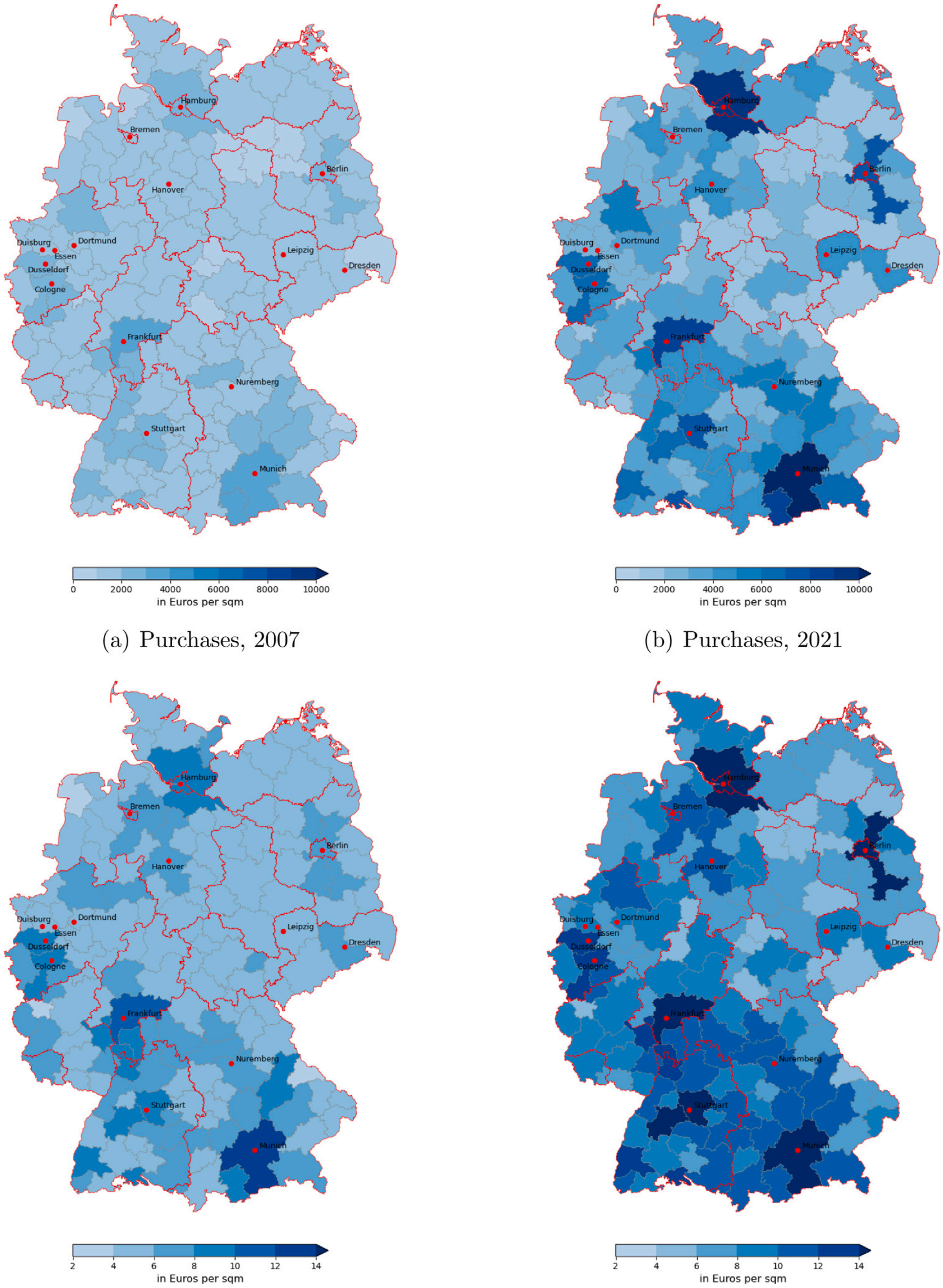
Fig. 3 provides an overview of 8255 postcode areas in Germany following the previous structure. Cities and locations with abundant recreational amenities appear at the top of the ranking whereas remote and low-density places are characterized by the low prices. Overall, purchase prices have changed between −1.1% to 11.0% per year implying an increase in inequality (coefficient of variation) of 57.5%. On the rental market, annual prices have changed between −2.6% and 7.6% translating into an increase in inequality of 22.0%.

3.4. Validation

In this section, we provide three validation tests that concern the properties of the algorithm and the data that we use.

[7] Address-based geocodes are accessible on site at RWI: (RWI and Immobilienscout24, 2021a,b,c,d).

[8] Note that this problem is less common in other countries where data are available for census tracts. However, long-lasting protests against census collections mean that census data become very patchy after 1971—the next waves are 1987 and then 2011—and census tracts are not consistently assigned.



(a) Purchases, 2007

(b) Purchases, 2021

(c) Rents, 2007

(d) Rents, 2021

Fig. 2. Local labor markets. Note: Unit of observation in panels (a)–(d) is 141 local labor markets as defined by Kosfeld and Werner (2012). We report correlation coefficients between purchase and rent indices in Appendix C.

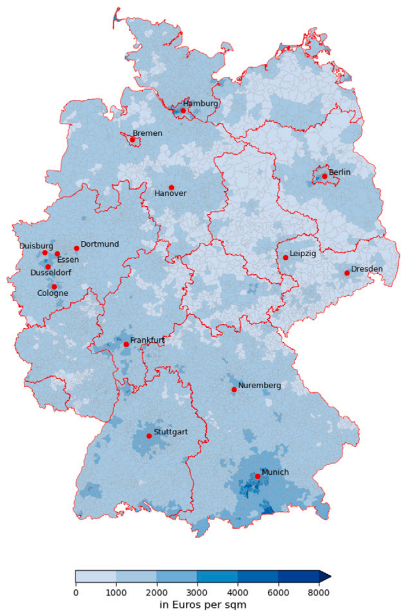
3.4.1. Spatial discontinuities

The algorithm is designed to allow for spatial discontinuities in levels and trends between sub-markets to be defined by the user. As we discuss in Section 2.2.6, we illustrate this approach using the Berlin Wall and assign different values to the sub-market variable M_i, depending on whether a transaction happened in the former East or West of Berlin or outside of Berlin. In Fig. 4, we use a conventional spatial boundary discontinuity set-up to inspect if there is discontinuous variation as one approaches the Berlin Wall coming from the west (negative distance values correspond to a location in former West Berlin). Interestingly, we see a positive discontinuity associated with a location on the Eastern side of the former Berlin Wall in 2007. This is consistent with an ambitious and successful urban renewal program that targeted the eastern parts of the city from the 1990s and 2000s (Ahlfeldt et al., 2017). Once the visible traces of the division period had been removed, place-based policies have become less spatially biased and consequently, the discontinuity fades away as time proceeds. The important takeaway from Fig. 4 is that the definition of sub-markets allows users to prevent the algorithm from smoothing across boundaries where discontinuities are to be expected.

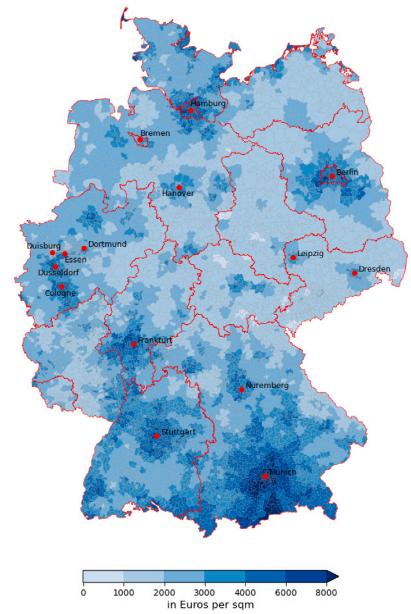
3.4.2. Out-of-sample vs. within-sample prediction

In this section, we subject our micro-geographic indices to a fairly demanding out-of-sample prediction exercise based on 500 × 500 m hexagons for Berlin. The idea is to use data from a fraction of these micro grids to predict our index for the remaining ones.

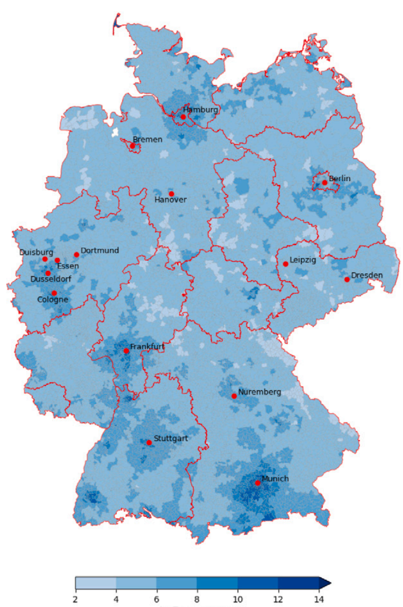
This is an interesting exercise because our algorithm is designed to fit a conditional mean non-parametrically in densely populated



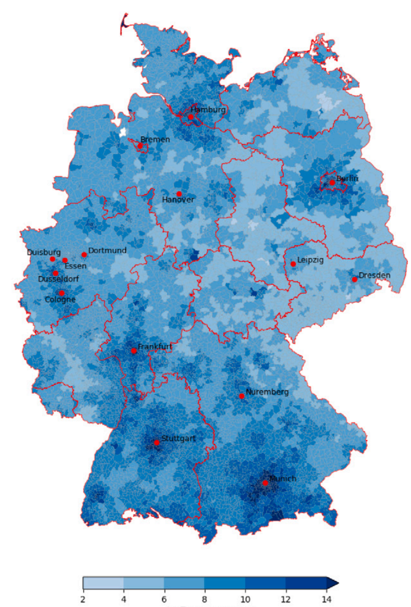
(a) Purchases, 2007



(b) Purchases, 2021



(c) Rents, 2007



(d) Rents, 2021

Fig. 3. Postcodes. Note: Units of observation in panels (a)–(d) are 8255 postcode areas in Germany. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie). We report correlation coefficients between purchase and rent indices in Appendix C.

areas while it extrapolates spatial trends to predict index values in sparsely populated areas. We claim that the latter feature results in strong out-of-sample predictive power which is essentially why we trust our algorithm to fill gaps on a map of index values that would otherwise remain blank. Before we can recommend the algorithm for other applications, it is useful to test its out-of-sample predictive power. To this end, we focus on the 500-meter grid for Berlin and drop about three quarters of the hexagons. Specifically, we design the sampling such that we leave at least one queens contiguity buffer between any estimation hexagon and the nearest overidentification hexagon. Fig. 5 illustrates the sampling design.

Next, we re-run the algorithm on property transactions keeping only this one quarter of Berlin hexagons (estimation sample) and predict the index for the other three quarters of hexagons (overidentification sample). Fig. 5 visualizes the index based on this drastically reduced estimation sample. Evidently, there is a close resemblance to the index estimated on the full sample in Section A.3. We find a convincingly tight fit along the 45-degree line between the within-sample predictions and the out-of-sample predictions across all hexagons in the overidentification sample in Fig. 6. It is reassuring to see that the algorithm does a good job predicting values in areas with sparse data.

3.4.3. Asking vs. transaction prices

In this section, we present a validation test that concerns the reliability of the asking price data we use. To this end, we use the universe of residential condominium transactions in Berlin observed from 2007–2017 which we obtained from the committee of valuation experts (*Gutachterausschuss für Grundstückswerte*). For a discussion of these data,

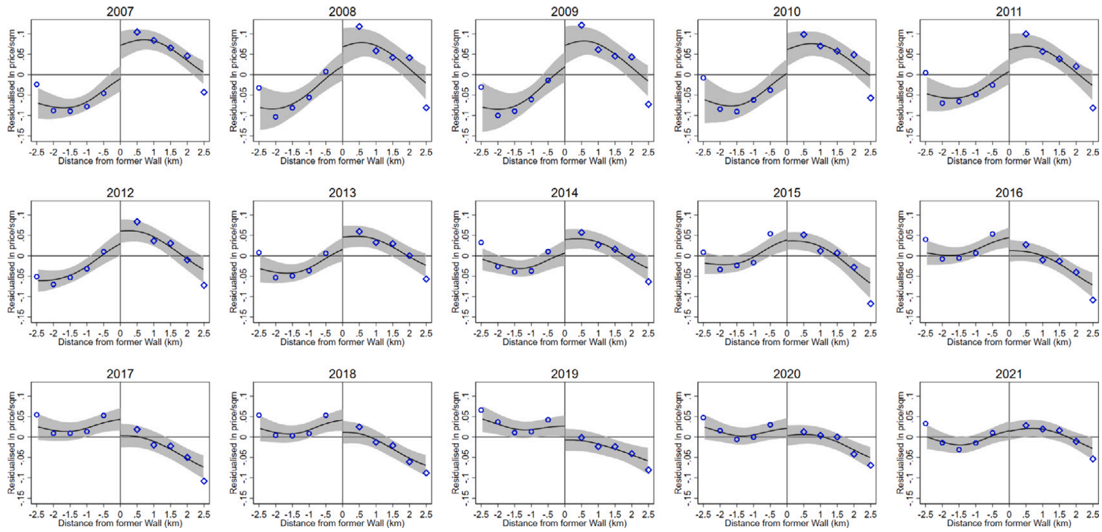
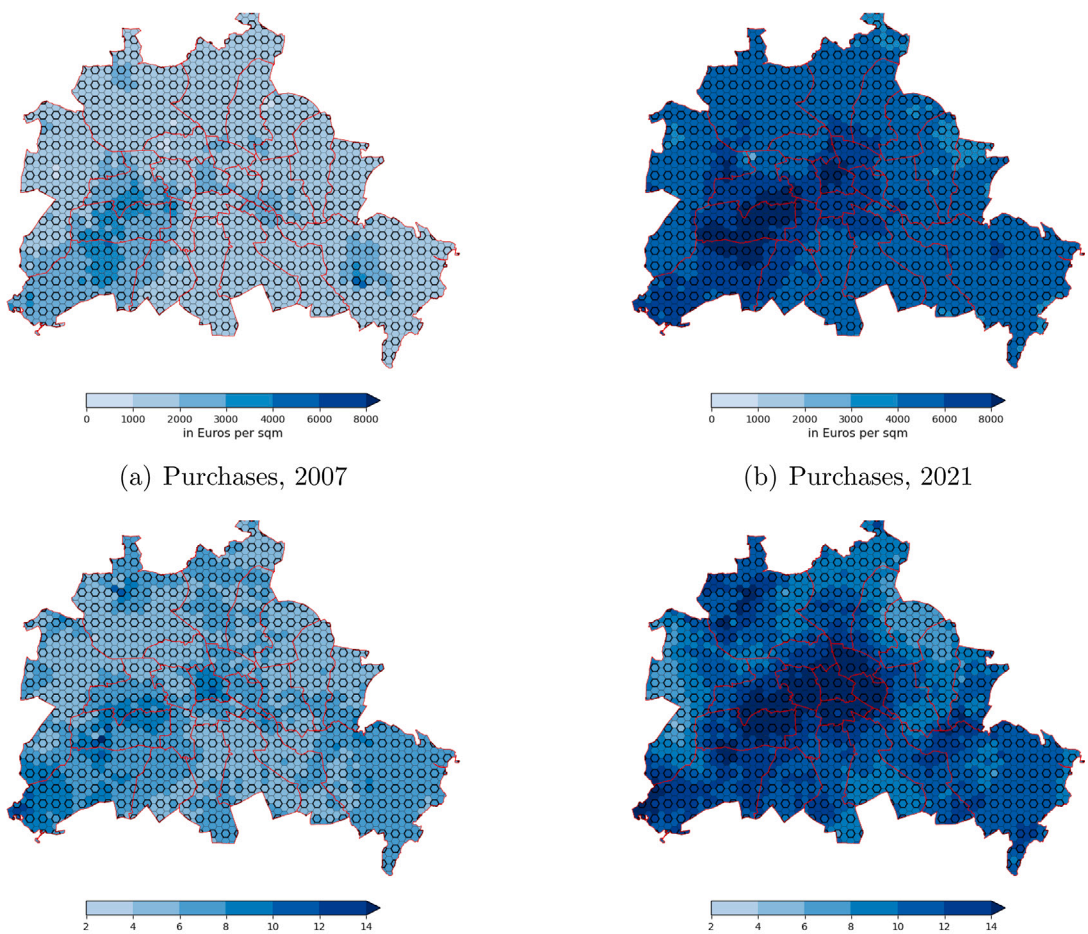


Fig. 4. Spatial discontinuities at Berlin Wall. Note: We restrict the sample to transactions within 2.5 km of the former Berlin Wall within the Berlin territory. The running variable “distance from former Wall” takes negative values for transactions in former West Berlin. We residualize predicted log *purchase prices* in a regression of log price against fixed effects defined for combinations of years and 1-km bin in the Y-coordinate. Point estimates (black solid lines) and 95% confidence intervals (gray shaded areas) are from separate local polynomial regressions of residualized log property prices against the running variable on both sides of the threshold. Local polynomial regressions are of degree 0 and use a Gaussian kernel with bandwidth of 0.25. Blue dots are. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



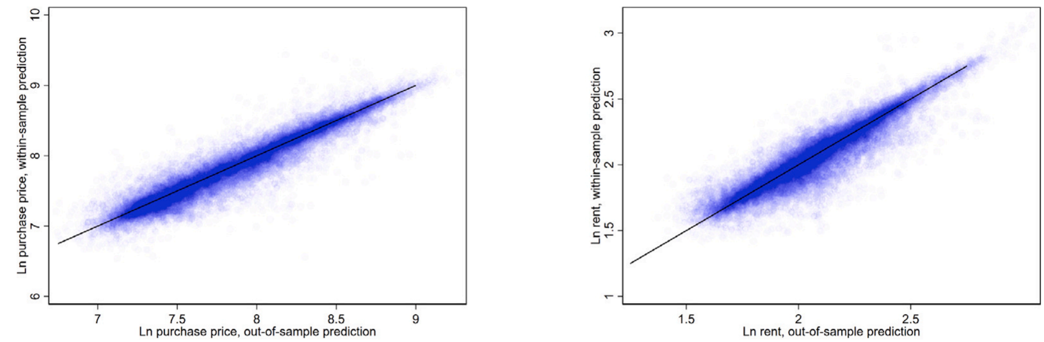
(a) Purchases, 2007

(b) Purchases, 2021

(c) Rents, 2007

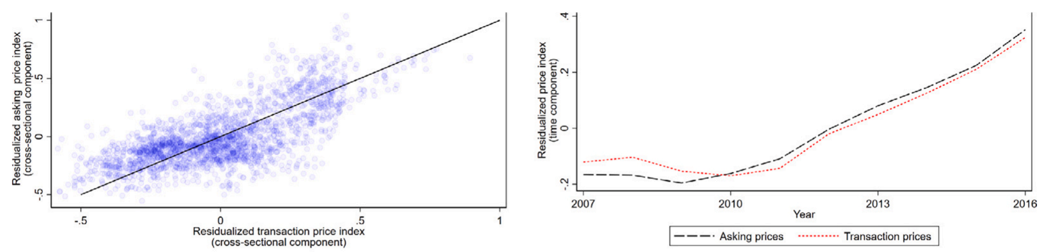
(d) Rents, 2021

Fig. 5. Hexagons Berlin: Validation. Note: The estimation is based on the black hexagons (25% of 1953 units). The index is predicted for all other hexagons. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).



(a) Purchases (b) Rents

Fig. 6. Validation Exercise–Overidentification. Note: Unit of observation in panels (a)–(d) are 1953 hexagons with a diameter of 500 m in Berlin. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie). Hexagons are based on own calculations. The figures show the correlation between out-of-sample predictions for purchase prices (a) or rents (b) and actual prices. The line is the 45-degree line.



(a) Cross-sectional correlation (b) Time trends

Fig. 7. Asking price and transaction price indices for Berlin. Note: We generate a transaction price index using actual transaction prices from Berlin observed from 2007 to 2016 for the hexagons shown in Fig. 6 using the approach and parameter values discussed in Sections 2.1 and 2.2 (the same as we apply to asking prices. To decompose the transactions price and the asking price index (using years 2007–2016) into cross-sectional and time components, we run auxiliary regressions of the predicted log price per square meter against year and hexagon fixed effects. The black solid line in the left panel is the 45-degree line.

we refer to Ahlfeldt et al. (2017). Using these data, we generate a transactions price index for the hexagonal grid cells displayed in Fig. 6 using exactly the same methodology as we applied to asking prices. To evaluate how the transactions price index compares to the asking price index in space and time, we decompose both indices into a cross-sectional and a time component by regressing the log of predicted per-square-meter prices on grid and year dummies. The left panel of Fig. 7 shows that both indices are positively correlated in space. The right panel shows that both indices track each other closely over time. Given that transaction prices are not universally accessible in Germany, the arguably most important insight from Fig. 7 is that asking prices provide a decent approximation of market trends, which echoes evidence from Ireland Lyons (2019) or Canada (Han and Strange, 2016). While we view the methodology as our primary contribution, this finding also lends validity to the asking price indices we create. Indirectly, the fact that we generate similar indices from completely different data sources also lends some validity to our proposed methodology.

4. Novel stylized facts

The application of the algorithm introduced in Section 2 to micro-geographic data on rental and purchase prices in Section 3 has generated a real estate data set that is unprecedented in terms of spatial detail and coverage of the German buyer and renter markets. In this section, we provide a first exploration of this data set with the aim to uncover stylized facts that may motivate further research.

4.1. Consistent spatial windows in rent and price indices

The algorithm we have introduced in Section 2.1 endogenously identifies a spatial window $A^j \in \{A^1, A^2, A^3, A^4\}$ and a fixed effect threshold parameter $T^j \in \{T^1, T^2, T^3, T^4\}$. Depending on the local density of observations, $\{A^j, T^j\}$ may differ in a purchase price and a rental price index. Since we are interested in comparing purchase price and rental price indices in this section, we use a version of the algorithm that takes $\{A^j, T^j\}$ as given. We select these parameters according to the following simple rule:

$$\bar{A}^j = \max\left(A^j_{purch}, A^j_{rent}\right)$$
$$\bar{T}^j = \max\left(T^j_{purch}, T^j_{rent}\right),$$

where *purch* and *rent* index parameter values retrieved from the applications of the original algorithm to the purchase and rental markets. Feeding $\{\bar{A}^j, \bar{T}^j\}$ into the algorithm that takes these parameters as given, but is otherwise identical, we then obtain purchase price and rental price indices that are suitable for the evaluation of price-to-rent ratios since they are based on exactly the same spatial windows. In practice, this precautionary measure turns out to be computationally heavy, but not particularly impactful since the algorithm chooses similar values for $\{A^j, T^j\}$ in the purchase and rental price sample (see Table A 1 in the appendix).

4.2. Cross-sectional correlations

A large body of literature has established that a broad range of locational features such as accessibility, natural amenities, or neighborhood quality capitalize into property values (see Cheshire and Sheppard, 1995; Ahlfeldt, 2011, for typical examples). The intuition is straightforward. The standard urban model predicts that, assuming perfect mobility, any locational advantage is offset by a correspondingly higher cost of housing to maintain a constant utility within the city. The monocentric city model focuses on commuting cost as approximated by distance from an exogenous central business district (CBD) (Alonso, 1964; Mills, 1967; Muth, 1969), but the logic extends to any other amenity (or disamenity).

With this in mind, we correlate our rent and price indices with several locational characteristics in Fig. 8. We provide bin scatter plots based on percentiles for a clearer presentation, but the underlying data comprise 8255 postcodes and the entire country, which is a fairly broad coverage within a literature that mostly focuses on particular cities (see Hill, 2013, for a survey). We look at three different dimensions which roughly represent a locations' attractiveness due to its (i) proximity to economic activities; (ii) consumption amenities and possibilities to interact socially; and (iii) natural amenities. To measure proximity to economic activities, we use the distance (in km) from the CBD of the local labor market area that nests the postcode. To measure a location's supply of consumption amenities, we use the number of geo-tagged photos shared in social media in a postcode. Overall, we are using 1.5 million pictures taken in the early 2010s. The measure captures visually appealing content (e.g. landmarks or scenic views) but also locations like bars and restaurants where people like to socialize. For further detail, we refer to Ahlfeldt et al. (2020). Third, to approximate natural amenities, we calculate the Vegetation Continuous Fields (VCF) product using Google Earth Engine (DiMiceli et al., 2017). Based on satellite images over the period 2000–2014, the measure approximates the percentage share of an area (here postcodes) covered by trees. We condition this measure on a Normalized Difference Vegetation Index (NDVI) so we compare regions with an equal degree of vegetation but different degrees of tree coverage. We think of the measure of tree coverage as a proxy for access to natural amenities like forests or leafy parks. Notice all panels show partial correlations of prices or rents and amenity measures that are regression-adjusted. Specifically, we condition each measure on all the other amenity measures and further absorb local LLM effects.

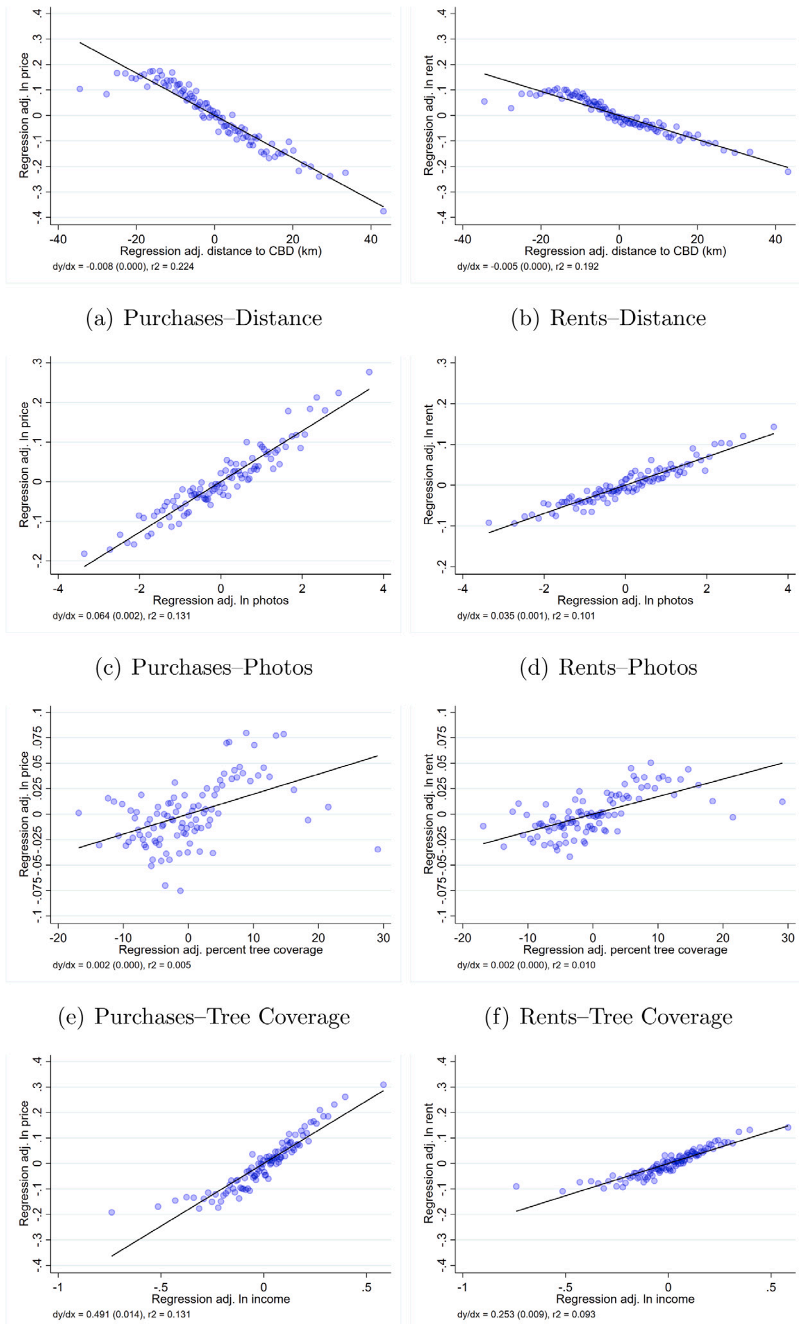
Although the cross-sectional multivariate regression is a workhorse tool in the hedonic price literature, we caution against causal interpretations of the partial correlations since there may be omitted variables correlated with the covariates we consider. In fact, distance from the CBD, by its very nature, is supposed to capture a multitude of factors that make traveling to city centers worthwhile, for professional and recreational purposes. Likewise, we use photos as a "big data" proxy for many factors that make places amenable to social interactions. We do not claim that the mere fact that someone shares a picture on social media adds to the value of a location, at least not in a quantitatively relevant way. Yet, the partial correlations are interesting because most examples in the literature focus on purchase prices within individual cities, whereas we compare hedonic implicit prices from purchase prices and rental prices covering an entire country. Hence, we report the marginal effects estimated from the underlying raw data ($\partial y/\partial x$) along with the standard errors (in parentheses) and the *partial* R^2.

In line with the predictions of the monocentric city model, prices (a) and rents (b) decrease as we are moving away from the CBD suggesting that people value living close to the center of economic activity. The respective slope coefficients suggest that prices (rents) decrease by 8% (5%) for every 10 km further away from the CBD. Panels (c) and (d) show a positive and also tight correlation between the number of photos taken and prices or rents, respectively. A 10-percent increase in the number of photos taken implies a 0.6% (0.3%) increase in price (rents), underlining the amenity value of proximity to social interactions. Since we are holding distance from the CBD constant, the significant effect of the photo variable reveals that the geography of consumption amenities is—unsurprisingly—not perfectly approximated by a linear distance gradient. Panels (e) and (f) show a positive though more noisy relationship between the percentage of tree coverage (conditional on overall vegetation) and prices or rents. At face value, the correlations suggest that a 10 percentage point increase in tree coverage increases prices (rents) by 0.02% (0.017%) percent. We attribute the high level of noise in part to the imprecisely measured tree coverage from satellite images with a 250 m spatial resolution and in part to tree coverage in the postcode being an incomplete measure of access to natural amenities.

The perhaps most interesting stylized fact that jointly emerges from panels (a)–(f) in Fig. 8 is that the point estimates we obtain for the purchase price specifications are of consistently greater magnitudes than those obtained for rents. This implies a relatively larger capitalization of amenity values into purchase prices. To our knowledge, this is a new stylized fact (at least for Germany) that we uncover with the help of our new micro-geographic house price indices. Within the standard framework of financial economics, we can rationalize this stylized fact via variation in expected rental growth or risk. Landlords and home-buyers may expect the value of amenities to rise and, thus, be willing to pay a premium in the expectation of greater (imputed) rents in the future. They may also understand that future demand shocks will trigger larger price adjustments in supply-inelastic markets (Büchler et al., 2021). Alternatively, properties in better locations may be perceived as safer assets because neighborhoods with high amenity values tend to be more stable over time as documented by Lee and Lin (2018). Denser places also tend to be more liquid, which can justify a lower yield. Finally, search frictions can generate non-uniform spatial price-to-rent ratios (Chapelle et al., 2022).

Another interesting variable to correlate property prices with is neighborhood income. Because of residential sorting, income is likely determined by the same variables as purchase prices and rents, including those that we cannot observe. In principle, the correlation can go both ways since preference-based sorting depends on the relative willingness-to-pay of different income groups. Taking the classic example—distance from the CBD, which we have documented to be negatively correlated with prices—the rich will live in the center if they value centrality more than the poor. In the standard model, this will be true if the income elasticity of commuting cost exceeds the income elasticity of housing demand. However, it could also be the other way round, which would result in rich people living on large parcels with plenty of interior and exterior space in leafy suburbs as often observed in North American cities. Empirically, it does not seem as if one force universally dominates the other (Wheaton, 1977). In some classic contributions, it has been assumed that the rich are pulled to the center unless they face a (temporary) advantage in accessing faster transport modes (LeRoy and Sonstelie, 1983). In other classics, the opposite is assumed unless the city center exhibits some amenity value such as an attractive historic fabric (Brueckner et al., 1999). For German cities, our expectation is that central cities are generally relatively rich as most downtown areas are fairly vibrant due to a walkable historic urban structure (and often a historic building stock) and public transit is generally well developed throughout metropolitan areas. Assuming that other amenities such as access to natural amenities or urban consumption amenities are normal goods, we would expect demand to increase if the income elasticity is larger than one. Hence, we have the rather unambiguous expectation that income and real estate prices should be positively correlated in Germany.

Using average disposable household income at the postcode level, which we obtain from the *GfK* (*Gesellschaft für Konsumforschung*), we find this expectation to be met by evidence in panels (g) and (h) of Fig. 8. In these panels, we condition on LLM fixed effects, but not on other amenities, because income is an endogenous variable that itself depends on amenities. The estimated slope coefficients $\partial y/\partial x < 1$ are expected because richer households only spend part of the greater expenditure on consuming housing services of greater quality (better location) whereas the other part will go into quantity (bigger houses). Yet, the elasticity that relates the log of purchase price to the log of neighborhood income is about twice as large as the respective elasticity for the log of rent. Given that home buyers and renters do not differ as dramatically in social strata in Germany as in many other countries, the difference in the elasticities is difficult to reconcile with differences



(a) Purchases–Distance (b) Rents–Distance

(c) Purchases–Photos (d) Rents–Photos

(e) Purchases–Tree Coverage (f) Rents–Tree Coverage

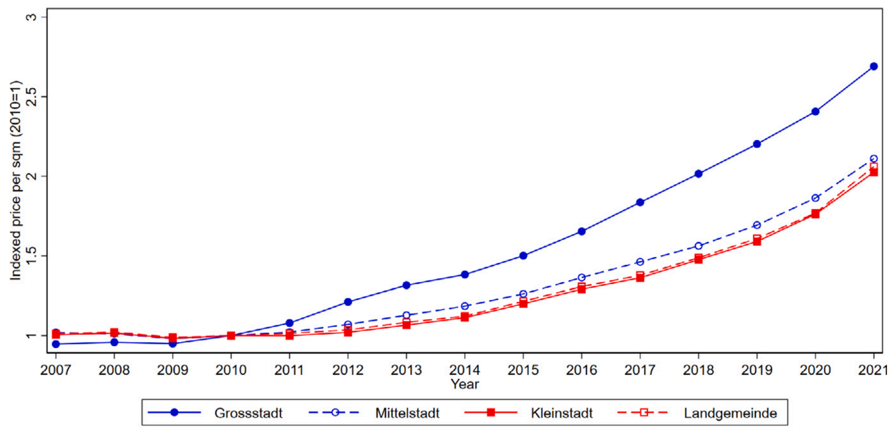
(g) Purchases–Income (h) Rents–Income

Fig. 8. Stylized facts I—Cross-sectional correlations. Note: Unit of observation in panels (a)–(h) are 8255 postcodes. The figures shows the correlation between (i) distance (in km) to the CBD as a proxy for access to economic activities; (ii) the log number of photographs taken as a measure for consumption amenities (iii) tree coverage (in percent) as a measure for the presence of natural leisure time amenities; and (iv) income, with prices (left panels) or rents (right panels).

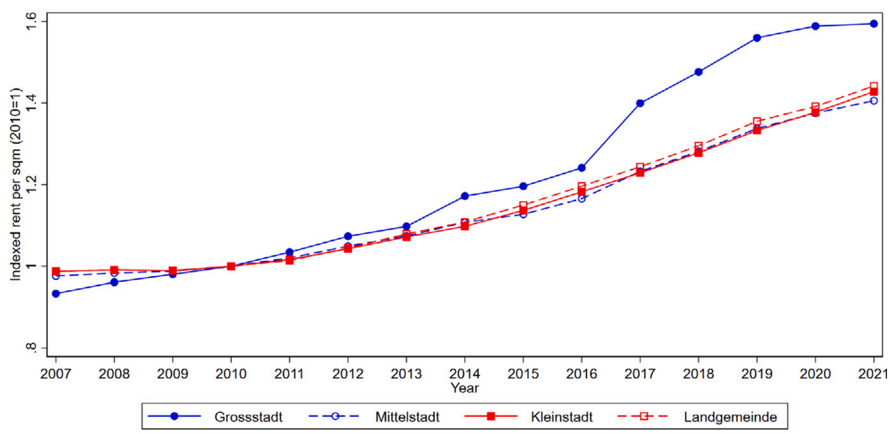
in consumption preferences alone. A plausible alternative explanation is that home buyers spend relatively more of their higher income on home quality (rather than quantity) because of the greater risk-adjusted return they expect in better neighborhoods. In any case, it appears that a deeper exploration of the spatial determinants of price-to-rent ratios in Germany is a promising area for research.

4.3. Temporal trends

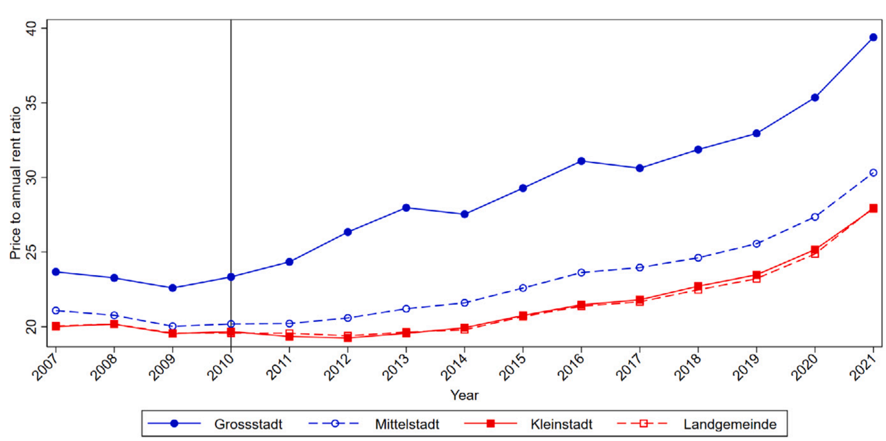
Having explored cross-sectional differences in buyer prices and rental prices, we now turn to the temporal dimension of the new indices. First, we show time series graphs of price and rent data for the period 2007–2021 for four different types of cities, large cities (*Großstadt*), small cities (*Kleinstadt*) and rural areas (*Landgemeinden*) in panels (a) and (b) of Fig. 9. We index the respective time series to 2010. The first insight is that there is no equivalent to the U.S. subprime mortgage crisis in Germany. To the contrary, low interest rates in the aftermath of the European Sovereign Debt Crisis and the lack of global investment opportunities triggered a steep increase in prices which was not matched by a corresponding increase in rents, at least initially. Computing the ratio between the buyer price and rental price indices (prior to normalization), panel (c) confirms that price-to-rent ratios were generally on the rise since 2009. This, by itself, is not a particularly striking finding since lower mortgage interest rates reduce



(a) Purchases

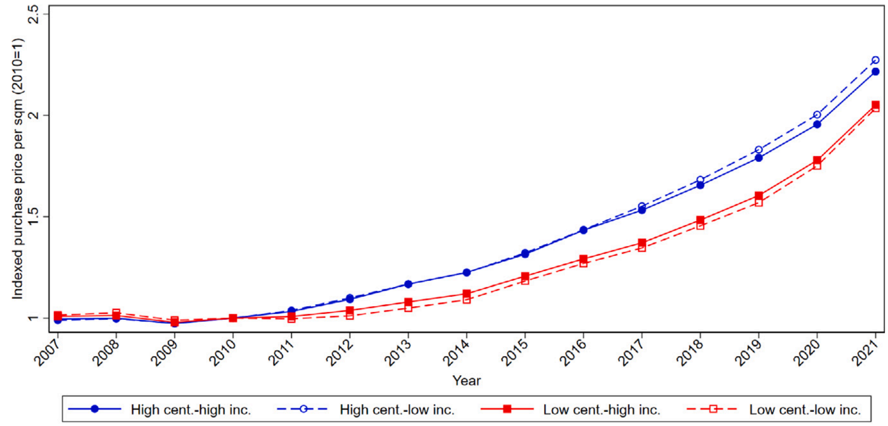


(b) Rents

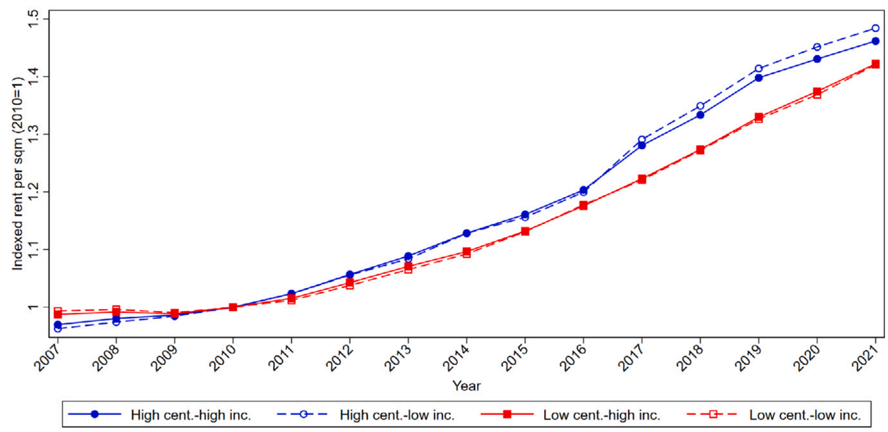


(c) Price-to-rent ratio

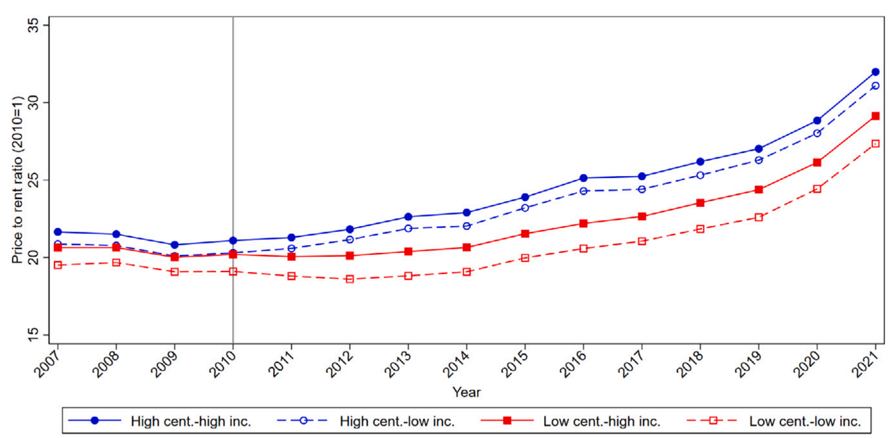
Fig. 9. Purchase price and rent trends between cities. Note: Settlement types are defined at the level of municipal associations. The figures shows the development of prices (a) rents (b) and the price-to-rent ratio (c) over the years 2007–2018 for four different types of cities, large cities (Grossstadt), medium sized cities (Mittelstadt), small cities (Kleinstadt) and rural areas (Landgemeinden).



(a) Purchases



(b) Rents



(c) Price-to-rent ratio

Fig. 10. Purchase price and rent trends within cities. Note: Centrality and income measured at the postcode level. High (low)-centrality postcodes are postcodes with a below (above) median distance from the CBD (normalized by the mean distance within LLMs). High (low)-income postcodes are postcodes with an above (below) median income (normalized by the mean income within LLMs). All trends are the averages across all postcodes within a centrality-income group. In panels (a) and (b), indices are normalized to have a mean of one in 2010.

the cost of capital, mapping to higher initial investments at constant rents and yields. The diverging trend in denser areas is, however, an interesting stylized fact.

Price-to-rent ratios averaged at about 25 in 2010, a much higher level than in the less agglomerated parts of the country. This is consistent with the stylized evidence from within cities introduced in Section 4.2 which points to price-to-rent ratios that are generally higher in more expensive areas, be it because buyers expect greater returns in the future, or lower risk.[9] Fig. 9 adds that the divergence of price-to-rent ratios in denser areas has increased over time. One interpretation is that rational forward-looking investors (Clayton, 1996), starting from 2010, adjusted their already positive expectations for rental growth upwards. This would be consistent with growing demand for density in Germany (Ahlfeldt et al., 2020) and elsewhere (Couture and Handbury, 2020). Indeed, panel b) reveals that rental growth accelerated in large cities a couple of years later.

Given the large capital inflows into the German real estate market, which represented one of the few ”safe havens” past the U.S. subprime mortgage crisis and the European Sovereign Debt crisis, it is also tempting to connect the surge in the price-to-rent ratio in large cities to foreign investment (Badarinza and Ramadorai, 2018). If foreign investments are biased towards larger cities, be it because of these markets are more liquid, less fragmented in terms of ownership, or simply because they are “on the map”, an inflow of international capital will tend to reinforce spatial differentials in the price-to-rent ratio.

Hilber and Mense (2021) argue that increases in the price-to-rent ratio can be triggered by expectations that are formed based on stronger positive responses to positive demand shocks in supply-inelastic markets. While large German cities are plausibly more supply-inelastic than smaller cities, the divergence of trends in buyer and rental prices did not start in a high-growth environment, suggesting a role for alternative explanations in the German context. Indeed, a perhaps more obvious explanation could be tightening rental price regulations in the largest German cities in the second half of the 2010s (Mense et al., 2019; Breidenbach et al., 2022). The sudden increase in rents towards the end of the 2010s could reflect a change in composition as newly constructed units were exempt from the new regulations.

Bridging the gap between Fig. 8, which considers variation in prices in a cross-section within cities, and Fig. 9, which considers variation in prices between cities over time, we look into price-to-rent ratios over time within cities in Fig. 10. To this end, we distinguish postcodes along two dimensions: Centrality and income. The defining criteria are simply whether a postcode is above or below the median distance from the CBD or the median disposable household income within its host LLM. Fig. 10 presents normalized buyer and rental price trends as well as the average price-to-rent ratios for the two-by-two combinations of these attributes. Otherwise, the presentation follows Fig. 9. The main insight from panels (a) and (b) is that centrality is the primary determinant of the trend in buyer prices and rental prices within cities. As with the between-city comparison, denser areas within cities appreciated faster. While for the level of the price-to-rent ratio, centrality is important, income also matters. In fact, high-centrality low-income areas had the same average price-to-rent ratio as low-centrality high-income areas up until 2010. After that, centrality starts dominating income as a determinant of the relative growth of purchase prices. Naturally, we can apply the same explanations for the divergence of buyer price and rental price trends as in the between-city comparison. Investors might have been willing to accept higher price-to-rent ratios because they expected greater future returns in central parts of German cities. Indeed, the relative pattern of rental growth in Fig. 10, panel (b) (central vs. non-central) is strikingly similar to the relative pattern of rental growth in Fig. 9, panel (b) (large cities vs. smaller cities). Similarly, spatially biased foreign investment could rationalize the pattern given that central cities are generally more liquid markets, have favorable building stock (more multi-storey buildings), and are likely better known to non-local investors. Finally, rental price regulations in the largest cities are more binding in the central, most expensive (as shown in Section 4.2), parts of the largest cities. In any case, uncovering the determinants of the spatial bias in the price-to-rent ratio in levels and trends appears to be a promising research area. Germany may be of interest in international comparison given a home ownership rate that is low by the standards of similarly developed countries. Our indices represent an asset to those wishing to embark on this mission.

[9] Previous work reports similar patterns for e.g. London (Halket et al., 2020; Bracke, 2015), Paris (Chapelle et al., 2022) or Shanghai (Chen et al., 2022) where the price-to-rent ratios are higher in denser and thus typically more expensive locations. The important difference is that we are looking at a temporal dimension and consider an entire country instead of selected cities.

5. Conclusion

This paper introduces a new algorithm that transforms prices of geolocated property transactions into a mix-adjusted balanced-panel house price index for arbitrary spatial units. While the spatial units can be of arbitrary size, the aggregation method itself is not arbitrary but well founded in urban economic theory and spatial methods. The strength of the algorithm is that it combines parametric and non-parametric estimation techniques to provide a tight local fit where data are abundant, and reliable extrapolations where data are sparse.

We will publish the underlying prediction algorithm along with suggestions for the critical parameter choices together with this paper. This allows other researchers who have access to individual property transaction data to employ our method and create their own indices. Our exemplary application further generates spatial price indices that are unique in their micro-geographic coverage of the German buyer and renter market since 2007. We believe that the algorithm and indices published along with this paper will facilitate applications of quantitative spatial models which have been held back by suitable real estate data that combine micro-geographic variation and comprehensive coverage. We also hope that the observed divergence of price-to-rent ratios in more densely populated areas will spur research into the underlying determinants, possibly using our data sets.

The use case for housing policy might be even stronger. Just to name a few potential applications, our indices could inform policy makers about the success of urban development, renewal, or heritage preservation measures, housing affordability issues, or emerging bubbles. The latter is key to assessing future risks to financial stability, which falls under the domain of the European Systemic Risk Board (ESRB) since 2010. Taking Germany as a case in point, the perhaps most obvious application would also have the highest impact. Government bill 19/26918, posted in February 2021, discusses a reform of regulations regarding local rent indices. The underlying motivation for the reform is that existing rent indices are often not up-to-date and lack a proper theoretical foundation. This has consequences for the assessment of rent control policies and legal disputes over rent price increases as part of the comparative rent control system. The issue is of some urgency as stressed by politicians from both sides of the political spectrum. As an example, Johannes Fechner of the social democratic party (SPD) recently criticized that 80 of the 200 largest German cities failed to publish the mandatory rent indices. Corroborating this criticism, Jan-Marco Luczak of the conservative CDU called for an academically founded rent index. Our point-pattern based algorithm that is based on insights from decades of economics research offers a readily and universally applicable solution to the problem of creating a comprehensive micro-geographic rent index for Germany.

Data availability

The micro data underlying the construction of German indices is confidential. We provide a public code directory that contains the core code and a synthetic data set. It also contains instructions on how to access the raw data. The public code directory can be found at https://doi.org/10.7807/immo:red:ahs:v1.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.regsciurbeco.2022.103836.

References

Ahlfeldt, Gabriel M., 2011. If Alonso was right: Modeling accessibility and explaining the residential land gradient. J. Reg. Sci. 51 (2), 318–338.

Ahlfeldt, Gabriel M., Bald, Fabian, Roth, Duncan, Seidel, Tobias, 2020. Quality of life in a dynamic spatial model. CEPR Discussion Paper, 15594.

Ahlfeldt, Gabriel M., Maennig, Wolfgang, Richter, Felix J., 2017. Urban renewal after the berlin wall: A place-based policy evaluation. J. Econ. Geogr. 17 (1), 129–156. http://dx.doi.org/10.1093/jeg/lbw003.

Ahlfeldt, Gabriel M., Pietrostefani, Elisabetta, 2019. The economic effects of density: A synthesis. J. Urban Econ. 111 (February), 93–107. http://dx.doi.org/10.1016/j.jue.2019.04.006.

Ahlfeldt, Gabriel M., Redding, Stephen J., Sturm, Daniel M., Wolf, Nikolaus, 2015. The economics of density: Evidence from the Berlin Wall. Econometrica 83 (4), 2127–2189. http://dx.doi.org/10.3982/ECTA10876.

Albouy, David, Lue, Bert, 2015. Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life. J. Urban Econ. 89, 74–92. http://dx.doi.org/10.1016/j.jue.2015.03.003, URL http://www.sciencedirect.com/science/article/pii/S0094119015000200.

Allen, Treb, Arkolakis, Costas, 2014. Trade and the topography of the spatial economy. Q. J. Econ. 129 (3), 1085–1140. http://dx.doi.org/10.1093/qje/qju016.

Almagro, Milena, Domínguez-Iino, Tomas, 2021. Location sorting and endogenous amenities: Evidence from Amsterdam. Working Paper.

Alonso, William, 1964. Location and Land Use. Harvard Univ. Press, Cambridge, MA.

Athey, Susan, Imbens, Guido W., 2019. Machine learning methods that economists should know about. Annu. Rev. Econ. 11 (1), 685–725. http://dx.doi.org/10.1146/annurev-economics-080217-053433.

Autor, David H., Palmer, Christopher J., Pathak, Parag A., 2014. Housing market spillovers: Evidence from the end of rent control in Cambridge, Massachusetts. J. Polit. Econ. 122 (3), 661–717. http://dx.doi.org/10.1086/675536.

Badarinza, Cristian, Ramadorai, Tarun, 2018. Home away from home? Foreign demand and London house prices. J. Financ. Econ. 130 (3), 532–555. http://dx.doi.org/10.1016/j.jfineco.2018.07.010.

Bailey, Martin J., Muth, Richard F., Nourse, Hugh O., 1963. A regression method for real estate price index construction. J. Amer. Statist. Assoc. 58 (304), 933–942. http://dx.doi.org/10.1080/01621459.1963.10480679, URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10480679.

de Bellefon, Marie-Pierre, Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, Gorin, Clément, 2021. Delineating urban areas using building density. J. Urban Econ. 125, 103226. http://dx.doi.org/10.1016/j.jue.2019.103226, URL https://www.sciencedirect.com/science/article/pii/S0094119019301032.

Bracke, Philippe, 2015. House prices and rents: Microevidence from a matched data set in central London. Real Estate Econ. 43 (2), 403–431. http://dx.doi.org/10.1111/1540-6229.12062.

Breidenbach, Philipp, Eilers, Lea, Fries, Jan, 2022. Temporal dynamics of rent regulations - the case of the german rent control. Reg. Sci. Urban Econ. 92.

Brueckner, Jan K., Thisse, Jacques-François, Zenou, Yves, 1999. Why is central Paris rich and downtown detroit poor? An amenity-based theory. Euro. Econ. Rev. 43 (1), 91–107. http://dx.doi.org/10.1016/S0014-2921(98)00019-1, URL http://www.sciencedirect.com/science/article/pii/S0014292198000191.

Büchler, Simon, Ehrlich, Maximilian v., Schöni, Olivier, 2021. The amplifying effect of capitalization rates on housing supply. J. Urban Econ. 126, 103370. http://dx.doi.org/10.1016/J.JUE.2021.103370.

Cellini, Stephanie Riegg, Ferreira, Fernando, Rothstein, Jesse, 2010. The value of school facility investments: Evidence from a dynamic regression discontinuity design. Q. J. Econ. 125 (1), 215–261. http://dx.doi.org/10.1162/qjec.2010.125.1.215.

Chapelle, Guillaume, Eyméoud, Jean-Benoît, Bruneel-Zupanc, Etienne, 2022. A Theory of Spatial Interactions. Working Paper.

Charles, Kerwin Kofi, Hurst, Erik, Notowidigdo, Matthew J., 2018. Housing booms, manufacturing decline and labour market outcomes. Econ. J. 129 (617), 209–248. http://dx.doi.org/10.1111/ecoj.12598.

Chay, Kenneth Y., Greenstone, Michael, 2005. Does air quality matter? Evidence from the housing market. J. Polit. Econ. 113 (2), 376–424. http://dx.doi.org/10.1086/427462, URL http://www.journals.uchicago.edu/doi/10.1086/427462.

Chen, Jie, Chen, Yu, Hill, Robert J., Hu, Pei, 2022. The user cost of housing and the price-rent ratio in shanghai. Reg. Sci. Urban Econ. 92, 103738. http://dx.doi.org/10.1016/j.regsciurbeco.2021.103738.

Cheshire, Paul, Sheppard, Stephen, 1995. On the price of land and the value of amenities. Economica 62 (246), 247–267. http://dx.doi.org/10.2307/2554906, URL http://www.jstor.org/stable/2554906.

Clapp, John M., 2004. A semiparametric method for estimating local house price indices. Real Estate Econ. 32 (1), 127–160. http://dx.doi.org/10.1111/j.1080-8620.2004.00086.x.

Clark, Colin, 1951. Urban population densities. J. R. Stat. Soc. Ser. A (General) 114 (4), 490–496. http://dx.doi.org/10.2307/2981088, URL http://www.jstor.org/stable/2981088.

Clayton, Jim, 1996. Rational expectations, market fundamentals and housing price volatility. Real Estate Econ. 24 (4), 441–470. http://dx.doi.org/10.1111/1540-6229.00699.

Cleveland, William S., Devlin, Susan J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. J. Amer. Statist. Assoc. 83 (403), 596–610. http://dx.doi.org/10.2307/2289282, URL http://www.jstor.org/stable/2289282.

Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, 2019. The costs of agglomeration: House and land prices in french cities. Rev. Econom. Stud. 86 (4), 1556–1589. http://dx.doi.org/10.1093/restud/rdy063.

Couture, Victor, Handbury, Jessie, 2020. Urban revival in America. J. Urban Econ. 119, 103267. http://dx.doi.org/10.1016/J.JUE.2020.103267.

Diamond, Rebecca, McQuade, Tim, Qian, Franklin, 2019. The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco. Amer. Econ. Rev. 109 (9), 3365–3394. http://dx.doi.org/10.1257/aer.20181289.

van Dijk, Dorinth W., Geltner, David M., van de Minne, Alex M., 2022. The dynamics of liquidity in commercial property markets: Revisiting supply and demand indexes in real estate. J. Real Estate Finance Econ. 64 (3), 327–360. http://dx.doi.org/10.1007/s11146-020-09782-5.

DiMiceli, Charlene M, Carroll, Mark L, Sohlberg, Robert A, Huang, Chengquan, Hansen, Matthew C, Townshend, John R G, 2017. Annual global automated MODIS vegetation continuous fields (MOD44B) at 250 m spatial resolution for data years beginning day 65, 2000–2014, collection 5 percent tree cover, version 6. University of Maryland, College Park, MD, USA, URL https://lpdaac.usgs.gov/documents/112/MOD44B_User_Guide_V6.pdf.

Dingel, Jonathan I., Miscio, Antonio, Davis, Donald R., 2021. Cities, lights, and skills in developing economies. J. Urban Econ. 125, 103174. http://dx.doi.org/10.1016/J.JUE.2019.05.005.

Duan, Naihua, 1983. Smearing estimate: A nonparametric retransformation method. J. Amer. Statist. Assoc. 78 (383), 605–610. http://dx.doi.org/10.2307/2288126, URL http://www.jstor.org/stable/2288126.

European Commission, 2013. Handbook on Residential Property Prices Indices (RPPIs). Publications Office of the European Union, Luxembourg.

Francke, Marc K., 2010. Repeat sales index for thin markets. J. Real Estate Finance Econ. 41 (1), 24–52. http://dx.doi.org/10.1007/s11146-009-9203-1.

Genesove, David, Mayer, Christopher, 2001. Loss aversion and seller behavior: Evidence from the housing market. Q. J. Econ. 116 (4), 1233–1260, URL http://www.jstor.org/stable/2696458.

Glaeser, Edward L., Gyourko, Joseph, Saiz, Albert, 2008. Housing supply and housing bubbles. J. Urban Econ. 64 (2), 198–217. http://dx.doi.org/10.1016/j.jue.2008.07.007, URL https://www.sciencedirect.com/science/article/pii/S0094119008000648.

Halket, Jonathan, Nesheim, Lars, Oswald, Florian, 2020. The housing stock, housing prices, and user costs: the roles of location, structure, and unobserved quality. Internat. Econom. Rev. 61 (4), 1777–1814. http://dx.doi.org/10.1111/iere.12475.

Han, Lu, Strange, William C., 2016. What is the role of the asking price for a house? J. Urban Econ. 93, 115–130. http://dx.doi.org/10.1016/j.jue.2016.03.008.

Hayunga, Darren K., Pace, R. Kelley, 2017. List prices in the US housing market. J. Real Estate Finance Econ. 55 (2), 155–184. http://dx.doi.org/10.1007/s11146-016-9555-2.

Heblich, Stephan, Redding, Stephen J., Sturm, Daniel M., 2020. The making of the modern Metropolis: Evidence from London. Q. J. Econ. 135 (4), 2059–2133. http://dx.doi.org/10.1093/qje/qjaa014.

Henkel, Marcel, Seidel, Tobias, Suedekum, Jens, 2021. Fiscal transfers in the spatial economy. Am. Econ. J. Econ. Policy 13 (4), 433–468.

Hilber, Christian A. L., Mense, Andreas, 2021. Why have house prices risen so much more than rents in superstar cities?. CEP Discussion Paper, 1743.

Hill, Robert J., 2013. Hedonic price indexes for residential housing: a survey, evaluation and taxonomy. J. Econ. Surv. 27 (5), 879–914. http://dx.doi.org/10.1111/j.1467-6419.2012.00731.x.

Hill, Robert J., Scholz, Michael, 2018. Can geospatial data improve house price indexes? A hedonic imputation approach with splines. Rev. Income Wealth 64 (4), 737–756. http://dx.doi.org/10.1111/roiw.12303.

Himmelberg, Charles, Mayer, Christopher, Sinai, Todd, 2005. Assessing high house prices: Bubbles, fundamentals and misperceptions. J. Econ. Perspect. 19 (4), 67–92. http://dx.doi.org/10.1257/089533005775196769, URL https://www.aeaweb.org/articles?id=10.1257/089533005775196769.

Hoffmann, Florian, Lemieux, Thomas, 2015. Unemployment in the great recession: A comparison of Germany, Canada, and the United States. J. Labor Econ. 34 (S1), S95–S139. http://dx.doi.org/10.1086/682424.

Knight, John R., Sirmans, C.F., Turnbull, Geoffrey K., 1994. List price signaling and buyer behavior in the housing market. J. Real Estate Finance Econ. 9 (3), 177–192. http://dx.doi.org/10.1007/BF01099271.

Kolbe, Jens, Schulz, Rainer, Wersing, Martin, Werwatz, Axel, 2015. Identifying Berlin's land value map using adaptive weights smoothing. Comput. Statist. 30 (3), 767–790. http://dx.doi.org/10.1007/s00180-015-0559-9.

Kosfeld, Reinhold, Werner, Alexander, 2012. Deutsche arbeitsmarktregionen - neuabgrenzung nach den kreisgebietsreformen 2007–2011. Raumforschung Raumordnung 70 (1), 49–64. http://dx.doi.org/10.1007/s13147-011-0137-8.

Lee, Sanghoon, Lin, Jeffrey, 2018. Natural amenities, neighbourhood dynamics, and persistence in the spatial distribution of income. Rev. Econom. Stud. 85 (1), 663–694. http://dx.doi.org/10.1093/restud/rdx018.

LeRoy, Stephen F., Sonstelie, Jon, 1983. Paradise lost and regained: Transportation innovation, income, and residential location. J. Urban Econ. 13 (1), 67–89. http://dx.doi.org/10.1016/0094-1190(83)90046-3.

Linden, Leigh, Rockoff, Jonah E., 2008. Estimates of the impact of crime risk on property values from Megan's laws. Amer. Econ. Rev. 98 (3), 1103–1127. http://dx.doi.org/10.1257/aer.98.3.1103, URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.98.3.1103.

Lopez, Esteban, Hewings, Geoffrey J.D., 2018. Housing price indices for small spatial units. Reg. Sci. Urban Econ. 70, 57–71. http://dx.doi.org/10.1016/j.regsciurbeco.2018.02.005, URL https://www.sciencedirect.com/science/article/pii/S0166046216303246.

Lyons, Ronan C., 2019. Can list prices accurately capture housing price trends? Insights from extreme markets conditions. Finance Res. Lett. 30, 228–232. http://dx.doi.org/10.1016/J.FRL.2018.10.004.

McMillen, Daniel P., 1996. One hundred fifty years of land values in chicago: A nonparametric approach. J. Urban Econ. 40 (1), 100–124. http://dx.doi.org/10.1006/JUEC.1996.0025, URL https://www.sciencedirect.com/science/article/abs/pii/S009411909690025X.

Meese, Richard, Wallace, Nancy, 1991. Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. Real Estate Econ. 19 (3), 308–332. http://dx.doi.org/10.1111/1540-6229.00555.

Mense, Andreas, Michelsen, Claus, Kholodilin, Konstantin A., 2019. Rent Control, Market Segmentation, and Misallocation: Causal Evidence from a Large-Scale Policy Intervention. Working Paper.

Mian, Atif, Sufi, Amir, 2014. What explains the 2007–2009 drop in employment? Econometrica 82 (6), 2197–2223. http://dx.doi.org/10.3982/ECTA10451.

Miller, Norman G., Sklarz, Michael A., 1986. A note on leading indicators of housing market price trends. J. Real Estate Res. 1 (1), 99–109, URL http://www.jstor.org/stable/44095186.

Mills, Edwin S., 1967. An aggregative model of resource allocation in a metropolitan area. Am. Econ. Rev. 57 (2), 197–210, URL http://www.jstor.org/stable/1821621.

Monte, Ferdinando, Redding, Stephen J., Rossi-Hansberg, Esteban, 2018. Commuting, migration, and local employment elasticities. Amer. Econ. Rev. 108 (12), 3855–3890. http://dx.doi.org/10.1257/aer.20151507, URL http://www.aeaweb.org/articles?id=10.1257/aer.20151507.

Mullainathan, Sendhil, Spiess, Jann, 2017. Machine learning: An applied econometric approach. J. Econ. Perspect. 31 (2), 87–106. http://dx.doi.org/10.1257/jep.31.2.87, URL https://www.aeaweb.org/articles?id=10.1257/jep.31.2.87.

Muth, R., 1969. Cities and Housing. Chicago University Press, Chicago.

OECD, 2012. Redefining Urban: A New Way To Measure Metropolitan Areas. Organisation for Economic Cooperation and Development.

Redding, Stephen J., Rossi-Hansberg, Esteban, 2017. Quantitative spatial economics. Annu. Rev. Econ. 9 (1), 21–58. http://dx.doi.org/10.1146/annurev-economics-063016-103713.

Roback, Jennifer, 1982. Wages, rents, and the quality of life. J. Polit. Econ. 90, 1257–1278.

Rosen, Sherwin, 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. J. Polit. Econ. 82 (1), 34–55. http://dx.doi.org/10.1086/260169.

RWI and Immobilienscout24, 2021a. RWI real estate data- apartments for rent - suf. RWI-GEO-RED. version 1. RWI -Leibniz institute for economic research dataset. http://dx.doi.org/10.7807/immo:red:wm:surf:v4.

RWI and Immobilienscout24, 2021b. RWI Real Estate Data- Apartments for sale - suf. RWI-GEO-RED. Version 1. RWI -Leibniz Institute for Economic Research. Dataset. http://dx.doi.org/10.7807/immo:red:wk:surf:v4.

RWI and Immobilienscout24, 2021c. RWI Real Estate Data- houdrd for sale - suf. RWI-GEO-RED. Version 1. RWI -Leibniz Institute for Economic Research. Dataset. http://dx.doi.org/10.7807/immo:red:hk:surf:v4.

RWI and Immobilienscout24, 2021d. RWI Real Estate Data- houses for rent - suf. RWI-GEO-RED. Version 1. RWI -Leibniz Institute for Economic Research. Dataset. http://dx.doi.org/10.7807/immo:red:hm:surf:v4.

Schaffner, Sandra, 2021. FDZ data description: Real-estate data for Germany (RWI-geo-RED) - advertisements on the internet platform ImmobilienScout25. RWI Projektberichte.

Shen, Lily, Ross, Stephen, 2021. Information value of property description: A machine learning approach. J. Urban Econ. 121, 103299. http://dx.doi.org/10.1016/j.jue.2020.103299, URL https://www.sciencedirect.com/science/article/pii/S009411902030070X.

Silver, Mick, 2018. How to measure hedonic property price indexes better. EURONA 2018 (1).

Sims, David P., 2011. Rent control rationing and community composition: Evidence from massachusetts. B.E. J. Econ. Anal. Policy 11 (1), 1–30.

Sunding, David L., Swoboda, Aaron M., 2010. Hedonic analysis with locally weighted regression: An application to the shadow cost of housing regulation in Southern California. Reg. Sci. Urban Econ. 40 (6), 550–573. http://dx.doi.org/10.1016/j.regsciurbeco.2010.07.002.

Tsivanidis, Nick, 2019. Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio. Working Paper, URL https://www.nicktsivanidis.com/s/TsivanidisTransMilenio_62019.pdf.

Wheaton, William C., 1977. Income and urban residence: An analysis of consumer demand for location. Am. Econ. Rev. 67 (4), 620–631, URL http://www.jstor.org/stable/1813394.