CESifo WORKING PAPERS

9187
2021

July 2021

Micro-Geographic Property Price and Rent Indices

Gabriel M. Ahlfeldt, Stephan Heblich, Tobias Seidel

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9187*

Micro-Geographic Property Price and Rent Indices

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 density bias of price-to-rent ratios in levels and trends, within and between cities. Our method lends itself to the creation of comparable neighborhood-level qualified rent indices (*Mietspiegel*) across Germany.

JEL-Codes: R100.

Keywords: index, real estate, price, property, rent.

Gabriel M. Ahlfeldt
London School of Economics and
Political Sciences (LSE) / United Kingdom
g.ahlfeldt@lse.ac.uk

Stephan Heblich
University of Toronto / Canada
stephan.heblich@utoronto.ca

Tobias Seidel
University of Duisburg-Essen / Germany
tobias.seidel@uni-due.de

July 7, 2021
The usual disclaimer applies.

1 Introduction

Reliable indices that capture the market value of real property 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—the current general-purpose workhorse tool in spatial economics—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).[1] 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 the construction of micro-geographic purchase and rent price indices that uses spatial methods to overcome the limitations of sparse property 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 along with geographic coordinates, transaction dates, and property characteristics. The output is a balanced panel data set 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 estimation techniques. Conveniently, it allows the user to manage the *bias-variance trade-off* via the program syntax. Our contribution is to provide a reliable and transparent tool that generates spatial house price indices in an environment that is typically dominated by commercial data providers to whom their algorithms are the "secret souce". Upon final publication of this paper, we will 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 more than ten years.

The house price and rent index we propose combines several techniques that are

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

established in urban economics and data science. We start with the popular hedonic regression approach whose micro-foundations were 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., 2020a). We nest this approach that has become canonical in urban economics research into locally weighted regressions. This approach was originally suggested by Cleveland and Devlin (1988) and first adopted to studies of property 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*.

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. 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, whose diameter also 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, such as 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 whose values can be chosen in the programming syntax. We propose to proceed with conservative parameter values since we wish to avoid implausible outliers. Other users may choose different values —resulting, for example, in smaller spatial windows— that best suit their aversion to outliers. Users who are willing to formalize their objective function that trades off bias against variance may also delegate the identification of the critical parameter values to another algorithm. In this case, our approach becomes 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 for Germany to introduce its functionality. Our application makes use of geocoded data from the online platform *Immoscout24* for the period 2007–2018 that is largely representative of the rising pool of property price information on prices and rents that is accessible to researchers and data scientist 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 in 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 diameter of $500m$. To validate the accuracy of the spatial house price index, we exclude information from about three quarters of all hexagons and recalculate the index for all locations. A comparison between the actual and predicted values shows a tight fit that underlines the validity of our procedure.

The application to the case of Germany comes with the benefit that existing house price and rent indices are not available below the county level and often consist of average prices, possibly by house type. 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 that are adjusted for property characteristics and location. Since we develop the index from micro-data, we can also choose 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., 2020a, 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 account. We directly relate to an emerging literature that analyzes the determinants of price-to-rent ratios, albeit at a much lower level of spatial detail. Our micro-geographic rent and purchase price indices reveal new stylized facts that call for further analyses: There is a density bias in the price-to-rent ratio in levels and trends. Price-to-rent ratios tend to be higher in large cities. Within cities, they tend to be high in the more central parts. This density bias increased since 2010 when prices have started to outpace rents earlier and much more strongly in the largest agglomerations as well as in central-city neighborhoods.

The data set we share will allow researchers to delve into the origins of the spatially biased divergence that may relate to supply conditions (Glaeser et al., 2008; Hilber and Mense, 2021), credit constraints (Himmelberg et al., 2005), or foreign direct investment (Badarinza and Ramadorai, 2018), just to name a few. Hence, our contribution motivates and facilitates an entire research agenda.

More generally, our work connects to various important research strands that are concerned with either generating or using 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. Recent notable developments in 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), or machine learning to 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 in a way that supports 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 theory-consistent methodology for the generation of micro-geographic price and rent indices that can be viewed as canonical among urban economists.

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 cost of agglomeration (see Ahlfeldt and Pietrostefani, 2019, for a review), infer quality of life (Roback, 1982; Ahlfeldt et al., 2020a), 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 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 $j \in J$. For each j, we run a locally weighted regression (LWR) of the following type:

$$\begin{aligned}\ln \mathcal{P}_{i,t} = \; & a_t^j + \bar{S}_i b^j + \sum_z d_z^j (D_i^j \times I(z = t)) + e^j I(D_i^j > T^j)_i \\ & + f^j (X_i - X^j) + g^j (Y_i - Y^j) + \epsilon_{i,t}^j,\end{aligned}$$

where $\mathcal{P}_{i,t}$ is the purchase or rental price of a property i transacted in year t. $\bar{S}_i$ 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_z^j being the LWR j-specific gradient in year z. $I(.)$ 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 spatial gradients. $\epsilon_{i,t}^j$ is the residual term.

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

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

where $N^{(D_i^j \leq T^{s \in \{1,2,3,4\}})}$ 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:

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

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

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 price index for a target unit is then simply defined as:

$$
\hat{\mathcal{P}}_t^j = \exp(\hat{\alpha}_t^j),
$$

which we recover from the LWR-j-specific estimates of time-fixed effects α_t^j. To facilitate the computation of confidence bands, we also report standard errors

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

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

Via the parameters $\{A^1, A^2, A^3, A^4, N^A, T^1, T^2, T^3, T^4, N^T\}$, the user has flexible control over the *bias-variance trade-off*. Smaller values in all parameters will generally lead to greater spatial variation, 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 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.

3 Application

The procedure outlined in the previous section is entirely point-pattern based and does not rely on context-specific spatial units or administrative data. This makes it applicable in a wide range of geographic contexts. The programming syntax further allows users to freely choose parameters for the predictive model, thus making it easily adjustable to different applications and spatial resolutions. To illustrate the functionality of our prediction method, we apply the algorithm to five of the arguably most popular spatial layers in Germany that also corroborate the comprehensive labor market data provided by the Research Data Centre of the Federal Employment Agency. Specifically, we calculate price and rent indices at the level of (i) *local labor markets*, (ii) *counties*, (iii) *municipalities*, (iv) *postcodes*, and (v) *micro grids*. Data generated at these aggregation levels serve different purposes in the literature. Some units vary greatly in geographic size while others vary greatly in population which affects the average and the variability of the density of observations. In each case, we suggest suitable parameter values along with a a rationale for the specific choice and present a series of maps to illustrate the spatio-temporal variation generated by our algorithm. In a last step, we validate our method with out-of-sample predictions.

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 20 million properties listed for rent and an equal amount listed for purchase over the period 2007–2018. 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 further characteristics, e.g. information on the type of heating system. We use the following readily accessible scientific use files which is georeferenced at the level of 1km^2 grid cells in projected units of the ETRS coordinate system (address-based georeferences are accessible on site at RWI): RWI and Immobilienscount 24 (2021a,b,c,d). We refer to Schaffner (2021) for a detailed data description. 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 30m^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., 2020a; 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 sizable 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 centre of the city, where —assuming a monocentric city structure— commuting cost are zero.

3.2.2 Parameter choices

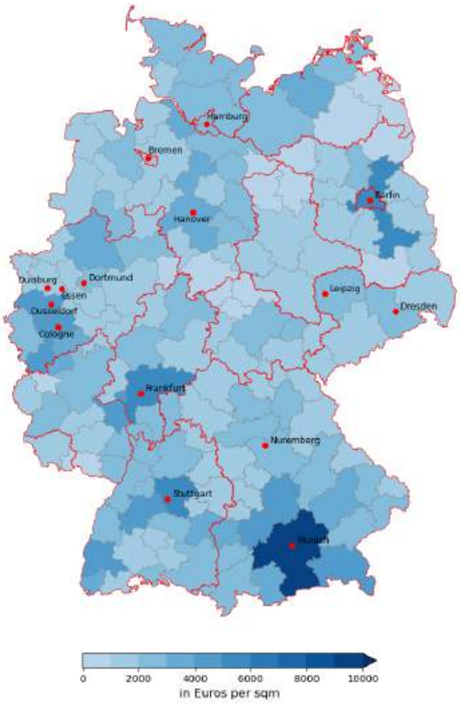
We follow Combes et al. (2019) and argue that a theory-consistent index that captures pure housing cost in a LLM should control for a parametric distance gradient that captures commuting costs in the spirit of the monocentric city model (see Alonso, 1964; Mills, 1967; Muth, 1969). For the spatial window 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 centre 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

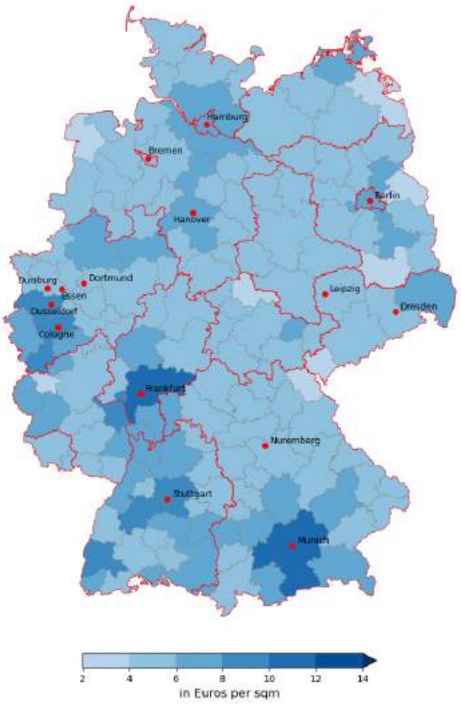
We present our results for the years 2007 and 2018 in Figure 1. Panels (a) and (b) depict prices for purchases while panels (c) and (d) are based on rental prices. The indices clearly reveal an increase in both the levels and the spatial dispersion of prices. The LLM *München* was leading the list in terms of purchase prices with 3,585€ (2007) and 9,393€ (2018) while *Elbe-Elster* (590€, 2007) and *Osnabrück* (642€, 2018) had the lowest prices per square meter. *Berlin* developed most dynamically with a growth rate over the period of 172.4% while prices declined by 6.9%

Figure 1: Local labor markets

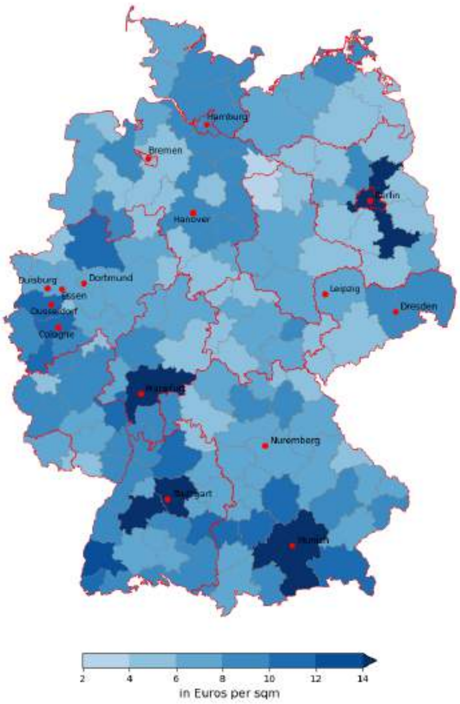
(a) Purchases, 2007



(b) Purchases, 2018



(c) Rents, 2007



(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) is 141 local labour markets as defined by Kosfeld and Werner (2012).

in *Würzburg*. Describing regional disparities in house prices based on the coefficient of variation, our index implies an increase in inequality by 64.5% between 2007-2018.

Turning to rental prices, *München* was the most expensive local labor market in both 2007 (11.89€) and 2018 (21.37€). *Südvorpommern* (3.25€, 2008) and *Lüchow-Dannenberg* (3.25€, 2018) were at the lower end of the ranking. Rents have grown by 123.3% in *Berlin* while they declined by 4.4% in *Lüchow-Dannenberg*. Rental price dispersion has increased by 38.2% over this period.

3.3 Application II: Counties

3.3.1 Context

In the German context, counties (NUTS3 regions in official EU nomenclature) define the least granular spatial unit where a variety of data are publicly available. Examples include the German *Regionaldatenbank* published by the Federal Statistical Office or commuting flows published by the Institute for Employment Research (IAB). Quantitative research conducted at this spatial level depends on publicly available house price indices that are often subject to the same criticism expressed above, namely that they are unweighted averages (see Seidel and Wickerath, 2020; Braun and Lee, 2021, in the German context). As with the LLM areas, our index provides a theory-consistent measure of housing cost by predicting prices at the economic center of a county (Combes et al., 2019). We use the jurisdictional classification in 2019 which comprises 401 counties. At this geographical level, additional information can be easily matched.

3.3.2 Parameter choices

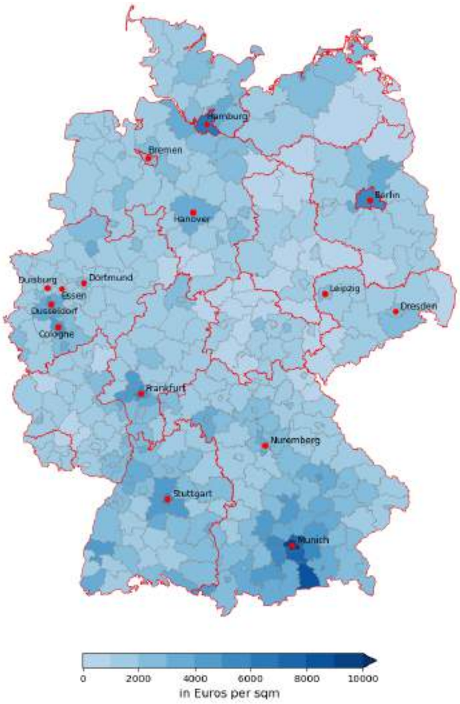
County-level data are often employed as an approximation for cities in the absence of more suitable data. To account for this, we recommend the same parametrization we employed for LLMs and employ it in our calculations. However, in some instances researchers may be genuinely interested in county-level variables without a particular urban model in mind. In these cases, we recommend estimating the municipality-level index (sub-units of counties) using the parametrization introduced in the next section and aggregating it to the county level, weighted by population.

3.3.3 Results

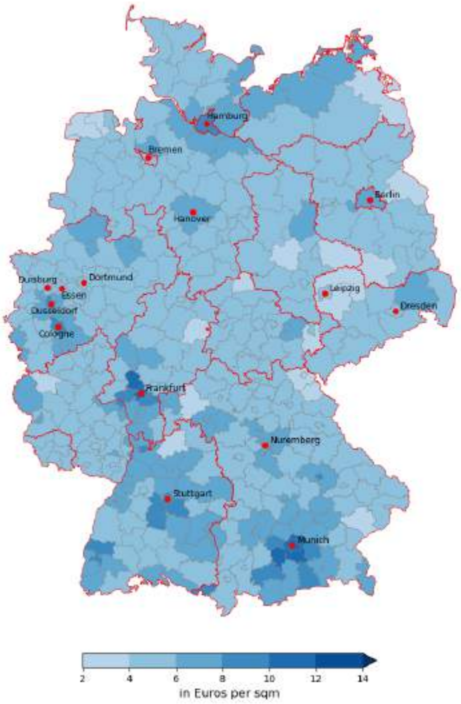
The general pattern of relative house prices relate to that observed in the previous part on local labor markets - albeit at higher resolution (see Figure 2). The most

Figure 2: Counties

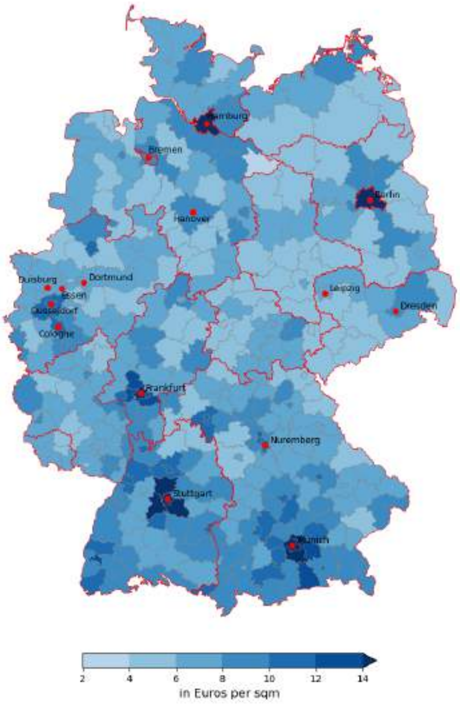
(a) Purchases, 2007



(b) Purchases, 2018



(c) Rents, 2007



(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) is 401 counties based on the jurisdictional definition in 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

expensive county with respect to purchase prices was, again, *München* (city) both in 2007 (3,585€) and 2018 (9,393€). *Vogtlandkreis* in Saxony was the least expensive county with square meter prices of 518€ (2007) and 606€ (2018), respectively. In terms of changes, purchase prices declined by 14.5% in *Gotha* (Thuringia) while they increased by 172.4% in *Berlin*. Taking the coefficient of variation as a measure of price dispersion, we find that overall inequality increased by 60.8%.

On the rental market, the city of *München* was leading the list in both years at 11.89€ (2007) and 21.37€ (2018). *Vorpommern-Greifswald* ranked at the lower end in 2007 with 3.24€ per square meter. In 2018, *Lüchow-Dannenberg* took that place with a rental price of 4.07€. The latter county was also characterized by the lowest growth rate in rents, namely a decline of 4.4%. *Berlin* was located at the top of the ranking also for the country classification with a growth rate of 123.3%. Rental price dispersion increased by 30.1%.

3.4 Application III: Municipalities

3.4.1 Context

There are about 11,000 municipalities (local administrative units, LAU, in EU nomenclature) in Germany that differ quite remarkably in their size, both across states and within states. At the extreme, the city state of *Berlin*, home to about 3.6 million inhabitants, and *Gröde* or *Dierfeld*, both home to 10 inhabitants each, are considered one municipality. Therefore, some states with extremely small municipalities such as Rhineland Palatinate grouped municipalities in municipal associations (*Verbandsgemeinden*) that share a common local administration. Because of the enhanced comparability across states, it is sensible to employ municipal associations (where they have been formed) in quantitative research (Ahlfeldt et al., 2020b). We follow this convention and, using the official classification for 2019, construct our house price index for 4,608 municipalities and municipal associations.

3.4.2 Parameter choices

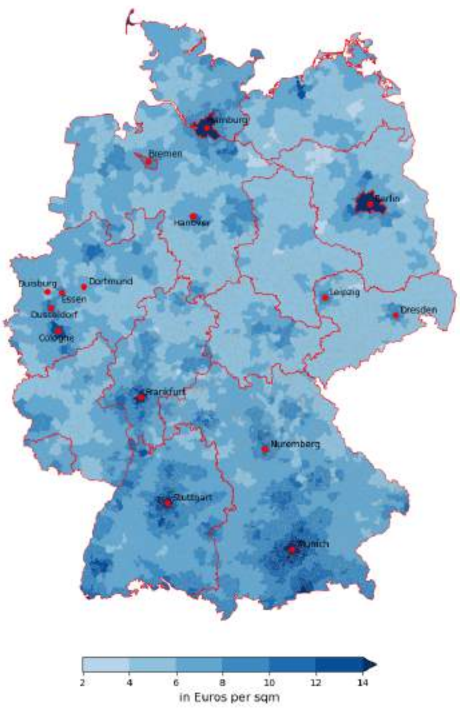
Municipalities that do not coincide with independent cities (like the extreme case Berlin) are significantly smaller than LLMs or counties. Consequently, the focus moves away from a theory-consistent index that adjusts for commuting costs and towards a purely empirical problem of predicting an index for a relatively small area within which there will typically not be enough observations to estimate a credible conditional mean. To increase the number ob observations, we consider distance buffers around the municipality of interest and add additional observations within

Figure 3: Joint municipalities (Verbandsgemeinden)

(a) Purchases, 2007

(b) Purchases, 2018

(c) Rents, 2007



(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 4,608 joint municipalities. These entities are grouped according to joint administration at the local level. The jurisdictional definition refers to 31 December 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

this buffer until we reach a minimum number of observations that guarantees a reliable estimate. Specifically, we make the following choices for the distance and size thresholds $\{A^1 = 10, A^2 = 25, A^3 = 50, A^4 = 100, T^1 = 10, T^2 = 15, T^3 = 20, T^4 = 50\}$ (all in km) and minimum transaction numbers $\{N^A = 10,000, N^T = 1,000\}$. These choices result in 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. The parametric distance control and the spatial fixed effect then ensure that we estimate an index that is specific to the municipality even if we have to use a relatively large window.

3.4.3 Results

Figure 3 illustrates nicely the evolution of house price changes, both for purchases and rents, at a high resolution. Eyeballing suggests that the largest cities have experienced the highest growth rates. Indeed, we find *Grünwalder Forst* near Munich (4,402€, 2007) and *München* (10,701€, 2018) at the top of the purchase price index. *Dahlen* (Saxony) and *Huy* (Saxony-Anhalt) had the lowest purchase prices per square meter at 496€ (2007) and 562€ (2018), respectively. *München* experienced growth rates of 185.1% while prices declined by 27.9% in *Peenetal/ Loitz* (Mecklenburg-Western Pomerania). In terms of the coefficient of variation, we find an increase in price dispersion of 50.9%.

Turning to the rental market, the least expensive municipalities were *Neunburg vorm Wald* (Bavaria) with a square meter price of 2.80€ (2007) and *Heiligengrabe* (Brandenburg) with 2.89€ (2018). *München* was leading the list in both years with respective rents of 12.34€ and 23.08€. *Berlin* experienced the highest rent growth of 141.3% while rents declined by 38.2% in *Treptower Tollensewinkel* (Mecklenburg-Western Pomerania). The coefficient of variation increased by 20.9%.

3.5 Application IV: Postcodes

3.5.1 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.[2] There are 8,255 postcodes that are designed to accommodate similar

[2]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

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 (note that the next application will address the problem of heterogeneous geographic size). 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.

3.5.2 Parameter choices

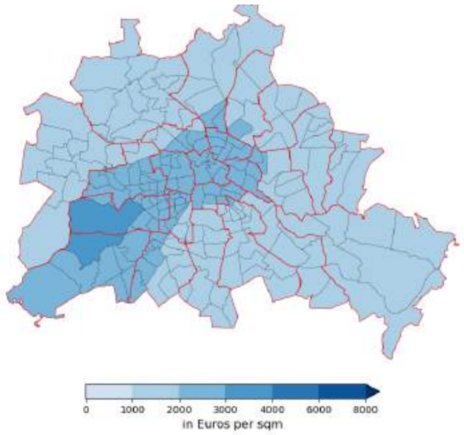
Similarly to municipalities, we mainly face an empirical problem of predicting an index for a relatively small area within which there will typically not be enough observations to estimate a credible conditional mean. As before, we overcome this limitation by using observations from neighboring municipalities. 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 $\{N^A = 10,000, N^T = 1,000\}$ 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.

3.5.3 Results

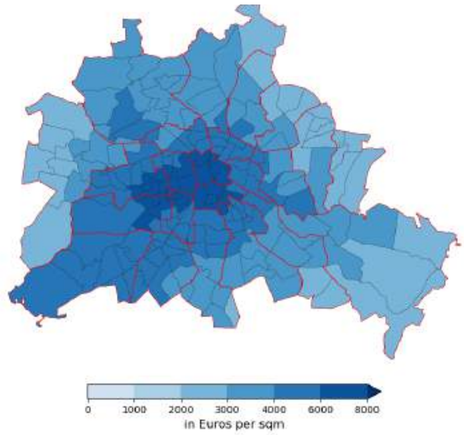
As the results at the postcode level look very similar to the index at the municipality level, we take advantage of the higher resolution and focus on Berlin that consists of 190 postcode areas (we provide maps for other German cities in the Online Appendix). Figure 4 shows four panels according to the previous structure. We observe that the center and the south-west tend to be the high-price areas and the development over time clearly reveals the attractiveness of the city center. Purchase prices have increased between 57-264% translating into an increase in inequality of 22.3%. On the rental market prices were raised between 17-141%. This, however, led to a more pronounced change in the inequality measure (coefficient of variation) by 49%.

patchy after 1971–the next waves are 1987 and then 2011–and census tracts are not consistently assigned.

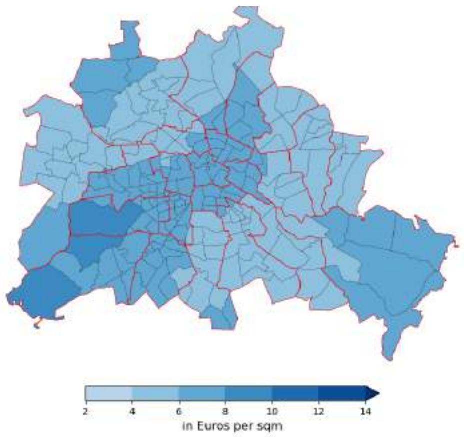
Figure 4: Postcodes Berlin



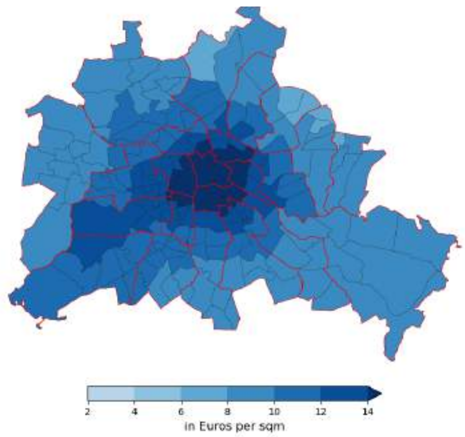
(a) Purchases, 2007



(b) Purchases, 2018



(c) Rents, 2007



(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 190 postcode areas in Berlin. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

3.6 Application V: Micro grids

3.6.1 Context

While postcode-level precision helps us zoom into cities, postcodes in Germany still tend to be larger than US census tracts or output areas in the UK. They are certainly much larger than the housing blocks that have been used to analyze the strengths and spatial scope of social and professional interactions (Ahlfeldt et al., 2015). To achieve even higher precision, we introduce a last application where we show how our algorithm can be applied to zoom into even finer grids of arbitrary shape. To this end, we construct a grid of hexagons with a diameter of 500 meters that covers, again, the entire Berlin city state. While we could generate an index at this level for the entire country, the returns to enhancing the spatial resolution would be confined to dense urban areas where spatial differences are particularly pronounced over short distances and the density of observations is sufficiently high.

3.6.2 Parameter choices

Applying the algorithm only to a dense city like Berlin, we can make parameter choices that aim at maximizing the flexibility of the index subject to the constraint that there remain sufficient degrees of freedom. Hence, we use small distance thresholds with the intention of only reverting to larger spatial windows and spatial fixed effects if observations are insufficient. We make the following choices for the thresholds: $\{A^1 = 5, A^2 = 10, A^3 = 25, A^4 = 50, T^1 = 1, T^2 = 2, T^3 = 5, T^4 = 10\}$ (all in km) for distance and $\{N^A = 10,000, N^T = 1,000\}$ for the number of transactions.

3.6.3 Results

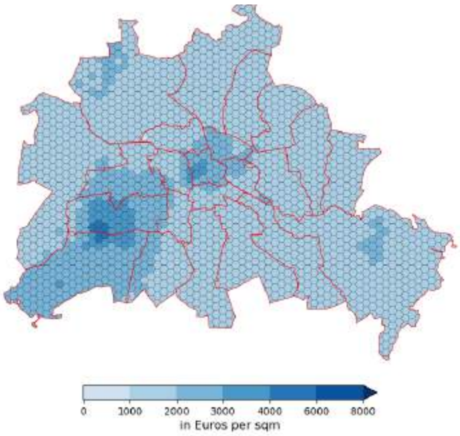
Naturally, the generic pattern of rents and purchase prices at the level of hexagons in Figure 5 resembles the postcode-level maps in Figure 4. Central areas rapidly appreciated, even relative to the most attractive wealthy suburbs in the south-west. This is a manifestation of the gentrification trends observed in cities around the world. However, more features of the spatial structure become apparent at the finer hexagon level. Purchase price maps reveal the duo-centric structure of the city, with prices peaking near the prestigious Boulevards *Kurfürstendamm* in former West Berlin and *Unter den Linden* in former East Berlin. Turning to rental price maps, we observe pockets of high rental prices outside the central district *Mitte* such as in *Kreuzberg* and the bordering districts *Neukölln* and *Friedrichshain*, a vibrant area that has become a hub of startup entrepreneurship (Moeller, 2018).

3.7 Validation

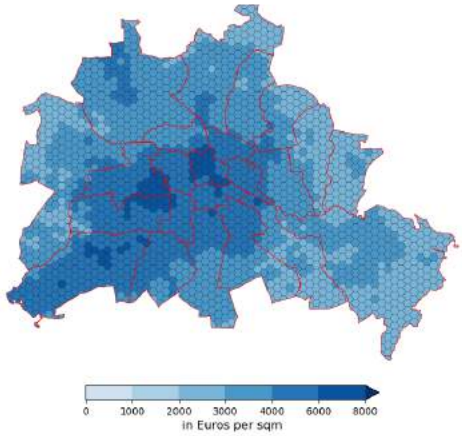
In this section, we subject our micro-geographic indices to a fairly demanding out-of-sample prediction exercise where we use data from a fraction of the hexagons introduced in Section 3.6 to predict our index for the remaining ones.

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

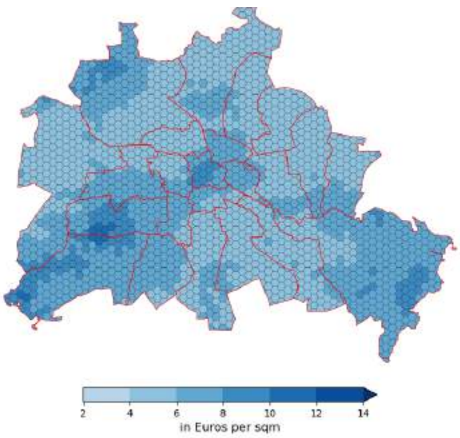
Figure 5: Hexagons Berlin



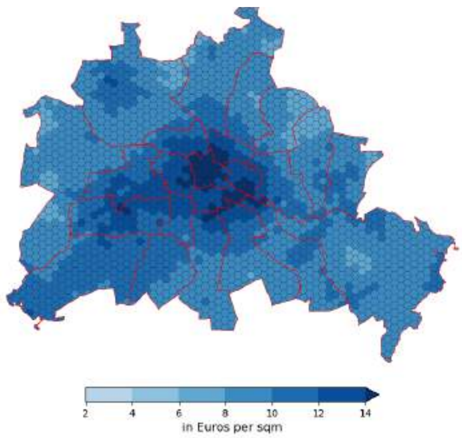
(a) Purchases, 2007



(b) Purchases, 2018



(c) Rents, 2007



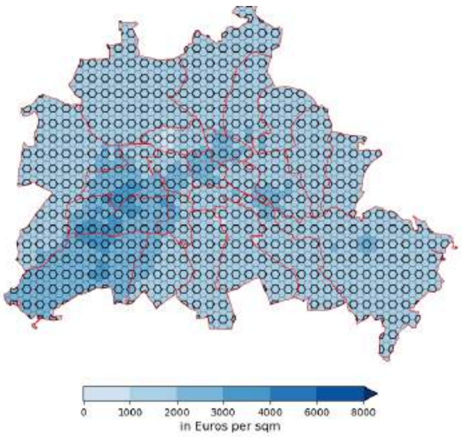
(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 1,953 hexagons with diameter of 500m in Berlin. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

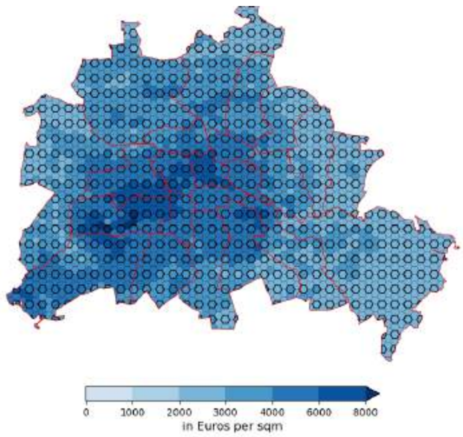
it is useful to test its out-of-sample predictive power. To this end, we focus on the 500-meter grid for Berlin and drop about three quarters of the hexagons. Specifically, we design the sampling such that we leave at least one queens contiguity buffer between any estimation hexagon and the nearest overidentification hexagon. Figure 6 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). Figure 6 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 3.6. 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 overidentifcation sample in Figure 7. It is reassuring to see that the algorithm does a good job predicting values in areas with sparse data.

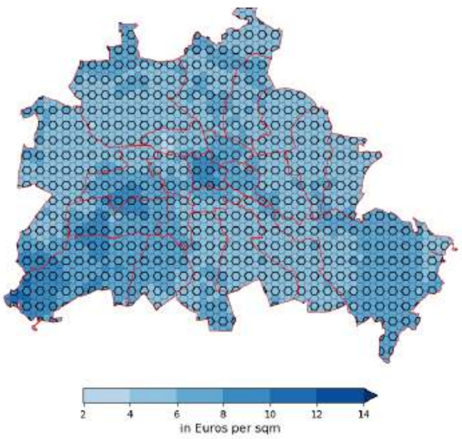
Figure 6: Hexagons Berlin: Validation



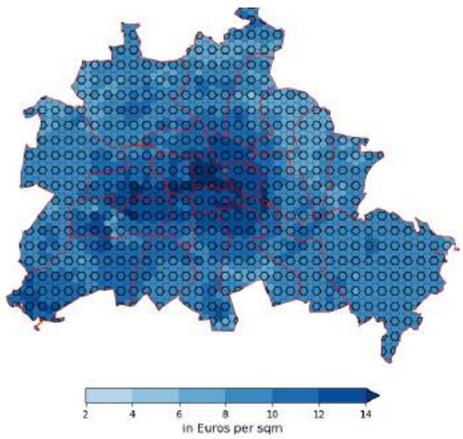
(a) Purchases, 2007



(b) Purchases, 2018



(c) Rents, 2007



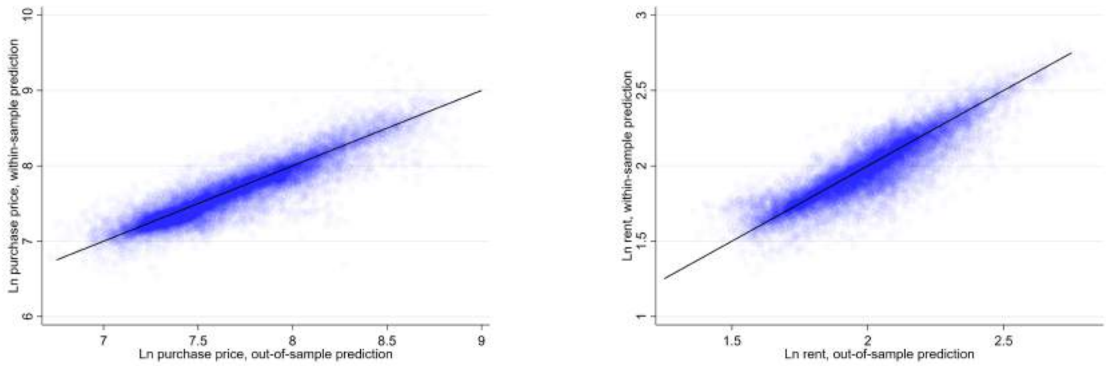
(d) Rents, 2018

Note: The estimation is based on the black hexagons (25% of 1,953 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).

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.

Figure 7: Validation Exercise–Overidentification



(a) Purchases (b) Rents

Note: Unit of observation in panels (a)-(d) 1,953 hexagons with diameter of 500m 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 shows the correlation between out-of-sample predictions for purchase prices (a) or rents (b) and actual prices. The line is the 45 degree line.

4.1 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).

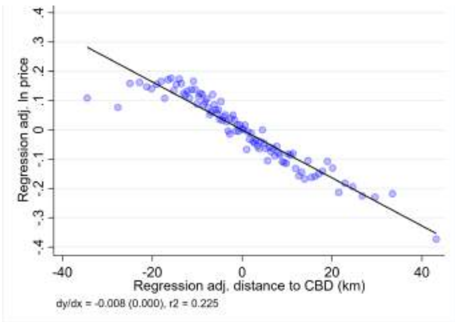
With this in mind, we correlate our rent and price indices with several locational characteristics in Figure 8. We provide bin scatter plots based on percentiles for a clearer presentation, but the underlying data comprise 8,255 postcodes and the entire country, which is a fairly broad coverage within 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. (2020a). 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. In these validation exercises, we show partial correlations of prices or rents and these amenity measures that are regression-adjusted. Specifically, we condition each measure on all the other amenity measures and further absorb local LLM effects.

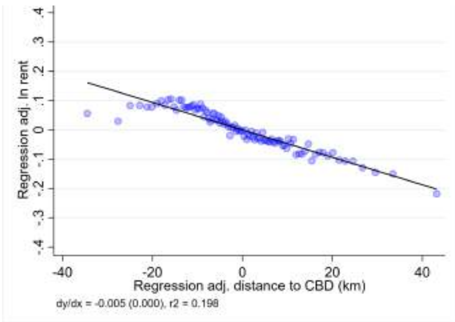
Although the cross-sectional multivariate regression is a workhorse tool in the hedonic price literature, we caution against causal interpretations of the partial correlations since there may be omitted variables correlated with the covariates we consider. In fact, distance from the CBD, by its very nature, is supposed to capture a multitude of factors that make traveling to city centres 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 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

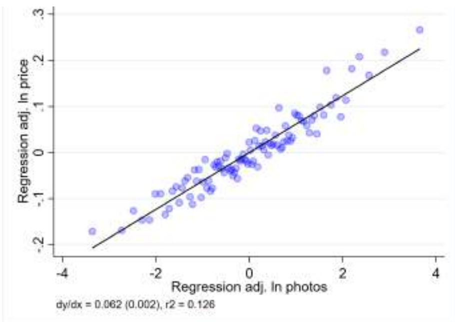
Figure 8: Stylized facts I–Cross-sectional correlations



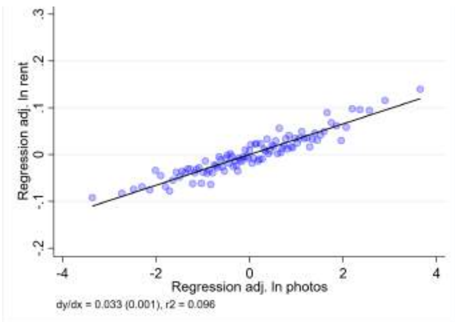
(a) Purchases–Distance



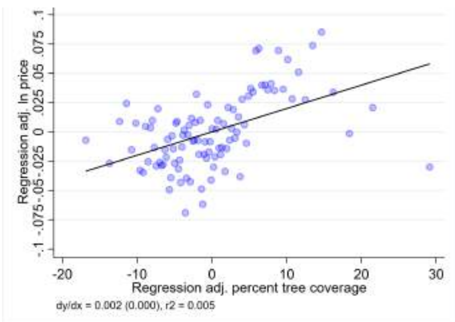
(b) Rents–Distance



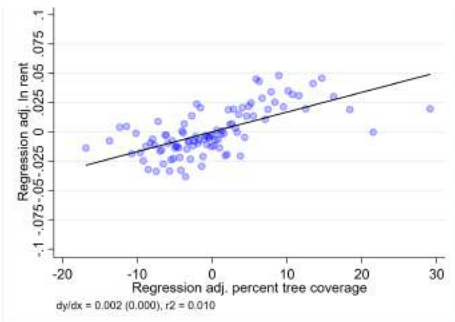
(c) Purchases–Photos



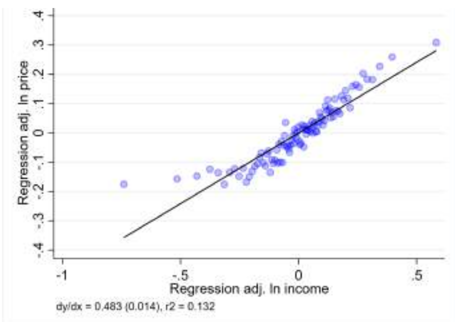
(d) Rents–Photos



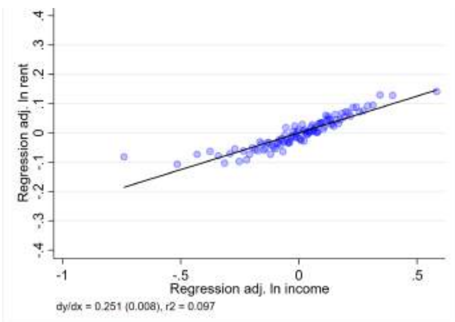
(e) Purchases–Tree Coverage



(f) Rents–Tree Coverage



(g) Purchases–Income



(h) Rents–Income

Note: Unit of observation in panels (a)-(h) are 8,255 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).

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 250m 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 Figure 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 that the yield is systematically related to the amenity value of locations within cities. 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 form of a lower yield in the expectation of greater (imputed) rents in the future. 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) using long-run income data.

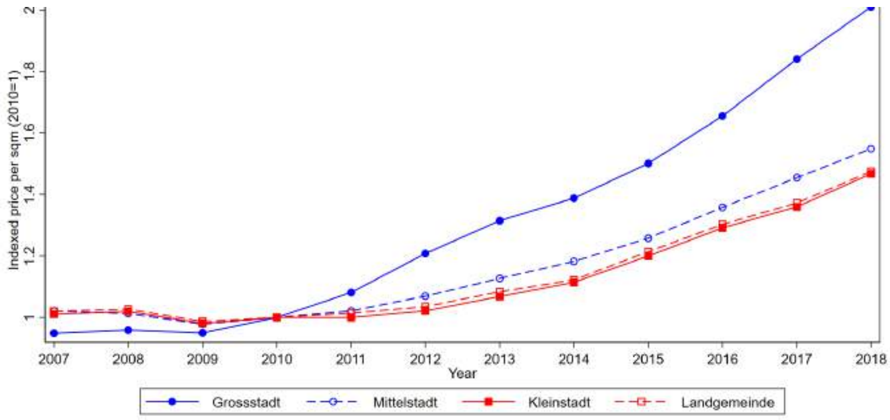
Another interesting variable to correlate property prices with is neighborhood income. Because of residential sorting, income is likely determined by the same variables as purchase prices and rents, including those that we cannot observe. In principle, the correlation can go both ways since preference-based 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 centre 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 centre 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 downtowns 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 Figure 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 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.

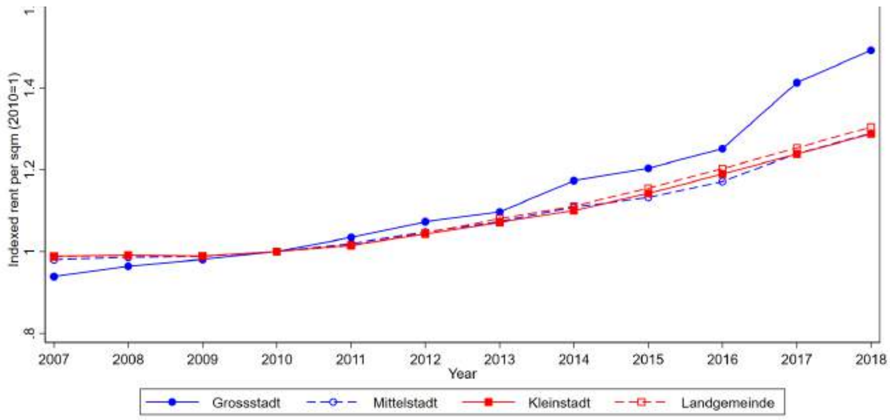
4.2 Temporal trends

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 2007–2018 for four different types of cities, large cities (*Grossstadt*), small cities (*Kleinstadt*) and rural areas (*Landgemeinden*) in panels (a) and (b) of Figure 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,

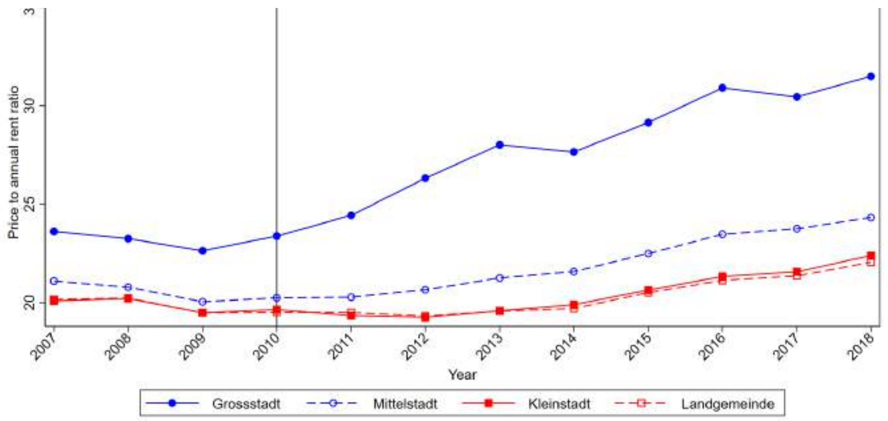
Figure 9: Purchase price and rent trends between cities



(a) Purchases



(b) Rents



(c) Price-to-rent ratio

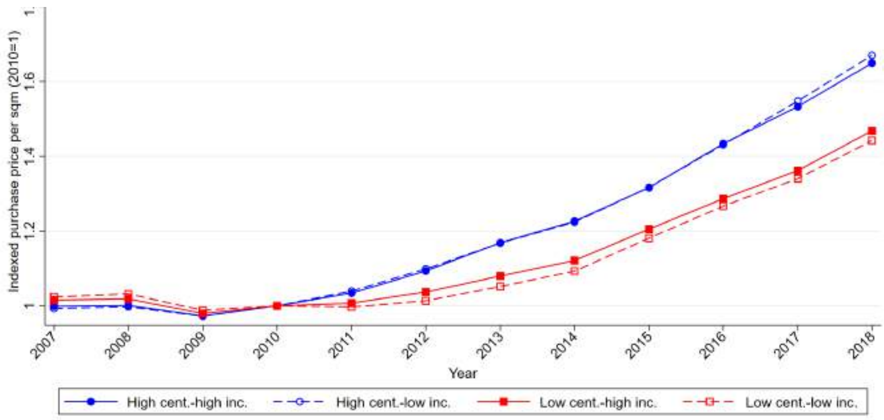
Note: Figures based on data for 4,608 municipal associations. The figures shows the development of prices (a) rents (b) and the price-to-rent ratio (c) over the years 2007–2018 for four different types of cities, large cities (Grossstadt), medium sized cities (Mittelstadt), small cities (Kleinstadt) and rural areas (Landgemeinden).

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

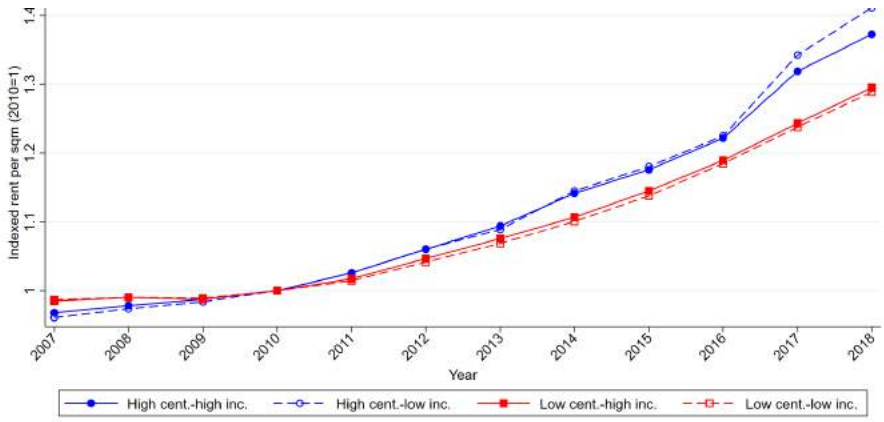
Price-to-rent ratios averaged at a near 25 in 2010, a much higher level than in the less agglomerated parts of the country. This is consistent with the stylized evidence from within cities introduced in Section 4.1 which points to price-to-rent ratios that are generally higher in more expensive areas, be it because buyers expect greater returns in the future, or lower risk. Figure 9 adds that the spatial bias in the price-to-rent ratio 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. 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 "save 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.

Bridging the gap between Figure 8, which considers variation in prices in a cross-section within cities, and Figure 9, which considers variation in prices between cities over time, we look into price-to-rent ratios over time within cities in Figure 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. Figure 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 Figure 9. The main insight from

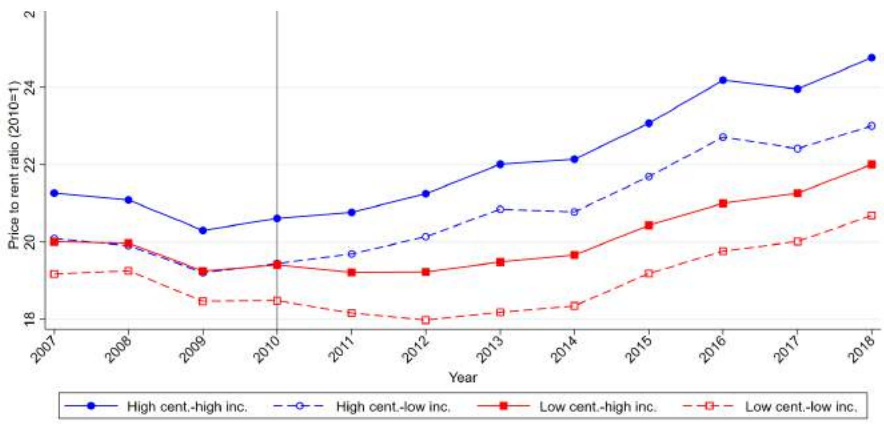
Figure 10: Purchase price and rent trends within cities



(a) Purchases



(b) Rents



(c) Price-to-rent ratio

Note: Figures based on data for 8,255 postcodes. 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.

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 lower initial yields because they expected greater future returns in central parts of German cities. Indeed, the relative pattern of rental growth in Figure 10, panel (b) (central vs. non-central) is strikingly similar to the relative pattern of rental growth in Figure 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-story buildings), and are likely better known to non-local investors. 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.

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.

Upon publication of this paper, we will publish the underlying prediction algorithm along with suggestions for the critical parameter choices, so others who have access to individual property transaction data can easily employ our method to create their own indices under their own parameter choices according to their own needs. A collateral of our exemplary application are spatial price indices that are unique in terms of the micro-geographic coverage of the German buyer and renter market since 2007. We hope that the algorithm and indices published 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 new stylized facts on the density-bias in German price-to-rent ratios in levels and trends 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.

References

Ahlfeldt, Gabriel, "If alonso was right: Modeling accessibility and explaining the residential land gradient," *Journal of Regional Science*, 2011, *51* (2), 318–338.

Ahlfeldt, Gabriel M. and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111* (February), 93–107.

_ , Fabian Bald, Duncan Roth, and Tobias Seidel, "Quality of life in a dynamic spatial model," *CEPR Discussion Paper*, 2020, *15594.*

_ , Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density : Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (4).

Ahlfeldt, Gariel M., Duncan Roth, and Tobias Seidel, "Employment-maximizing minimum wages," *Working paper*, 2020.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 5 2014, *129* (3), 1085–1140.

Almagro, Milena and Tomas Domínguez-Iino, "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Working paper*, 2020.

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

Athey, Susan and Guido W Imbens, "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 2019, *11* (1), 685–725.

Autor, David H, Christopher J Palmer, and Parag A Pathak, "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts," *Journal of Political Economy*, 2014, *122* (3), 661–717.

Badarinza, Cristian and Tarun Ramadorai, "Home away from home? Foreign demand and London house prices," *Journal of Financial Economics*, 12 2018, *130* (3), 532–555.

Bailey, Martin J, Richard F Muth, and Hugh O Nourse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, 12 1963, *58* (304), 933–942.

Braun, Stefanie and Gabriel S Lee, "The prices of residential land in German counties," *Regional Science and Urban Economics*, 2021, *89*, 103676.

Brueckner, Jan K, Jacques-FranÃ§ois Thisse, and Yves Zenou, "Why is central Paris rich and downtown Detroit poor?: An amenity-based theory," *European Economic Review*, 1 1999, *43* (1), 91–107.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein, "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *Quarterly Journal of Economics*, 2010, *125* (1), 215–261.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J Notowidigdo, "Housing Booms, Manufacturing Decline and Labour Market Outcomes," *The Economic Journal*, 2 2018, *n/a* (n/a).

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Cheshire, Paul and Stephen Sheppard, "On the Price of Land and the Value of Amenities," *Economica*, 1995, *62* (246), 247–267.

Clark, Colin, "Urban Population Densities," *Journal of the Royal Statistical Society. Series A (General)*, 1951, *114* (4), 490–496.

Clayton, Jim, "Rational Expectations, Market Fundamentals and Housing Price Volatility," *Real Estate Economics*, 12 1996, *24* (4), 441–470.

Cleveland, William S and Susan J Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, 1988, *83* (403), 596–610.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "The Costs of Agglomeration: House and Land Prices in French Cities," *The Review of Economic Studies*, 10 2019, *86* (4), 1556–1589.

Diamond, Rebecca, Tim McQuade, and Franklin Qian, "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, 9 2019, *109* (9), 3365–3394.

DiMiceli, Charlene M, Mark L Carroll, Robert A Sohlberg, Chengquan Huang, Matthew C Hansen, and John R G Townshend, "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*, 2017.

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

Glaeser, Edward L, Joseph Gyourko, and Albert Saiz, "Housing supply and housing bubbles," *Journal of Urban Economics*, 2008, *64* (2), 198–217.

Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm, "The Making of the Modern Metropolis: Evidence from London," *The Quarterly Journal of Economics*, 11 2020, *135* (4), 2059–2133.

Henkel, Marcel, Tobias Seidel, and Jens Suedekum, "Fiscal Transfers in the Spatial Economy," *American Economic Journal: Economic Policy*, 2021, *forthcomin.*

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

Hill, Robert J, "Hedonic price indexes for residential housing: a survey, evaluation and taxonomy," *Journal of Economic Surveys*, 12 2013, *27* (5), 879–914.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai, "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions," *Journal of Economic Perspectives*, 2005, *19* (4), 67–92.

Hoffmann, Florian and Thomas Lemieux, "Unemployment in the Great Recession: A Comparison of Germany, Canada, and the United States," *Journal of Labor Economics*, 12 2015, *34* (S1), S95–S139.

Kolbe, Jens, Rainer Schulz, Martin Wersing, and Axel Werwatz, "Identifying Berlin's land value map using adaptive weights smoothing," *Computational Statistics*, 2015, *30* (3), 767–790.

Kosfeld, Reinhold and Alexander Werner, "Deutsche Arbeitsmarktregionen -Neuabgrenzung nach den Kreisgebietsreformen 2007-2011," *Raumforschung und Raumordnung*, 2012, *70* (1), 49–64.

Lee, Sanghoon and Jeffrey Lin, "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income," *The Review of Economic Studies*, 1 2018, *85* (1), 663–694.

LeRoy, Stephen F. and Jon Sonstelie, "Paradise lost and regained: Transportation innovation, income, and residential location," *Journal of Urban Economics*, 1 1983, *13* (1), 67–89.

Linden, Leigh and Jonah E. Rockoff, "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 2008, *98* (3), 1103–1127.

Lopez, Esteban and Geoffrey J D Hewings, "Housing price indices for small spatial units," *Regional Science and Urban Economics*, 2018, *70*, 57–71.

McMillen, Daniel P., "One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach," *Journal of Urban Economics*, 7 1996, *40* (1), 100–124.

Meese, Richard and Nancy Wallace, "Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices," *Real Estate Economics*, 9 1991, *19* (3), 308–332.

Mian, Atif and Amir Sufi, "What Explains the 2007â2009 Drop in Employment?," *Econometrica*, 11 2014, *82* (6), 2197–2223.

Mills, Edwin S, "An Aggregative Model of Resource Allocation in a Metropolitan Area," *The American Economic Review*, 1967, *57* (2), 197–210.

Moeller, Kristoffer, "Culturally clustered or in the cloud? How amenities drive firm location decision in Berlin," *Journal of Regional Science*, 9 2018, *58* (4), 728–758.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 5 2017, *31* (2), 87–106.

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

Redding, Stephen J. and Esteban Rossi-Hansberg, "Quantitative Spatial Economics," *Annual Review of Economics*, 2017, *9* (1), 21–58.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90*, 1257–1278.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 1 1974, *82* (1), 34–55.

RWI and Immobilienscount 24, "RWI Real Estate Data- Apartments for Rent - suf. RWI-GEO-RED. Version 1. RWI -Leibniz Institute for Economic Research. Dataset. https://doi.org/10.7807/immo:red:wm:surf:v4," 2021.

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

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

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

Schaffner, Sandra, "FDZ Data description: Real-Estate Data for Germany (RWI-GEO-RED) - Advertisements on the Internet Platform ImmobilienScout25," *RWI Projectberichte, Essen.*, 2021.

Seidel, Tobias and Jan Wickerath, "Rush hours and urbanization," *Regional Science and Urban Economics*, 11 2020, *85*, 103580.

Shen, Lily and Stephen Ross, "Information value of property description: A Machine learning approach," *Journal of Urban Economics*, 2021, *121*, 103299.

Sims, David P, "Rent Control Rationing and Community Composition: Evidence from Massachusetts," *The B.E. Journal of Economic Analysis & Policy*, 2011, *11* (1), 1–30.

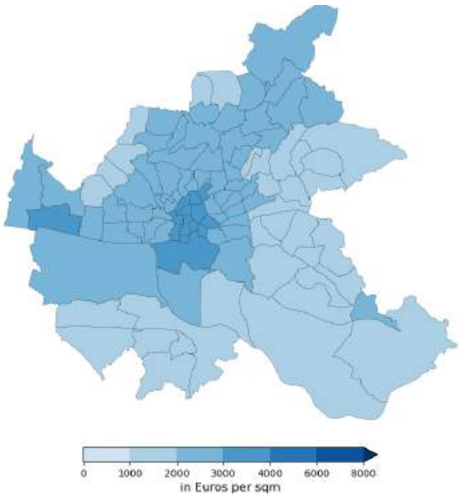
Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenio," *Working Paper*, 2019.

Wheaton, William C, "Income and Urban Residence: An Analysis of Consumer Demand for Location," *The American Economic Review*, 7 1977, *67* (4), 620–631.

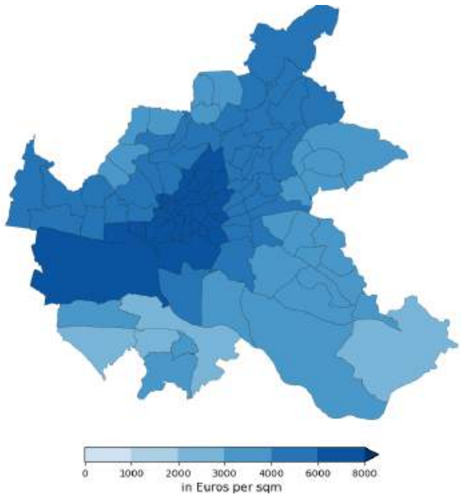
ONLINE APPENDIX—not for publication

1 Additional figures and tables

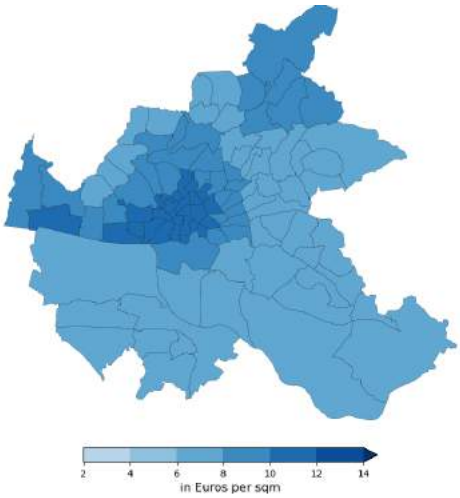
Figure A1: Postcodes Hamburg



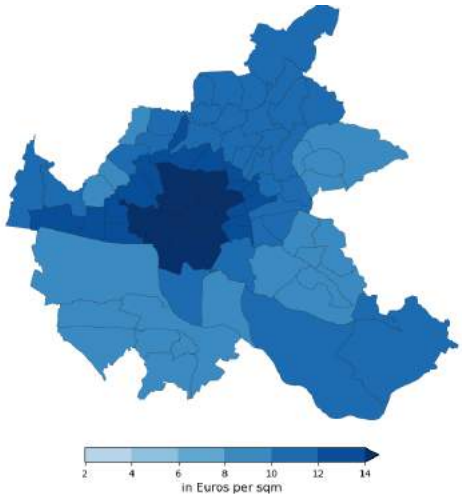
(a) Purchases, 2007



(b) Purchases, 2018



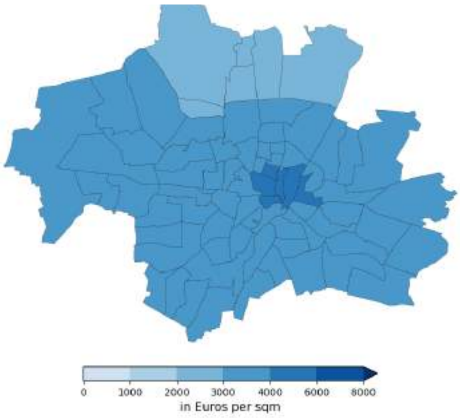
(c) Rents, 2007



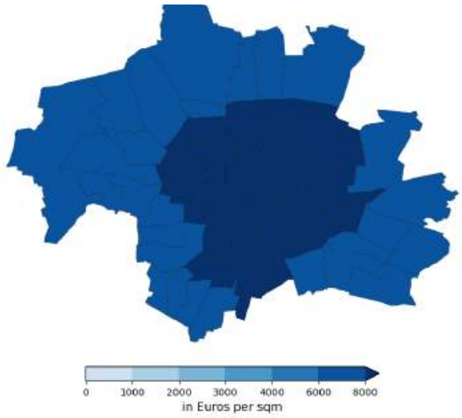
(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 101 postcode areas in Hamburg. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

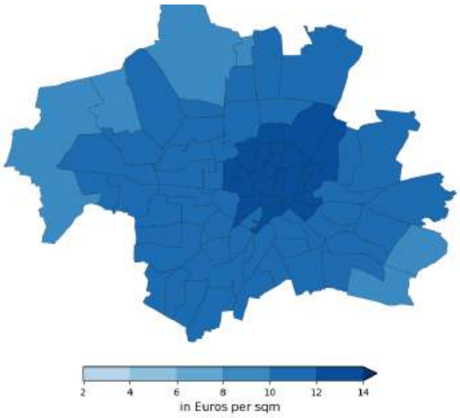
Figure A2: Postcodes Munich



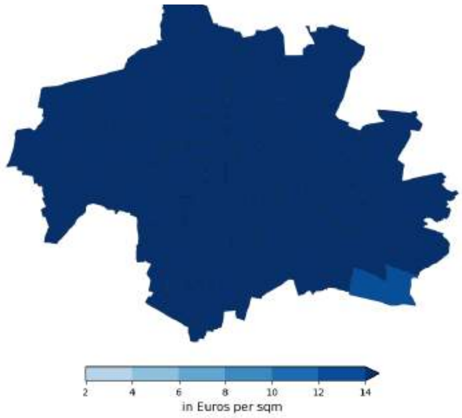
(a) Purchases, 2007



(b) Purchases, 2018



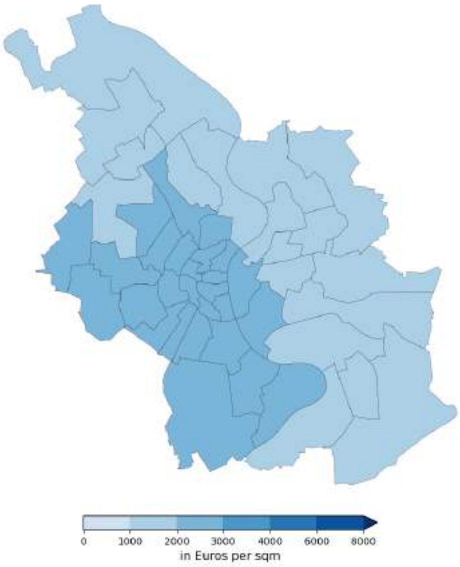
(c) Rents, 2007



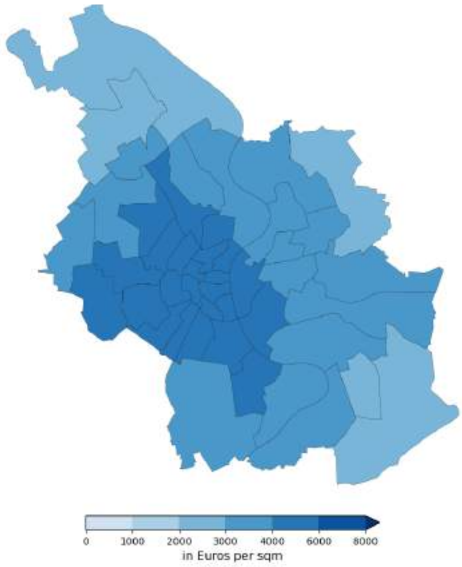
(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 74 postcode areas in Munich. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

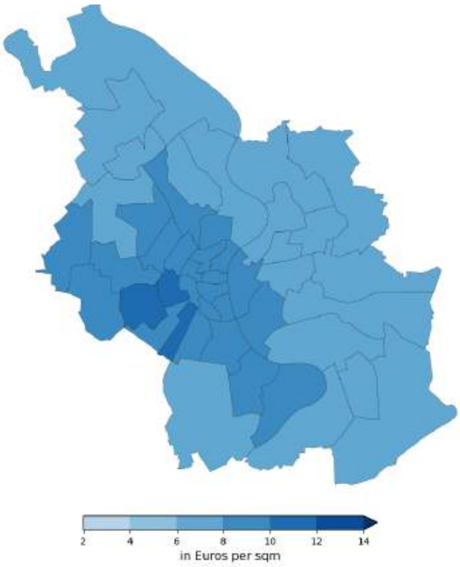
Figure A3: Postcodes Cologne



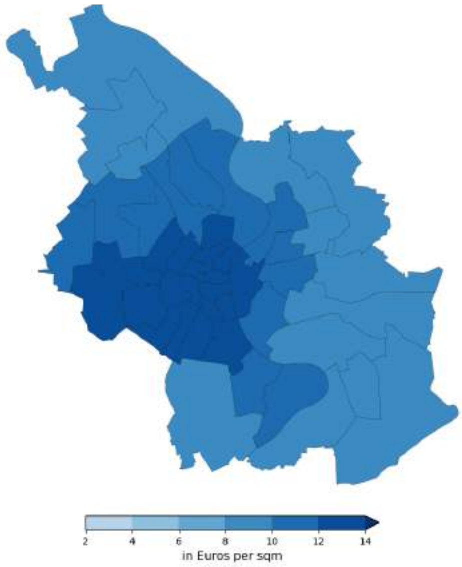
(a) Purchases, 2007



(b) Purchases, 2018



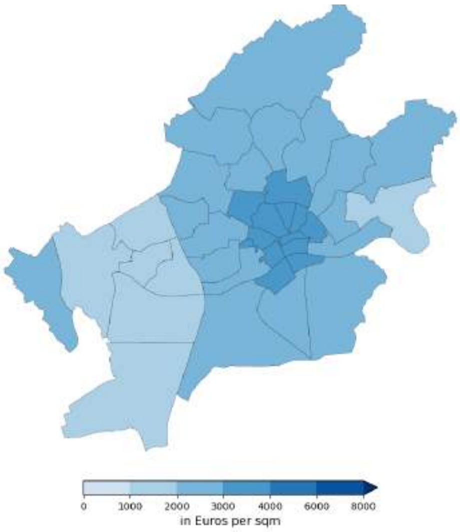
(c) Rents, 2007



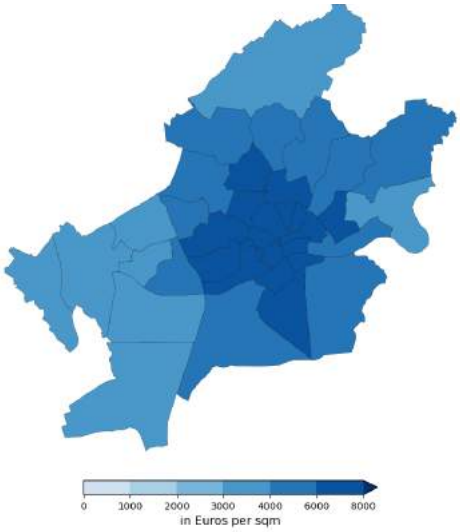
(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 45 postcode areas in Cologne. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).

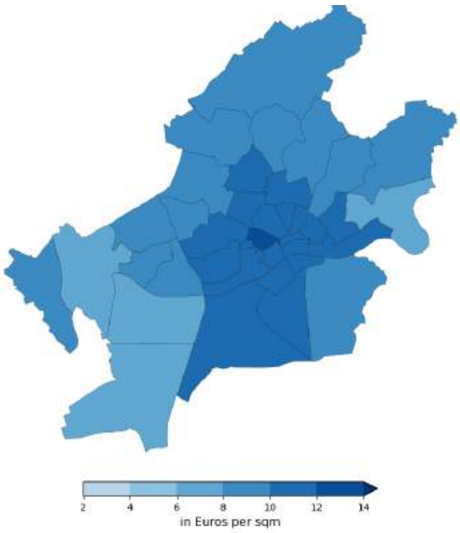
Figure A4: Postcodes Frankfurt



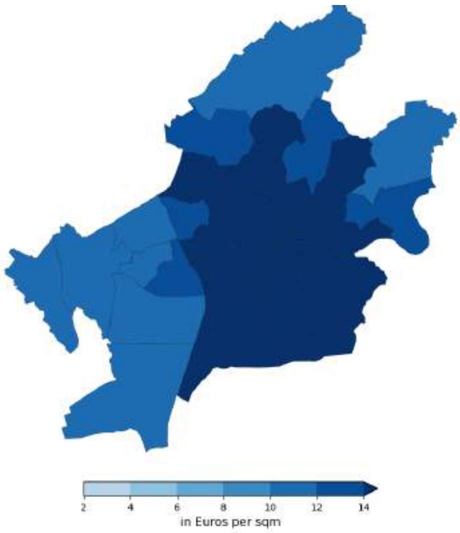
(a) Purchases, 2007



(b) Purchases, 2018



(c) Rents, 2007



(d) Rents, 2018

Note: Unit of observation in panels (a)-(d) 41 postcode areas in Frankfurt. The jurisdictional definition refers to 2019. Shapefiles are provided by the Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie).