RUHR

ECONOMIC PAPERS

Lea Eilers

Is My Rental Price Overestimated? A Small Area Index for Germany

rwi

#734

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #734

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-854-7

Ruhr Economic Papers #734

Lea Eilers

Is My Rental Price Overestimated? A Small Area Index for Germany

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788854
ISSN 1864-4872 (online)
ISBN 978-3-86788-854-7

Lea Eilers[1]

Is My Rental Price Overestimated? A Small Area Index for Germany

Abstract

Real estate prices are central to a range of themes that are, e.g., relevant for monetary policy, community development, environmental valuation, and economic planning more generally. This paper develops a real estate index based on apartment offer prices on the post code level for Germany, taking into account apartment heterogeneity and small sample sizes within regional areas as well as spatial and temporal dependencies. In a first step, a hedonic price function is estimated. In a second step, the residuals calculated from the hedonic function are used as direct estimates in a small area estimation (SAE). This technique is designed to yield estimates with a smaller variance in the context of small samples. The results show similarities between the estimates obtained from the residuals and SAE estimates. But the SAE models show non-negligible gains in accuracy for the coefficient of variance, i.e. the estimates are stabilized.

JEL Classification: R23, C01, R30

Keywords: Housing market; hedonic; small area estimation

December 2017

1 Lea Eilers, RWI. – I gratefully acknowledge the support of Thomas K. Bauer, Matthias Kaeding, Barbara Treude, and Colin Vance for helpful comments. – All correspondence to: Lea Eilers, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: lea.eilers@rwi-essen.de

1 Introduction

After years of moderately rising property prices, the German property market has experienced a strong increase in rental and purchasing prices since 2009. Especially in economically strong agglomeration areas, such as Berlin, Munich or Hamburg, rents have increased by five to ten percent each year.[1] The resulting discussion on affordable housing has induced politicians to shelter tenants with respect to rental payments. In a new law, the German government introduced a cap on rents (*Mietpreisbremse*). According to this law, the rental price of newly led apartments can only be ten percent higher than the local comparative rent.[2] Therefore, the comparative rent is of major interest in the rental-control design. But there is no generally valid definition nor calculation procedure, and each Federal State computes their own rental price index.

Embarrassments coming along with the calculation of a reliable rental price index have been shown 2014 in a court decision in Berlin. In a process of a rental increase (*Mieterhöhungsverfahren*) an independent expert concluded that the 2013 rental price index for Berlin (*Berlin Mietpreisspiegel 2013*) was insufficient in at least four points. First, the division of the entire urban area in only three residential areas was found to be arbitrary; second, the sample was not representative; third, gross rents were incorrectly converted into net rents; and fourth, extreme value adjustments were found to be inaccurate.

Against this background, this paper develops a way to calculate a regional rental index on the post code level. The index is based on rental prices of all apartments offered for rent within Germany as advertised on the internet platform *ImmobilienScout24* for the years 2014 and 2015. The study contributes to the literature by combining two different approaches in order to account for small sample size and apartment heterogeneity. Moreover, the computation of the index takes into account spatial and temporal interdependencies. Due to the applied method, all regions are comparable to each other and must not be interpreted to a baseline region.

Generating a nationwide regional price index for Germany faces two main challenges. Firstly, data on transacted or offered real estates are sparse. The reasons are twofold: (a) lack of data sources and (b) a relatively low transaction rate. The poor availability of regional real

[1] See <http://www.bpb.de/apuz/183444/herausforderungen-der-wohnungspolitik?p=all>.

[2] New buildings and comprehensively refurbished apartments are excluded. The areas for whom the rental control has to hold is defined by the federal states.

estate data restricts price indices to a relatively large regional scale or selected cities. Thereby, some price indices are based on simple averages or calculated using averages over real estate subgroups based on size, age or building type.[3] Secondly, instructive indices can only be computed if prices refer to apartments of comparable quality. Having heterogeneous apartments raises the question of whether a price change really reflects a price change or rather a change in the quality of local dwellings. In order to assure the comparability of apartments, a hedonic price function can be used, which explains the price of an apartment in terms of its price-determining attributes.[4]

As a consequence of these challenges, indices on small regional scales are only available for regions with a sufficient sample size. To tackle the problem of a small sample size within a regional area, I apply a small area estimation (SAE) approach. The advantage of the SAE approach is enhanced efficiency of the estimates in the case of a small sample size. To that end, the model assumes a fixed relationship between the statistics of interest and a set of covariates, which is available for all areas under consideration. This can be used to reduce standard errors and calculate more precise estimates. To encounter the problem of heterogeneous real estates, I estimate a hedonic price function and define the estimated residuals obtained from the hedonic regression as the price index. The calculated averaged residual can be seen as an indicator for unexplained characteristics that are different for each region and point in time.[5] Put differently, they identify regions in which the model under- or overpredicts apartment prices, which indicates that individuals pay lower or higher prices for the apartments than the amount attributable to observed characteristics.

The advantage of combining the hedonic price equation and the SAE technique is that the resulting index controls for both apartment heterogeneity and small sample sizes. In the context of the Berlin court decision, this index meets all points of criticism. First, the index is calculated on the post code level which makes a small-scale distribution within the city

[3] Examples include the *Deutscher Eigentums-Immobilienindex* (DEIX), *BORISplus.NRW*, *BBSR Angebotsmieten/-preise* and *BulwienGesa Immobilienindex*. These indices draw a picture of national developments but fail to reveal insights regarding within county or city developments (see Schürt [2010]).

[4] Examples for indices based on hedonic price equations are *Hypoport Hauspreisindex* (HPX), *vdp Immobilienpreisindizes*, *Niedersächsischer Immobilienindex* (NIDEX), *Destatis HPI–Schlüsselfertiger Neubau und bestehende Wohngebäude*, *empirica Preisdatenbank* and *F+B Marktmonitor*. These indices are calculated on the basis of different data sources and are available for different regional levels (see Schürt [2010]).

[5] This follows directly from the equilibrium result derived in Epple and Platt [1998] and Epple and Sieg [1999] that is, higher prices (beyond those predicted by the observable housing attributes and neighborhood characteristics) correspond to better unobserved neighborhood amenities.

possible. Second, the used dataset should be representative for those using the internet to search an apartment since it contains all advertised apartments offered on ImmobilienScout[6]. Third, the net rent is reported on the platform and must not be calculated. Fourth, the index adjust for extreme values and, based on the hedonic function, incorporates real price changes.

SAE techniques in the context of real estate research have been applied in two other papers that are closely related to this study: Articus and Burgard [2014] use German data but a slightly different strategy, and Pereira and Coelho [2013] use Portuguese data and a similar strategy. Both studies rely on data on the NUTS-3 level, while this study is able to define an index on the post code level. This study extends the previous literature by taking the heterogeneity of apartments into account. The calculated averaged residual can be seen as an indicator for unexplained characteristics determining the price that are different for each area. Moreover, this study uses apartment offer prices which show the latest adjustments in the market while inventory rents, used by Articus and Burgard [2014], against the background of the low transaction rate in Germany, would not reflect the current market situation.

The application of different small area estimation (SAE) techniques identifies the spatio-temporal Fay-Herriot (STFH) models as best fitting the data. The STFH model incorporates past time periods and bordering post code areas and shows substantial gains in accuracy for the coefficient of variance under the model-based estimators. Moreover, the direct estimates (the calculated residuals from the hedonic price function) are not very different from their model-based counterparts, but the model-based counterparts are generally more accurate. Geographically, the most expensive regions in Germany are in the south, with Munich and vicinity having the highest rental prices, together with Frankfurt, Stuttgart and some parts of Cologne, Berlin and Hamburg.

The remainder of this paper is structured as follows. Section 2 presents a description of the methodology. Section 3 describes the data construction and provides descriptive statistics. Estimation results are presented in Section 4 and Section 5 concludes.

[6] According to its website, *ImmobilienScout24* receives about 1.5 million different properties either for rent or for sale per month. It has more than 2 billion page impressions per month, with over 100,000 property sellers. Of those using the internet to search the market for real estate, about 88.5 percent use this portal. The platform covers about 35.7 percent of all rental contract conclusions in Germany.This information is available on the website of *ImmobilienScout24*: <https://www.immobilienscout24.de/>. Accessed 27 November 2014.

2 Empirical Strategy

2.1 Hedonic function

I start with a hedonic apartment price model defining the dependent variable as the log apartment offering price while the independent variables are apartment attributes and time indicators.

The hedonic apartment price function is given by

$$\log\left(p_{ijt}\right) = \mathbf{h}'_{ijt}\boldsymbol{\eta} + \tau_t + \nu_{ijt}, \tag{1}$$

where p_{ijt} is the price of apartment i in post code area j at time t. The vector $\mathbf{h}_{ijt}$ contains the characteristics of the apartment with $\boldsymbol{\eta}$ representing a vector of parameters to be estimated and $\boldsymbol{\tau}_t$ representing a vector of time fixed effects. The error term, ν_{ijt}, represents unobserved factors, including time- and region-varying factors that influence average house prices. The hedonic pricing model hence decomposes the rental price of an apartment into various components, such as physical and environmental attributes. The estimated parameters of the model provide information about the relative contribution of each of these apartment features. The neighborhood averages of the residuals after estimating Equation (1) via OLS can be seen as a price index, and will be used as the dependent variable in the small area estimation. Following Epple and Platt [1998] and Epple and Sieg [1999], there is a positive relationship between real estate prices and neighborhood quality. Higher prices (beyond those predicted by the observable housing attributes and neighborhood characteristics) correspond to better unobserved neighborhood amenities. Hence, the residual captures price determining characteristics that could not be explained by observable attributes.

The unit-specific residuals are given by $\hat{\nu}_{ijt} \equiv \log\left(p_{ijt}\right) - \widehat{\log\left(p_{ijt}\right)}$ where $\widehat{\log\left(p_{ijt}\right)}$ represents the predicted values of the apartment price. It follows that the average residual over each post code per year is obtained by

$$\hat{\hat{\nu}}_{jt}^{Dir} = (1/M_i)\sum_{i=1}^{M_i}\hat{\nu}_{ijt} \tag{2}$$

where $\hat{\hat{\nu}}_{jt}^{Dir}$ is the direct estimate in the following small area estimation. M_i is the number of observation per post code. The average residual is calculated regardless of the number of observations within a post code area.

2.2 Fay-Herriot Estimator

The estimation for small regional areas using disaggregated data suffers from small sample sizes. Therefore, using simple averages of the residuals over the post code areas as estimates for regional indices could lead to unacceptably large standard errors.[7] Empirical techniques, such as model-based small area estimation, are designed to produce reliable estimates even for very small sample sizes by making use of additional information.

The standard area level model without spatio- and temporal interactions, as defined by Fay and Herriot [1979], is given as follows[8]:

$$\hat{\bar{v}}_j^{Dir} = \mathbf{n}_j'\boldsymbol{\beta} + \lambda_j + \varepsilon_j, \tag{3}$$

where $\hat{\bar{v}}_j^{Dir}$ is the direct estimate of the average residual $\hat{\bar{v}}_{jt}^{Dir}$ obtained from Equation (2). $\mathbf{n}_j$ denotes a vector of auxiliary information on the post code level j with the corresponding regression coefficient $\boldsymbol{\beta}$. λ_j denotes an area specific random effect with the known model variance $\sigma^2_{\varepsilon,j}$. This random effect caputures the variation between the regions, which is determined by the differences in the auxiliary variables and not explained by $\mathbf{n}_j'\boldsymbol{\beta}$. ε_j is the error term, which in this case is the sample error of the direct estimate with unknown variance σ^2_λ. It is assumed that the errors λ_j and ε_j are independently distributed as $\lambda_j \overset{iid.}{\sim} \mathcal{N}(0, \sigma^2_\lambda)$ and $\varepsilon_j \overset{ind.}{\sim} \mathcal{N}(0, \sigma^2_{\varepsilon,j})$.

When estimating SAE models, the interest is not so much on the model coefficients themselves but rather in the prediction of the target variable from the model. The Empirical Best Linear Unbiased Predictor (EBLUP) for the parameter of interest (see Rao [2003]), $\bar{v}$, is given by

[7] The number of observations per post code area and year vary between 2 and 1759 observations, with a mean of 97 observations.

[8] For simplicity, the index t is removed from the equations if only one year is taken into account.

$$
\begin{aligned}
\hat{\hat{v}}_j^{FH} &= \mathbf{n}_j'\hat{\boldsymbol{\beta}} + \hat{\lambda}_j \\
&= \mathbf{n}_j'\hat{\boldsymbol{\beta}} + \hat{\gamma}_j(\hat{\hat{v}}_j^{Dir} - \mathbf{n}_j'\hat{\boldsymbol{\beta}}) \\
&= \hat{\gamma}_j\hat{\hat{v}}_j^{Dir} + (1-\hat{\gamma}_j)\mathbf{n}_j'\hat{\boldsymbol{\beta}}
\end{aligned} \tag{4}
$$

with

$$
\hat{\gamma}_j = \frac{\hat{\sigma}_\lambda^2}{\hat{\sigma}_{\epsilon,j}^2 + \hat{\sigma}_\lambda^2}. \tag{5}
$$

The Fay-Herriot estimator for the post code average residual of the hedonic house price function is $\hat{\hat{v}}_j^{FH}$. The parameters $\hat{\beta}$ and $\hat{\lambda}_j$ denote the Best Linear Unbiased Estimator (BLUE) for β and the Best Linear Unbiased Predictor (BLUP) for λ_j, respectively. Equation (4) shows that $\hat{\hat{v}}_j^{FH}$ is a weighted sum of the synthetic estimator $\mathbf{n}_j'\hat{\boldsymbol{\beta}}$, obtained from the fixed part of the model, and the direct estimator $\hat{\hat{\boldsymbol{v}}}_j^{\boldsymbol{Dir}}$. This composition is weighted by the area specific shrinkage factor $\hat{\gamma}_j$ setting the model variance $\sigma_{\hat{\lambda}}^2$ in relation to the total variance $\sigma_{\epsilon,j}^2 + \sigma_{\hat{\lambda}}^2$ given in Equation (5). If the model variance is small compared to the design variance, $\hat{\gamma}_j$ is close to zero and the synthetic estimator dominates. Intuitively, the higher $\hat{\gamma}_j$, the lower the confidence in the model-based estimator [Rao 2003; Articus and Burgard 2014]. The relevant measure to judge the quality of model-based small area estimates is the mean squared error (MSE), that is $MSE(\hat{\hat{v}}_j^{FH}) = E(\hat{\hat{v}}_j^{FH} - v_j)^2$.

2.3 Spatial Estimator

Apartment prices as well as neighborhood characteristics are spatially correlated, decreasingly so with spatial distance, meaning that apartments located nearby each other tend to have similar prices. Therefore, spatial dependence among the random area effects, λ_j, needs to be taken into account. According to Salvati [2004], the Fay-Herriot model proposed in Equation 3 can be extended with spatially correlated effects in order to account for spatial relationships between neighboring post code areas, thereby leading to more reliable predictions.

The vector λ_j of Equation (3) of neighborhood effects allows a first order simultaneous autoregressive, SAR(1), process, that is

$$\boldsymbol{\lambda_j} = \rho^{SP}\boldsymbol{W}\boldsymbol{\lambda_j} + \epsilon_j^{SP} \tag{6}$$

with $\epsilon_j^{SP} \sim \mathcal{N}(0, \sigma^{2SP})$. ρ^{SP} is an autoregression spatial parameter with $\rho^{SP} \in (-1, 1)$ and $\boldsymbol{W}$ is a $J \times J$ proximity matrix obtained by a row-wise standardization of an initial matrix with zeros on the diagonal and the remaining entries equal to one when the row domain have a first oder common border (queen-matrix) with the column domain [Anselin 1988]. ϵ_j^{SP} is a vector of independet error terms with zero mean and constant unknown variance $\sigma^2_{\epsilon_j^{SP}}$.

The EBLUP for the spatial Fay Herriot model presented in Equation (6) are given in Petrucci and Salvati [2004]. A detailed description on how to estimate the covariance matrix can be found in Salvati [2004] and in Pratesi and Salvati [2008]. The MSE for the spatial model depends on the two parameters $(\sigma_\varepsilon^2, \rho^{SP})$ and is calculated as in Singh, Shukla and Kundu [2005].

2.4 Spatio-temporal Estimator

Apartment prices are available for various years on post code level and therefore the SFH model is further extended by including random effects for the time periods nested within the regional level to account for temporal correlation. Marhuenda, Molina and Morales [2013] proposed this model as the spatio-temporal Fay-Herriot model (STFH).

The model is given by

$$\hat{\nu}_{jt}^{Dir} = \boldsymbol{n}'_{jt}\beta + \lambda_{jt} + \eta_{jt} + \varepsilon_{jt} \tag{7}$$

and λ_{jt} follow, as in Equation (6), an SAR(1) process. For each post code area j, the vector $\boldsymbol{\eta_j} = (\eta_{j1}, ..., \eta_{jT})'$ is i.i.d. following the first order autoregressive, AR(1), process

$$\eta_{jt} = \rho^{time}\eta_{j,t-1} + \epsilon_{jt}^{time} \tag{8}$$

with $\epsilon_{jt}^{time} \sim \mathcal{N}(0, \sigma^{2,time})$. In the case of spatio-temporal Fay-Herriot models, a parametric bootstrap procedure is used to find the Mean Squared Prediction Error (MSPE). A detailed description on how the bootstap works can be found in Pereira and Coelho [2013]. The EBLUP and MSE are calculated as in Marhuenda, Molina and Morales [2013].

3 Data Description

The empirical analysis is based on a unique dataset. First, I use individual real estate data from the RWI-GEO-RED, which contains information on all German apartments and houses for sale and rent that were advertised on *ImmobilienScout24*. Second, these data are combined with neighborhood data on the post code level obtained from the RWI-GEO-GRID. These data contain information on socio-economic characteristics of the post code area.

The estimation sample for the hedonic pricing model consists of apartments for rent. The dependent variable measures offer prices. The rental price is measured in Euro and enters into the regression as its log.[9] For the analysis, the price last offered is used, and apartments in construction are deleted from the sample.

The explanatory variables capture apartment characteristics, such as the age of the apartment in years, its size, equipment variables such as balcony, fitted kitchen, garden, elevator and/or cellar, as well as a set of categorical variables indicating its type and state. These variables constitute the vector $\mathbf{h}_{ijt}$ and are used as covariates in Equation (1) to predict apartment prices. Descriptive statistics of apartment characteristics are reported in Table 1.[10]

Neighborhood information is obtained from the RWI-GEO-GRID data [RWI 2016*a*,*b*,*c*,*d*,*e*,*f*]. The RWI-GEO-GRID data are based on data by *microm Micromarketing-Systeme und Consult GmbH*, a commercial micro- and geomarketing provider. Microm uses more than a billion individual data points for the aggregation of their dataset. For this study, the data are aggregated on the post code level.

The variable share of foreigners is based on an analysis of first and surnames with respect to their linguistic origin. The evaluation refers to the head of the household. The unemployment rate is the share of the unemployed in the population of those working or searching for employment. The purchasing power reflects the average household income on the post code level. It comprises information on labor supply, capital wealth, rental income minus taxes and social security contributions, including social transfers such as unemployment benefits, child-allowances and pensions. Regular payments, e.g. for rent, electricity or insurance premiums are not subtracted from the purchasing power.

[9] The sample is trimmed by dropping the observation at the highest and lowest one percent concerning the rental price, the number of rooms, the overall living area and the age of the apartment.

[10] For a further description of the real estate data see an de Meulen, Micheli and Schaffner [2014].

TABLE 1
SUMMARY STATISTICS FOR APARTMENT CHARACTERISTICS

Variable	Mean	Std. Dev.
Price (sq.m)	6.93	2.47
Number of Rooms	2.62	0.89
Area	72.49	24.9
Age	50.01	33.68
Has cellar	0.69	0.46
Has elevator	0.2	0.4
Has garden	0.21	0.41
Has balcony	0.68	0.47
Has fitted kitchen	0.37	0.48
State: Like New/ First move in	0.13	0.34
State: Renovated	0.24	0.43
State: Modernized, well-kept	0.38	0.49
State: Not Renovated or not stated	0.24	0.43
Type: Top floor	0.13	0.34
Type: Loft, Maisnette, Penthouse, Terrace flat	0.07	0.25
Type: Floor apartment	0.12	0.33
Type: Apartment	0.48	0.5
Type: Mezzanine, Basement	0.04	0.19
Type: Other	0.16	0.37

NOTES.—The number of observations is 1,166,507, pooled for 2014 and 2015.
SOURCE.—Author's calculations based on RWI-GEO-RED.

The variable house type indicates the most common size of the buildings in the street segmented in which the house is located: (1) single- and two-family homes in heterogeneous or homogeneous road sections, (2) 3-9 family homes and (3) blocks of flats with 10 or more households. The share of dwelling house report the absolut number of dwellings within a post code. The payment default variable describes the statistical probability of payment default for each post code area in Germany. Summary statistics for the neighborhood information are reported in Table 2. For a detailed description see microm [2017] and Budde and Eilers [2014].

The analysis focuses on the period from 2014 to 2015. Apartment price indices are only calculated when the averaged residual is available for both years. The analysis relies on 6,005 out of 8,208 post code areas in Germany. Out of these 6,005 post code areas more than 3,300 in 2015 and 3,400 in 2014 have less than 50 observations, showing the relevance of small sample sizes on the post code level in the rental market in Germany. Post code areas with less than 50 observations are mainly located in rural areas (see Figure 5).

TABLE 2
SUMMARY STATISTICS FOR NEIGHBORHOOD CHARACTERISTICS

Variable	Mean	Std. Dev.	Min.	Max.
Share dwelling house	2975	2169	125	17656
Share single- and two-familiy houses	0.49	0.21	0.01	0.9
Share 3-9 family homes	0.34	0.1	0.04	0.73
Share building blocks with 10 or more households	0.15	0.18	0	0.93
Share low payment default	0.39	0.18	0	0.89
Share medium payment default	0.34	0.08	0.01	0.68
Share high payment default	0.31	0.07	0.06	0.78
Purchasing power per household	44870	8001	20857	104934
Share foreigners	0.06	0.05	0	0.31
Unemployment rate	0.06	0.035	0	0.22

NOTES.—The number of observations is 12,010, pooled for 2014 and 2015.
SOURCE.—Author's calculations based on RWI-GEO-GRID.

4 Results

4.1 Hedonic Price Model

Table 3 displays the results obtained from estimating the hedonic house price equation (Equation 1). Column (1) presents the estimation results for 2014 and Column (2) presents the estimation results for 2015.

Generally, the coefficient estimates conform to expectations. For instance, larger units are associated with higher prices, and apartments without cellars, balcony or fitted kitchen are rented for a lower price. It is also notable that the age of the unit has a nonlinear effect on the renting price. Relative to the average apartment, flats located on the top floor and on the ground floor are associated with lower prices while units in the category "Loft, Maisonette, Penthouse, Terrace flat" are rented for significantly higher prices. Renovated apartments as well as modernized or well-kept apartments and not renovated apartments are associated with a lower renting price than a newly constructed unit. The purpose of estimating the hedonic pricing model is to predict housing prices, which the model does reasonably well as indicated by the R^2 of about 68 percent. The differences in the predicted and actual house prices are then averaged at the post code level.

The regional distribution of the residuals is shown in Figure 1. The darker areas indicate neighborhoods where the hedonic price model underpredicts prices, i.e., indicating post code areas in which individuals are willing to pay more than the amount attributable to observed

TABLE 3
APARTMENT PRICE REGRESSIONS

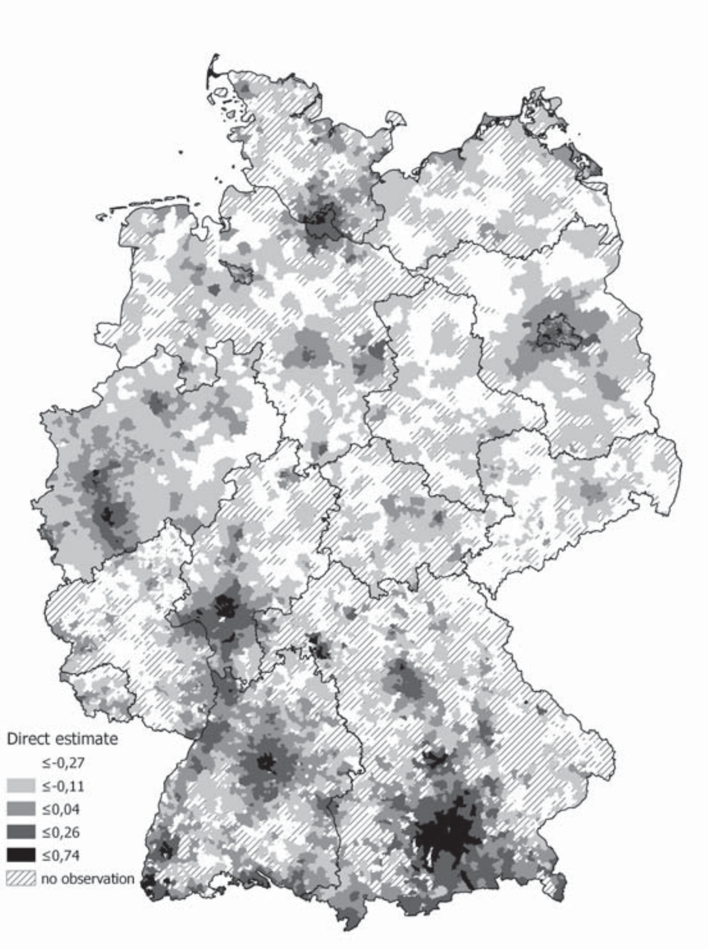
	(1) 2014	(2) 2015
Number of Rooms	-0.0463***	-0.0576***
	(0.0029)	(0.0027)
Has cellar	0.0284***	0.0123***
	(0.0035)	(0.0033)
Has elevator	0.0968***	0.0964***
	(0.0049)	(0.0048)
Has garden	-0.0377***	-0.0330***
	(0.0037)	(0.0036)
Has balcony	0.0688***	0.0542***
	(0.0029)	(0.0029)
Has fitted kitchen	0.1769***	0.1851***
	(0.0042)	(0.0041)
Area	0.0230***	0.0240***
	(0.0003)	(0.0003)
Area (sq)	-0.0001***	-0.0001***
	(0.0000)	(0.0000)
State: Renovated	-0.0431***	-0.0472***
	(0.0044)	(0.0048)
State: Modernized, well-kept	-0.0857***	-0.0784***
	(0.0038)	(0.0042)
State: Not Renovated or not stated	-0.1180***	-0.1113***
	(0.0048)	(0.0050)
Age	-0.0116***	-0.0112***
	(0.0005)	(0.0005)
Age sq.	0.1962***	0.1885***
	(0.0101)	(0.0108)
Age cu.	-0.0009***	-0.0009***
	(0.0001)	(0.0001)
Type: Top floor	-0.0213***	-0.0141***
	(0.0028)	(0.0031)
Type: Loft, Maisonette, Penthouse, Terrace flat	0.0308***	0.0344***
	(0.0034)	(0.0036)
Type: Floor apartment	-0.0200***	-0.0140***
	(0.0027)	(0.0027)
Type: Mezzanine, Basement	0.0124***	0.0162***
	(0.0042)	(0.0056)
Type: Other	-0.0363***	-0.0302***
	(0.0048)	(0.0047)
Observations	589,487	576,457
R^2	0.689	0.678

NOTES.—Constant is included but not reported. The reference category for the state indicators is "like new / first move-in"; for the type indicators, it is "Apartment". Standard errors are robust to clustering at the post code level and are presented in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
SOURCE.—Author's calculations based on RWI-GEO-RED.

housing and neighborhood characteristics. The lighter areas correspond to neighborhoods where the model overpredicts prices.

In general, Figure 1 indicates that the model underpredicts apartment rents in cities, most notably in Munich, Stuttgart, Frankfurt, Hamburg and Berlin. This indicates that there are

FIGURE 1
DIRECT ESTIMATE AT THE POST CODE LEVEL 2015



NOTES.—Post code areas are restricted to those having observation in 2014 and 2015. The share is presented at the post code level for 2015.
SOURCE.—Author's calculations based on RWI-GEO-RED 2015.

important unobservable factors in these regions that drive apartment rents upwards. These factors could, for example, take the form of agglomeration or network effects that are usually observed in cities and other densely populated areas. Conversely, one can also see that overprediction of the price predominantly occurs in more rural areas.

4.2 Small Area Estimation

The application of the SAE models realizes gains in precision in the context of small sample sizes and decreases the large standard errors of traditional direct estimates which leads to more reliable estimates of the rental price index.

All models presented in Section 2 are estimated using the auxiliary variables presented in Table 2. Standard information criteria are used to identify the model fitting the data best. Hence, the spatio-temporal Fay-Herriot (STFH) model is identified to produce the most reliable estimates for a rental price index on post code areas in Germany (see Table A.1).

Table 4 presents descriptive statistics of the small area estimates and the coefficient of variance for the direct estimates as well as for the different models applying small area techniques. The SFH and STFH model are based on a first order common border spatial queen matrix, so that the spatial weight is defined only by those neighborhoods that share a common border.[11]

TABLE 4
DESCRIPTIVE STATISTICS

	Estimates				Coefficient of variance			
	Direct	FH	SFH	STFH	Direct	FH	SFH	STFH
Minimum	-0.75	-0.67	-0.70	-0.66	-896,900	-25,390	-17,470	-49,939
Lower quartile	-0.24	-0.23	-0.25	-0.25	-105.0	-23.76	-23.6	-19.49
Median	-0.11	-0.11	-0.11	-0.13	-50.1	-10.14	-10.0	-9.60
Mean	-0.08	-0.08	-0.08	-0.10	-146.3	-13.35	57.9	-18.53
Upper quartile	0.05	0.04	0.04	0.02	53.8	4.78	4.9	3.50
Maximum	0.74	0.71	0.72	0.69	167,200	45,980	376,700	16,144

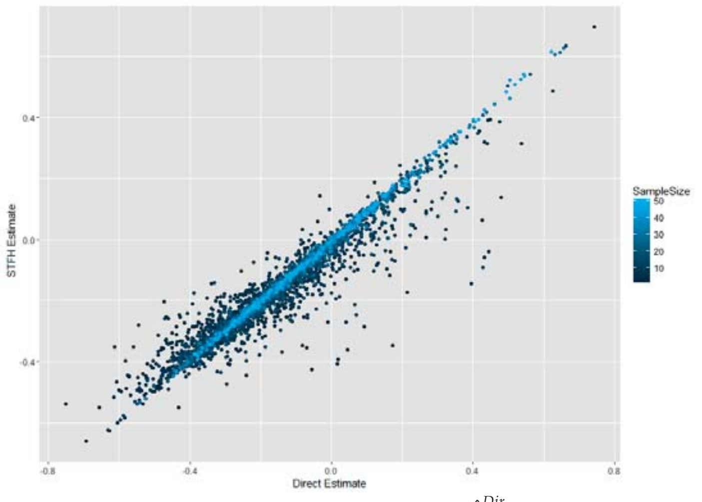
NOTES.—Direct denotes the averaged residual over the post code area, while FH defines the Fay-Herriot model, SFH the spatial Fay-Herriot model and SFTH the spatio-temporal Fay-Herriot model.
SOURCE.—Author's calculations based RWI-GEO-RED and RWI-GEO-GRID.

On the one hand, the direct estimate and the estimates of the SAE-models are quite similar, but the estimates of the STFH model are slightly more stable compared to the direct estimates and shows the smallest minimum and maximum value. On the other hand, there are clear differences in the coefficient of variance (CV) between the four models. While the CV of the direct estimates ranges between -896,900 to 167,200, the CV of the SAE-models are much more precise. The smallest range is observed in the STFH model with a minimum of -49,939 and maximum value of 16,144. Moreover, the STFH model presents a much lower mean and me-

[11] For that reason, post code areas without any neighbor are dropped from the sample. Since we cannot observe apartment offers in all post code areas, some post code areas do not have any bordering neighbor. This is true for 33 post code areas within this sample. The results are estimated using the R-package "sae" by Molina and Marhuenda [2015] based on Marhuenda, Molina and Morales [2013].

dian CV compared to the CV of the direct estimates, but the median is quite similar for all SAE-models. The estimates as well as the CV calculated for the STFH model on the post code level are presented in Figures 2 and 3 for areas having less than 50 offered apartments in 2015.

FIGURE 2
COMPARISON OF THE DIRECT ESTIMATES AND STFH ESTIMATES



NOTES.—Estimates based on Equation 7 for the direct estimates ($\hat{\bar{v}}_j^{Dir}$) and the STFH estimates. The figure presents the results for post codes having less than 50 observations in 2015.
SOURCE.—Author's calculations based on RWI-GEO-RED and RWI-GEO-GRID.

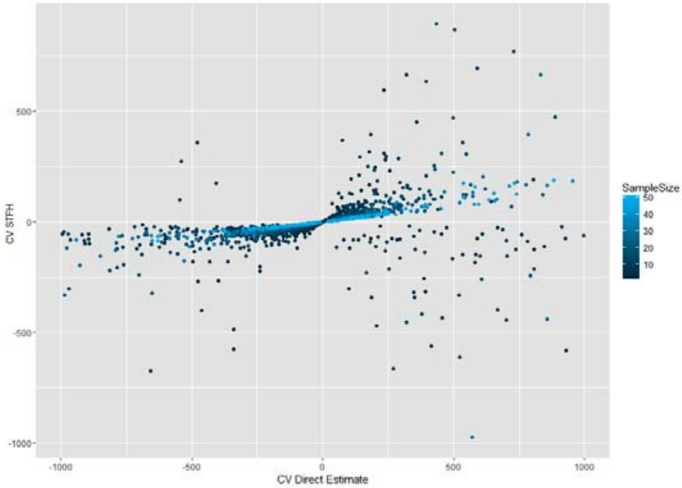
Figure 2 compares the direct estimates calculated as in Equation 2, presented on the *abscissa*, and the STFH estimates computed by means of Equation 7, presented on the *ordinate*. The darker the dots the less apartments are offered in the corresponding post code area in 2015. The figure shows that the SAE estimates are not very different from their direct counterparts, but they are generally more accurate, especially for small sample sizes. With increasing sample size, the parameters of the direct estimates and the estimates of the STFH model become more similar as the direct estimates get more weight.

Indicated by Table 4 and supported by Figure 3, non-negligible gains in accuracy are obtained for the CV under the model-based (STFH) estimators. The application of the STFH model, i.e. taking into account information from related small areas and past time periods, improves the estimates, especially for post code areas with small sample sizes.[12]

The regional distribution of the rental price index based on the STFH model, presented in

[12]Figures 6 and 7 show the results for the full sample.

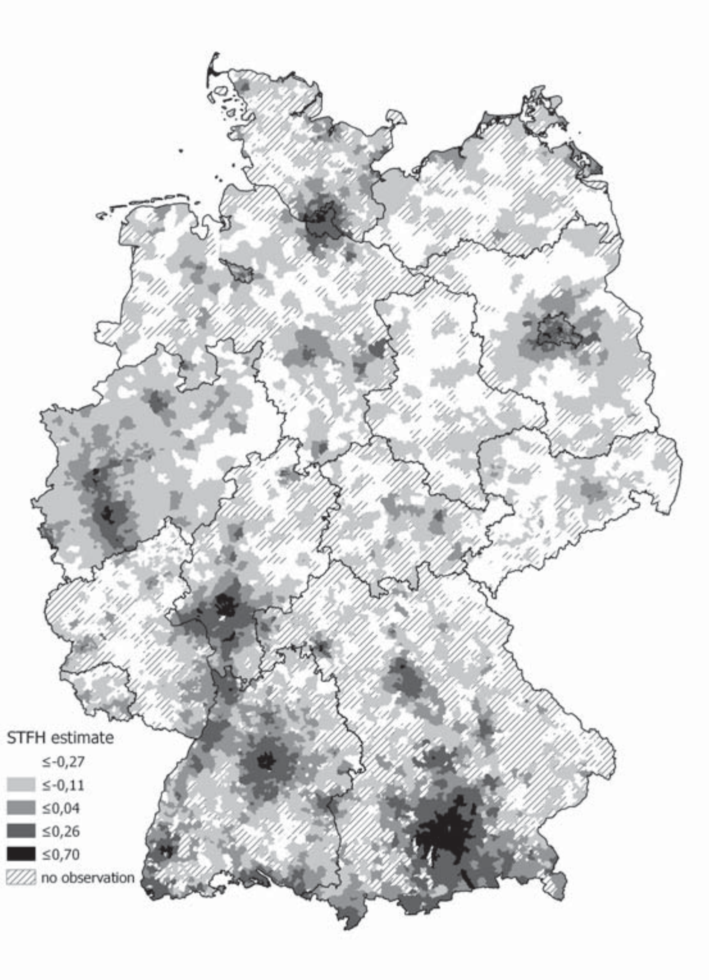
FIGURE 3
COMPARISON OF THE COEFFICIENTS OF VARIANCE (CV)



NOTES.—Coefficients of variance for the direct estimates ($\hat{v}_j^{Dir}$) and the STFH estimates. The coefficient of variance is the ratio of the standard deviation to the mean. The figure presents the results for post codes having less than 50 observations in 2015. Two extreme outliers are dropped from the figure for better readableness.
SOURCE.—Author's calculations based on RWI-GEO-RED and RWI-GEO-GRID.

Figure 4, shows how the model over- and underpredicts prices. Darker areas are underpredicted prices, while lighter areas show overpredicted prices. Munich and areas in the vicinity are the most obvious areas showing underpredicted prices, meaning that individuals are willing to pay more than the amount attributable to observed apartment and neighborhood characteristics. Likewise, the STFH model underpredicts prices in Stuttgart, Frankfurt and some neighborhoods of Berlin, Hamburg and Cologne. These regions are characterized by a high turnover rate and therefore have a sufficient number of observations to calculate reliable indices. Post code areas located in rural areas tend to have overpredicted prices, indicating that individuals pay lower prices for the apartments than the amount attributable to observed apartment and neighborhood characteristics would expect. These are also the areas which have a relatively low turnover rate in apartments and would not lead to reliable indices under the direct estimates. For example, the rural areas in the north and south of the Federal State Brandenburg show overpredicted prices in the STFH model but not for the direct estimates. Especially for these areas applying SAE techniques gives an improvement in real estate research on small areas.

FIGURE 4
SPATIO-TEMPORAL FAY-HERRIOT ESTIMATOR AT THE POST CODE LEVEL



NOTES.—Post code areas are restricted to those having observation in 2014 and 2015 and at least one neighbor.
SOURCE.—Author's calculations based on RWI-GEO-RED 2015 and 2014.

The estimated coefficients of the auxiliary variables are of less importance in the setting of SAE estimation but are presented in Table A.2. The estimates of all auxiliary information are significant on the 1 percent level but differ in sign and strength between the different models. A positive sign of coefficient means that the increase of that particular indicator in an area results in an increase in the price index.

5 Conclusion

In economically strong agglomeration areas, increasing segregation has become a major topic in political debates since segregation can be severed by self-selection of individuals into certain neighborhoods based on preferences and purchasing power. As a consequence on the discussion on affordable housing, politicians introduced a cap of rents in some cities to shelter tenants with respect to rental payments. This cap of rents defines an upper limit for the rent increase of newly led apartments based on a comparative rent. The calculation of this comparative rent is organized differently in each Federal State and could, as seen in the court decision 2014 in Berlin, be insufficient for different reasons.

This paper applies a way to calculate a rental price index for apartments. To that end, a two-step procedure is applied to generate a reliable rental price index on post code areas. Firstly, a hedonic price function that explains the price of an apartment characteristics is estimated. Afterwards, the calculated residual of the hedonic price function can be interpreted as an indicator for unexplained characteristics determining rental prices. These results are not reliable estimators because they are based on a relatively small sample size caused by the low turnover rate of rental apartments in Germany. Therefore, in the second-step, SAE techniques are applied to estimate reliable price indices for all available post code areas. These techniques are designed to produce reliable information even in the case of small sample sizes.

The presented index is based on two different data sources. Firstly, all apartments offered for rent within Germany as advertised on the internet platform *ImmobilienScout24* between 2014 and 2015 are used. Secondly, the SAE model additionally relies on socio-economic characteristics on the post code level obtained from the RWI-GEO-GRID.

Testing different SAE models, the spatio-temporal Fay-Herriot estimator is found to be the most efficient model in estimating a nationwide rental price index. The results show that non-negligible gains in accuracy are obtained for the coefficient of variance under the model-based estimators that borrow information from related small areas and past time periods as well as neighborhood characteristics. Moreover, the direct estimates (residuals from the hedonic price function) are not very different from their model-based counterparts, but they are generally less accurate, mainly for small sample size.

References

an de Meulen, Philipp, Martin Micheli and Sandra Schaffner. 2014. Documentation of German real estate market data: Sample of real estate advertisements on the internet platform ImmobilienScout24, 2007–2013. Technical report RWI - Leibniz-Institute for Economic Research.

Anselin, Luc. 1988. *Spatial econometrics: methods and models*. Vol. 4 Springer Science & Business Media.

Articus, Charlotte and Jan Pablo Burgard. 2014. A Finite Mixture Fay Herriot-type model for estimating regional rental prices in Germany. Technical report Research Papers in Economics.

Budde, Rüdiger and Lea Eilers. 2014. Sozioökonomische Daten auf Rasterebene: Datenbeschreibung der microm-Rasterdaten. Technical report RWI Materialien.

Epple, Dennis and Glenn J Platt. 1998. "Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes." *Journal of Urban Economics* 43(1):23–51.

Epple, Dennis and Holger Sieg. 1999. "Estimating equilibrium models of local jurisdictions." *Journal of Political Economy* 107(4):645–681.

Fay, Robert E and Roger A Herriot. 1979. "Estimates of income for small places: an application of James-Stein procedures to census data." *Journal of the American Statistical Association* 74(366a):269–277.

Marhuenda, Yolanda, Isabel Molina and Domingo Morales. 2013. "Small area estimation with spatio-temporal Fay–Herriot models." *Computational Statistics & Data Analysis* 58:308–325.

microm. 2017. *Datenhandbuch 2017*. microm Micromarketing-Systeme und Consult GmbH.

Molina, Isabel and Yolanda Marhuenda. 2015. "sae: An R Package for Small Area Estimation." *The R Journal* 7(1):81–98.
URL: *http://journal.r-project.org/archive/2015-1/molina-marhuenda.pdf*

Pereira, Luis Nobre and Pedro Simões Coelho. 2013. "Estimation of house prices in regions with small sample sizes." *The Annals of Regional Science* 50(2):603–621.

Petrucci, Alessandra and Nicola Salvati. 2004. Small Area Estimation considering spatially correlated errors: the unit level random effects model. Technical report Working Paper.

Pratesi, Monica and Nicola Salvati. 2008. "Small area estimation: the EBLUP estimator based on spatially correlated random area effects." *Statistical Methods & Applications* 17(1):113–141.

Rao, J. N. K. 2003. *Small Area Estimation*. John Wiley Sons, Inc.

RWI, microm. 2016*a*. *RWI-GEO-GRID: Socio-economic data on grid level (Wave 5) - House type*. Vol. 1 RWI Leibniz Institute for Economic Research.
URL: *http://fdz.rwi-essen.de/doi-detail/id-107807micromhaustypv4.html*

RWI, microm. 2016*b*. *RWI-GEO-GRID: Socio-economic data on grid level (Wave 5) - Purchasing power.* Vol. 1 RWI – Leibniz Institute for Economic Research. Dataset. https://doi.org/10.7807/microm:kkp:V4.

RWI, microm. 2016*c*. *RWI-GEO-GRID: Socio-economic data on grid level (Wave 5) - Share of foreigners.* Vol. 1 RWI – Leibniz Institute for Economic Research. Dataset. https://doi.org/10.7807/microm:auslaender:V4.
URL: *http://fdz.rwi-essen.de/doi-detail/id-107807micromauslaenderpv4.html*

RWI, microm. 2016*d*. *RWI-GEO-GRID: Socio-economic data on grid level (Wave 5) - Unemployment rate.* Vol. 1 RWI – Leibniz Institute for Economic Research. Dataset. https://doi.org/10.7807/microm:alq:V4.
URL: *http://fdz.rwi-essen.de/doi-detail/id-107807micromalqv4.html*

RWI, microm. 2016*e*. *RWI-GEO-GRID: Socio-economic data on grid level (Wave 5) Payment index*. RWI Leibniz Institute for Economic Research.
URL: *http://fdz.rwi-essen.de/doi-detail/id-107807micromzahlindexv4.html*

RWI, microm. 2016*f*. *RWI-GEO-GRID: Socio-economic data on grid level (Wave 5) Population*. RWI Leibniz Institute for Economic Research.
URL: *http://fdz.rwi-essen.de/doi-detail/id-107807micromeinwohnerv4.html*

Salvati, Nicola. 2004. Small Area Estimation by Spatial Models: The Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP). Working Paper 2004/03 University of Florence.

Schürt, Alexander. 2010. Synopse Immobilienpreisbeobachtung in Deutschland 2010. AnforderungenDatengrundlagenVerfahrenProdukte. Working paper Bundesinstitut für Bau–, Stadt-und Raumforschung.

Singh, Bharat Bhushan, Girja Kant Shukla and Debasis Kundu. 2005. "Spatio-temporal models in small area estimation." *Survey Methodology* 31(2):183.

Appendix A Tables

TABLE A.1
GOODNESS OF FIT

	Dependent variable: $\hat{v}_j^{Dir}$		
	ll	AIC	BIC
Fay-Herriot model	3,831	-7,641	-7,574
Spatial Fay-Herriot model	1,558	-3,094	-3,020
Spatio-temporal Fay-Herriot model	8,885	-17,745	-17,657

NOTES.—ll denotes the log- likelihood value, AIC is the Akaike information criterion and BIC is the Bayesian information criterion.
SOURCE.—Authors' calculations based RWI-GEO-RED and RWI-GEO-GRID.

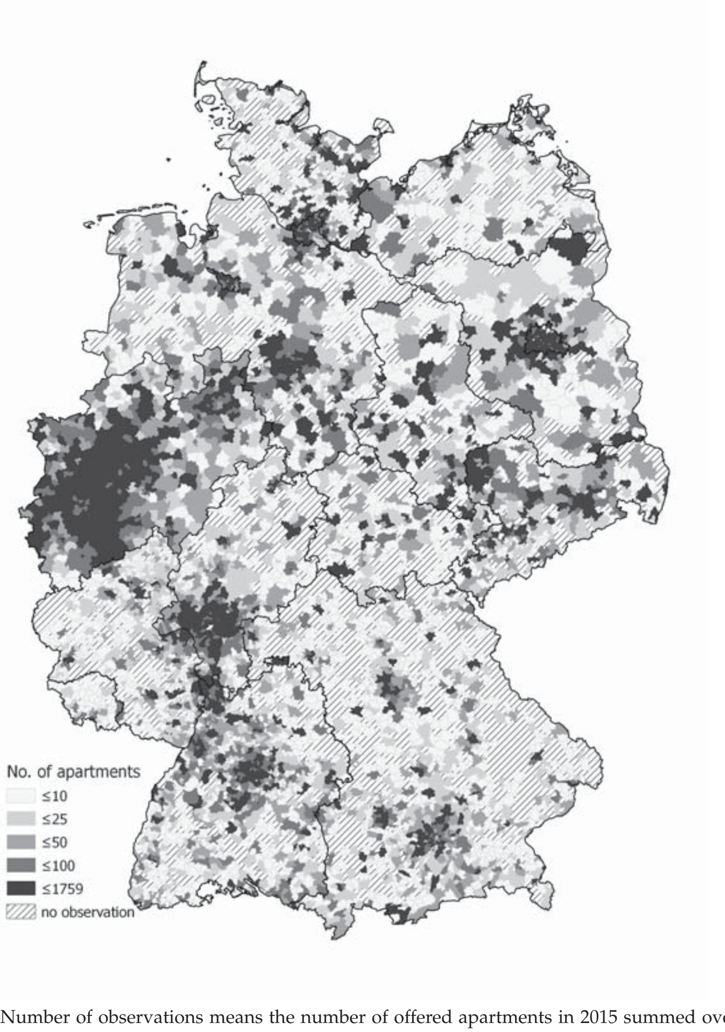
TABLE A.2
REGRESSION COEFFICIENTS

	Dependent variable: $\hat{v}_j^{Dir}$		
	Fay-Herriot (1)	spatial Fay-Herriot (2)	spatio-temporal Fay-Herriot (3)
Share foreigners	0.995***	0.939***	0.975***
	0.0525	0.0507	0.045
Unemployment rate	-0.009***	-0.0117	-0.010***
	0.0009	0.0039	0.0009
Purchasing power per household	0.01***	0.01***	0.009***
(÷ 1000)	0.0000	0.0000	0.0000
Share single- and two-family homes	-0.751***	-0.811***	-0.762***
	0.01714	0.1341	0.0163
Share 3-9 family homes	-0.325***	-0.429***	-0.322***
	0.0207	0.1329	0.019
Share low payment default	0.0786***	0.041***	0.092***
	0.0239	0.02909	0.024
Share medium payment default	0.274***	0.223***	0.226***
	0.0273	0.0295	0.0294
Share dwelling house (÷ 1000)	0.002***	0.002***	0.004***
	0.0000	0.0000	0.0000

NOTES.—Results are based on 12,010 observations pooled for the years 2014 and 2015.
SOURCE.—Authors' calculations based RWI-GEO-RED and RWI-GEO-GRID.

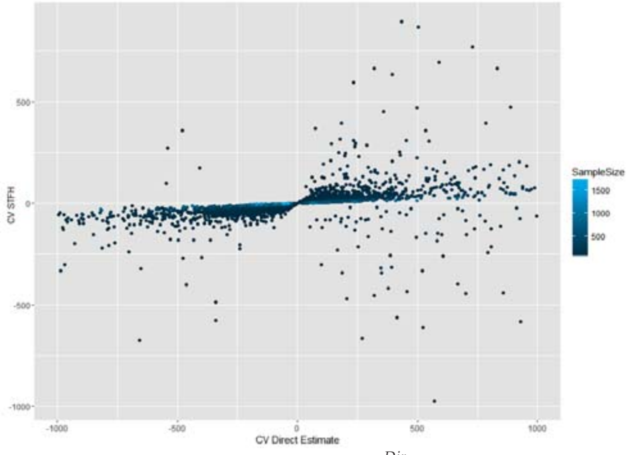
Appendix B Figures

FIGURE 5
NUMBER OF OFFERED APARTMENTS ON POST CODE LEVEL 2015



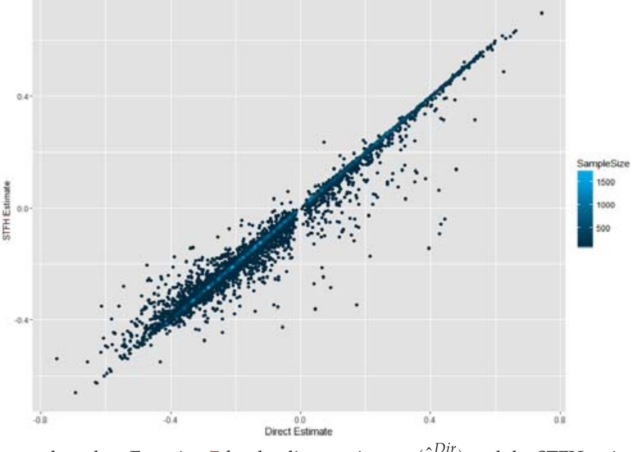
NOTES.—Number of observations means the number of offered apartments in 2015 summed over each post code. Using data for 2014 leads to a similar picture.
SOURCE.—Authors' calculations based on RWI-GEO-RED 2015.

FIGURE 6
COMPARISON OF THE COEFFICIENTS OF VARIANCE (CV)



NOTES.—Coefficients of variance for the direct estimates ($\hat{\bar{v}}_j^{Dir}$) and the STFH estimates. The coefficient of variance is the ratio of the standard deviation to the mean. The figure presents the results for 2015.
SOURCE.—Authors' calculations based on RWI-GEO-RED and RWI-GEO-GRID.

FIGURE 7
COMPARISON OF THE DIRECT ESTIMATES AND STFH ESTIMATES



NOTES.—Estimates based on Equation 7 for the direct estimates ($\hat{\bar{v}}_j^{Dir}$) and the STFH estimates.
SOURCE.—Authors' calculations based on RWI-GEO-RED and RWI-GEO-GRID.