

# Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria

*Marco Helbich<sup>1</sup>*

*Wolfgang Brunauer<sup>2</sup>*

*Eric Vaz<sup>3</sup>*

*Peter Nijkamp<sup>4</sup>*

<sup>1</sup> University of Heidelberg, Germany;

<sup>2</sup> UniCredit Bank Austria AG, Austria;

<sup>3</sup> Ryerson University, Canada;

<sup>4</sup> Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria

Marco Helbich, Wolfgang Brunauer, Eric Vaz, Peter Nijkamp

## Abstract

Modeling spatial heterogeneity (SH) is a controversial subject in real estate economics. Single-family-home prices in Austria are explored to investigate the capability of global and locally weighted hedonic models. Even if regional indicators are not fully capable to model SH and technical amendments are required to account for unmodeled SH, the results emphasize their importance to achieve a well-specified model. Due to SH beyond the level of regional indicators, locally weighted regressions are proposed. Mixed geographically weighted regression (MGWR) prevents the limitations of fixed effects by exploring spatially stationary and non-stationary price effects. Besides reducing prediction errors, it is concluded that global model misspecifications arise from improper selected fixed effects. Reported findings provide evidence that SH of implicit prices is more complex than can be modeled by regional indicators or purely local models. The existence of both stationary and non-stationary effects imply that the Austrian housing market is economically connected.

## 1 Introduction

In real estate research, it is well established that hedonic prices may vary across space such as stratifications of metropolitan areas, regions, and counties (e.g., Bourassa et al., 1999; Goodman and Thibodeau, 2003; Bischoff and Maennig, 2011; Helbich et al., 2013a). However, this parametric modeling approach has some restrictions: Spatial units have to be defined exogenously, SH is modeled in a discrete fashion where continuous changes across space can be expected, and usually the same definition of spatial units is used for all spatially varying effects (e.g., Redfearn, 2009; McMillen and Redfearn, 2010).

By preventing these limitations, non-parametric locally weighted regression procedures (LWR; Cleveland and Devlin, 1988) offer significant advantages (McMillen and Redfearn, 2010; McMillen, 2010). In this context, the slowly growing body of literature on LWR primarily use the geographically weighted regression (GWR; Fotheringham et al., 2002), which inherently assumes SH in the hedonic price function and *all* involved predictors (e.g., Bitter et al., 2007; Yu et al., 2007; Hanink et al., 2010). However, in situations where only some of the variables vary spatially, GWR results in inefficient estimations and possibly incorrect conclusions (Wei and Qi; 2012). This is particularly true for relatively small countries like Austria where real estate markets are economically connected through common federal policies such as governmental subsidies etc. Economic ties might also be relevant for structural housing features where Austrian-wide equilibrium conditions between supply and demand are a rational outcome. On the other side, spatially varying implicit prices are expected where local legislation and regulation (e.g., through spatial planning policies) are effective and/or where local supply consequence scarcity (e.g., plot area). Consequently, some price determining effects are expected to vary across space, while others are spatially homogenous. Both aspects are modeled simultaneously in the so-called MGWR (Fotheringham et al., 2002). Although the MGWR model seems to be rational, it has not yet been considered in real estate studies.

Therefore, the overall objective of this research is to explore SH in cross-sectional housing price functions. Using Austrian housing data, following contributions to the literature are established: First, the efficiency of fixed effects is evaluated in the context of global models, considering additionally iteratively technical corrections for spatial autocorrelation (SAC) and SH. Second, switching to fully local non-parametric modeling, it is demonstrated that SH is only inadequately captured by imposing spatially fixed effects and that systematic parameter variation is evident which deviates substantially from globally estimated marginal prices. Finally, due to an absence of empirical consensus which predictor enters the model globally or locally, the data-driven MGWR procedure is proposed. As such, following research questions are addressed:

- Are standard hedonic regressions equipped with spatial indicators suitable to model SH? Are technical corrections required to archive a well-specified model?
- Is SH beyond the regional fixed effects present? If this is the case, does a semi-local model outperform its global and fully local counterparts?

The rest of the paper is structured as follows. Section 2 reviews the theoretical foundations of hedonic pricing theory and discusses methodologies to account for SH. Following this, Section 3 introduces the empirical models. Section 4 presents the study area and the dataset. The results are summarized in section 5, before Section 6 highlights major conclusions and implications.

## 2 Background

### 2.1 Hedonic pricing theory

The theoretical basis of hedonic price modeling (Rosen 1974) is derived from Lancaster's (1966) consumer behavior theory, which argues that not the good itself creates utility, but its individual characteristics. As housing characteristics are non-separable and traded in bundles, real estate is usually treated as a heterogeneous good. Houses are valued for their utility-bearing characteristics with implicit prices, which can be considered as the component's specific prices (McDonald, 1997). Thus, a household implicitly chooses a set of different goods and services by selecting a specific object (Sheppard, 1997; Malpezzi, 2003). In the course of their purchase decisions, households aim to maximize their utility depending on their own social and economic characteristics. Households' utilities are also increasingly influenced by the absolute location of the house. Urban economic theory, in particular the monocentric Alonso-Muth-Mills model, provide a uniform framework to explain the spatial organization of housing (Anas et al. 1998). Briefly stated, the model claims that distance to the core city is the exclusively determinant causing spatial variation in housing prices and that prices, among other things, smoothly tend to decline with distance. Due to this reductionist use of commuting costs, the model has attracted some criticism. However, in a recent empirical study of real estate commodities, Ahlfeldt (2011) concludes that the model still performs satisfactorily.

Methodologically, a hedonic price function  $f$  describes the functional relationship between the real estate price  $P$  and associated physical characteristics  $x_1^p, \dots, x_n^p$  as well as neighborhood characteristics  $x_1^n, \dots, x_m^n$ . The former depicts the fabric of the dwelling (e.g., floor area); the latter defines the dwelling's surroundings, often based on census data (e.g., educational level; see Can, 1998). Usually the equation is estimated by means of multiple regression analysis. In practice, most empirical research uses a semi-log or log-log specification, having the advantage that prices vary with the quantity of housing characteristics and further corrects for heteroskedastic tendencies (Malpezzi

2003). Can and Megbolugbe (1997) assess that a wrong functional form results in unreliable and biased estimates.

## 2.2 Spatial effects in hedonic models

SAC and SH are two challenges in hedonic modeling. Since Dubin (1992; 1998) and a wealth of subsequent research (e.g., Can and Megbolugbe, 1997; Pace et al., 1998; LeSage and Pace, 2009; McMillen, 2010), it is now accepted that such spatial effects should be taken into account when estimating hedonic price functions. SAC describes the coincidence of locational and attribute similarity (Anselin 1988) caused by analogous neighborhood characteristics, similar socio-economic characteristics of their residents, and the quality of services (Dubin, 1992). As a consequence, the traditional ordinary least squares (OLS) estimator is inefficient and statistical inference is invalid (Dubin, 1998).

Although the price function assumes spatial equilibrium between supply and demand for dwelling characteristics, as real estate is fixed in space, and because of its durability and consistent properties, supply becomes largely inelastic (Malpezzi, 2003). Furthermore, different socio-economic and demographic conditions of households cause spatial variation in dwelling demand (Sheppard, 1997). Functional disequilibria, which manifest themselves in the emergence of heterogeneous market structures, might result as a consequence (McMillen and Redfearn, 2010). Therefore, SH should be considered in hedonic models. As LeSage and Pace (2009) pointed out, non-modeled heterogeneity can lead to biased results and falsely induced SAC.

## 2.3 Modeling spatial heterogeneity

SH can either be considered in a *discrete* way or in a *continuous* manner. Both approaches are discussed in this Section.

### 2.3.1 Discrete approaches

Discrete attempts model SH based on predefined spatial units (e.g., federal states) which are usually considered within the regression as fixed or random effects.

#### ***Fixed effects model***

Within the regression framework, fixed effects or spatial indicators for regions can be integrated, which let the intercepts vary over space. Slope heterogeneity can be controlled through spatial interaction effects of the spatial indicators with explanatory covariates (e.g., Kestens et al., 2004). However, they implicitly assume prior knowledge about the actual spatial process. Modeling heterogeneity using a large number of fixed effects can result in insufficient observations within regions for parameter estimations which, due to the loss of degrees of freedom, decrease the prediction accuracy. Therefore, it is common practice to interact in an ad-hoc fashion where only "one-variable-at-a-time" is considered (McMillen and Redfearn, 2010, p. 713). Thus, a trade-off between both data fidelity and reduction of prediction accuracy is required which is partially solved by the *random effects model* (Orford, 2000; Goldstein, 2011).

### **Random effects model**

The simplest case of a random effects model is the one-way intercept model, whereby intercepts vary spatially. Random effects can be approximated as a weighted average of the mean of the observations in the spatial units (corresponding to a dummy specification) and the overall mean of the whole country (Gelman and Hill, 2007). The weights are determined by the amount of information in a given region. For small subsample sizes with little information, estimates tend to equal the global mean, while for large subsamples estimates tend towards the "unpooled" dummy estimate. One advantage of this modeling strategy is that the closer the estimates are to the "pooled" model, the less effective degrees of freedom are used. However, if random effects are correlated with the predictors, the regression yields biased and inconsistent results. Similar to fixed effects, random effects models can be generalized by two-way structures (e.g., random time effects) and by letting predictors interact with the random effects, allowing different intercepts and/or slopes within each region (Goldstein, 2011).

### **Multilevel regression**

As stated in Orford (2000), considering real estate as nested within several levels of spatial units, turns the hedonic pricing model into a *multilevel* regression problem (e.g., Goldstein, 2011). Multilevel models generalize random effects models to more than one hierarchical level, in which coefficients are determined by a probability model (Gelman and Hill, 2007). This second-level model has parameters of its own (*hyperparameters*), which are also estimated from the data of the first-level model. However, in fixed and random effects, as well as in multilevel models, spatial units do not necessarily coincide with the true spatial (possibly non-stationary) data generating process (Dubin, 1998). Thus, one faces the modifiable areal unit problem (MAUP; see Fischer and Wang, 2011). In order to incorporate more homogeneous region delineations, statistical methods, such as cluster analysis (e.g., Bourassa et al., 1999) or regionalization (e.g., Helbich et al., 2013a), have been utilized. While Orford (2000) emphasizes the usefulness of multilevel modeling, Fotheringham et al. (2002) sharply criticize the discrete nature in which space is implemented, which further implies that the price function is discrete and homogeneous within an spatial unit. Moreover, abrupt and artificial discontinuities along the spatial units are present, consequently this suggests continuous modeling of SH.

## **2.3.2 Continuous approaches**

These techniques do not rely on exogenous assumptions concerning spatial units, and therefore provide a data-driven modeling of parameter instability.

### **Polynomial regression and spatial expansion model**

Dubin (1992) discusses the use of polynomial regression to investigate continuous large-scale variation. The expansion model developed by Cassetti (1972) relates to polynomial regression. Parameters vary as a function of the coordinates and are allowed to drift spatially. Comparing non-stationarity modeling approaches, Bitter et al. (2007) conclude that the expansion model is suitable to analyze previous known broad trends but fails to investigate complex unknown local spatial relationships. Similar, Farber and Yeates (2006) state that complex patterns are poorly depicted by polynomial functions. Specifically, following Dubin (1992), this approach does not seem useful for hedonic modeling; because (a) polynomials tend to get distorted at the edges of the study area, (b) higher order polynomial terms are highly multicollinear, (c) results are uncoupled from the spatial distribution of the entities, and (d) are too smooth for modeling local variation in prices.

### **Local weighted regression**

Compared to the mentioned models, non-parametric LWR (Cleveland and Devlin, 1988) offer substantial advantages, highly adaptable for spatial housing data (McMillen and Redferan, 2010). Embedded in the LWR framework, the conditional parametric model, also termed GWR (Fotheringham et al., 2002), is an explicitly local model and circumvents problems discussed in the context of discrete modeling of heterogeneity and polynomial regression. GWR implicitly assumes continuously changing price functions and models SAC and SH (McMillen and Redferan, 2010). Currently, the suitability of GWR is extensively discussed: Wheeler and Tiefelsdorf (2005) note coefficient sign reversals. Jetz (2005 cited in Páez et al., 2011) hypothesizes that it might be induced by artificially localizing the model, causing a local omitting variable bias. Moreover, Wheeler and Tiefelsdorf (2005) have found correlations among estimated GWR parameters which may affect interpretations. Wheeler (2007) argues that local collinearity in the covariates are the reason, which inflates variances and in turn yield reverse parameter signs. Based on Monte-Carlo simulations, Farber and Páez (2007) examine the GWR bandwidth calibration. They conclude that globally fixed bandwidths do not appropriately reflect spatial processes, resulting in volatile regression coefficients. Extending Farber and Yeates (2006), who detected extreme coefficients which might emerge from local distributional irregularities, Cho et al. (2009) provide evidence that fixed bandwidths and less spatially dense data are potentially more vulnerable to extreme coefficients. Also McMillen and Redfearn (2010) favor a varying bandwidth. Recently, Páez et al. (2011) mitigate the doubt not to use GWR for inference. Criticizing previous studies (e.g., Wheeler and Tiefelsdorf, 2005) due to their limited sample sizes of less than 400 observations and their restricted research designs, they provide simulation-based evidence that spurious correlations are noticeably reduced by samples of more than 1,000 observations. A strong advantage of GWR is its flexibility (Farber and Yeates, 2006; McMillen 2010) and that the price function needs no prior assumption concerning the price determination process and its spatial variation (e.g., Yu et al., 2007). This corresponds to the suggestion of Ugarte et al. (2004), promoting more data-driven techniques. However, for some of the covariates the impact on the price can be global, while for others it may vary locally. Such constrains on the parameters can be imposed by politics or economic theory (Wei and Qi, 2012). This fact is usually neglected in GWR specifications (e.g., Bitter et al., 2007; Yu et al., 2007; Hanink et al., 2010), resulting in inefficient estimates (Wei and Qi, 2012). Therefore, based on Robinson (1988), Fotheringham et al. (2002) propose an extension of the basic GWR, resulting in the statistically more parsimonious MGWR, which is the focus of this research.

## **3 Methods**

This section introduces (a) global econometric models and (b) local as well as semi-local models, relevant for the subsequent empirical analysis.

### **3.1 Global models**

#### **3.1.1 Non-spatial reference model**

Formally, the OLS model is defined as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

where  $\mathbf{y}$  is an  $n \times 1$  vector of transaction prices,  $\mathbf{X}$  is an  $n \times k$  matrix of predictors,  $\boldsymbol{\beta}$  a  $k$ -vector of estimated parameters, and  $\boldsymbol{\varepsilon}$  an  $n \times 1$  error term with  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ . SH is modeled by using fixed effects, which reduces the extent of the prediction error, removes most of the systematic error, and leads to more reliable estimates (Can and Megbolugbe, 1997; Pace et al., 1998). However, as McMillen (2003) points out, unexplained similarities between neighboring houses can result in SAC, which is captured using spatial autoregressive models.

### 3.1.2 Spatial autoregressive model

The family of spatial autoregressive models, popularized by Anselin (1988), comprises two special cases, namely the lag and error model. The former extends the standard regression model by including a spatially lagged dependent variable (further on called SAR model), while the latter addresses SAC by defining a spatial autoregressive error process (Anselin, 1988). If only residual SAC is present, OLS estimates are inefficient, although unbiased and consistent. However, if a spatial lag process exists, OLS produces biased and inconsistent estimates. The Lagrange multiplier statistic (LM) on the estimated OLS residuals is applied to decide between these two alternatives.

The SAR, which is employed in this study, is written as:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

where  $\mathbf{y}$  is an  $n \times 1$  vector of prices,  $\mathbf{X}$  an  $n \times k$  predictor matrix,  $\boldsymbol{\beta}$  a  $k$ -vector of estimated coefficients,  $\mathbf{W}$  an  $n \times n$  weight matrix,  $\rho$  a spatial lag of the dependent variable  $\mathbf{y}$ , and  $\boldsymbol{\varepsilon}$  the *iid* error term. Equation (2) shows that this model adjusts for SAC in the dependent variable, interpretable as spillover effects. Bivand et al. (2008) discuss several options for the specification of  $\mathbf{W}$ , including  $k$ -nearest neighbors, neighborhood contiguity by distance, and graph-based approaches. However, caution is necessary, as an incorrect specification can itself result in residual SAC (Florax and Rey, 1995). A comprehensive treatment is provided by LeSage and Pace (2009).

### 3.1.3 Spatial two stage least squares with spatial heteroskedasticity and autocorrelation consistent standard errors

Homoskedasticity, as stipulated in both the OLS and the SAR model, is a strong and restrictive assumption (Kelejian and Prucha, 2007). Recently, however, Piras (2010) provided functions for robust inference in the presence of spatial heteroskedasticity and autocorrelation (SHAC) across spatial units estimated by a spatial two stage least square procedure (S2SLS):

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} = \mathbf{R} \boldsymbol{\zeta} \quad (3)$$

where  $\boldsymbol{\zeta}$  is a vector of disturbances and  $\mathbf{R}$  is an  $n \times n$  non-stochastic matrix with unknown elements. The other symbols correspond to equation (2). For details it is referred to Piras (2010).

Clearly, global models have the drawback that they assume homogeneous behavior of the parameters over space, which may be a misspecification of reality and consequently may be locally biased. Therefore, locally weighted regressions provide significant advantages (McMillen, 2010).

## 3.2 Local and semi-local models

### 3.2.1 Geographically weighted regression

GWR performs a series of weighted least squares regressions on subsets of the data, where the influence of an observation  $i$  decreases with the Euclidean distance to a regression point  $j$ . These distance-dependent weights are determined by a kernel function, including the bi-square, Gaussian, or tricube function. More important than the kernel type for model estimation is the range of the input data (bandwidth) to which data points are considered for each LWR (Fotheringham, et al. 2002). Besides assuming a predefined and fixed bandwidth, an adaptive bandwidth has been proven to be highly suitable in practice (McMillen and Redfearn, 2010), reflecting a relation between the density of the regression points and the bandwidth (i.e., denser points yield smaller bandwidths and vice versa). Often adaptive bandwidth selection is determined by cross-validation (CV; Farber and Páez, 2007). This reduces large estimation variances in sparsely sampled areas. To achieve parameter surfaces of the coefficients on non-observed locations, geostatistical algorithms (Pebesma, 2004) can be employed (e.g., Yu et al., 2007).

### 3.2.2 Mixed geographically weighted regression

MGWR (Fotheringham et al., 2002) is a statistically more parsimonious version of the GWR, where some coefficients (usually those with non-significant variation over space) are kept constant, saving degrees of freedom and improving efficiency of estimators of coefficients (Wei and Qi, 2012). Such a model is written in the form as follows:

$$y_i = \sum_{j=1}^k \beta_j x_{ij} + \sum_{l=k+1}^m \beta_l(u_i, v_i) x_{il} + \varepsilon_i \quad (4)$$

where  $y_i$  is the logarithmically transformed sales price of observation  $i$ ,  $i \in 1, \dots, n$ ,  $\beta_j$  are the global coefficients of covariates  $x_{ij}$ ,  $j \in 1, \dots, k$ , and are the  $\beta_l(u_i, v_i)$ ,  $l \in k+1, \dots, m$ , the local coefficients of covariates  $x_{il}$  at the coordinates  $(u_i, v_i)$  of observation  $i$ . Leung et al. (2000a) propose a statistic to distinguish spatial stationary and non-stationary covariates.

The MGWR calibration procedure can be achieved by a multi-step algorithm (Fotheringham et al., 2002), a two-step procedure (Mei et al., 2006), and a constrained type (Wei and Qi, 2012). Due to easier implementation, the multi-step algorithm is applied:

1. Each global covariate  $x_1, \dots, x_k$  and the response variable  $\mathbf{y}$  are regressed on the local covariates  $x_{k+1}, \dots, x_m$  using the basic GWR algorithm and the resulting residuals  $\tilde{x}_1, \dots, \tilde{x}_k$  and  $\tilde{\mathbf{y}}$  are determined.
2. The residuals of the response are regressed on the residuals of the global covariates using OLS:  $\tilde{y}_i = \sum_{j=1}^k x_{ij} \beta_j + \varepsilon_i^{global}$ , which yields the estimated global coefficients  $\hat{\beta}_1, \dots, \hat{\beta}_k$ .
3. The estimated global effects are then subtracted from the original response variable, resulting in a partial residual of the response:  $\tilde{y}_i = y_i - \sum_{j=1}^k x_{ij} \hat{\beta}_j$ .
4. Finally, these adapted prices are regressed on the spatially varying covariates by using the basic GWR method, resulting in the varying GWR coefficients:  $\tilde{y}_i = \sum_{l=k+1}^m \beta_l(u_i, v_i) x_{il} + \varepsilon_i^{local}$ .

## 4 Study area and data

### 4.1 Housing data

Housing data for Austria are provided by the UniCredit Bank Austria AG. The dataset consists of 3,887 geocoded single-family-homes. Figure 1 shows their spatial distribution. Clearly, observations are non-uniformly distributed and show a concentration in the eastern federal states. Individual (logged) transaction prices of house purchases recorded in Euro have been collected from 1998-2009 together with physical properties of the houses in order to estimate the value of the collateral for mortgages.

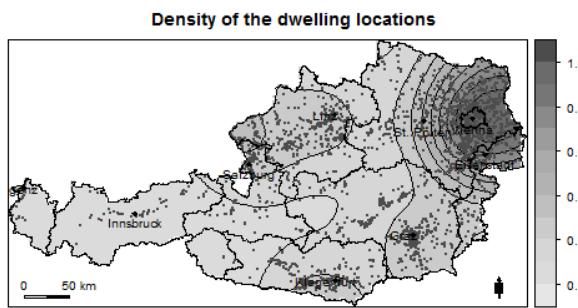


Fig. 1: Spatial distribution of houses depicted as points and as normalized kernel density estimation. The larger black point symbols represent the nine Austrian federal state capitals within each federal state (black boundary lines).

Official census data published by Statistics Austria and Michael Bauer Research GmbH for 2001 and 2009, respectively, describe the socio-economic characteristics of each municipality and enumeration district (thereafter called neighborhood covariates). Table 1 provides an overview of the data.

### 4.2 Covariates and their expected effects

#### 4.2.1 Physical covariates

Nine physical covariates, measuring the size of the house and describing the quality of the house, are included. A positive effect of the floor area and the size of the plot the house is built upon is expected. In line with Malpezzi (2003), both covariates are logarithmically transformed to account for the multiplicative structure. Furthermore, the market value of a house is dependent on its structural condition and architecture; that is, the efficiency of heating, the presence of a garage and basement as a positive asset, as opposed to a traditional top-floor attic which might reduce the price *ceteris paribus* due to the limitation of the amount of usable area.

#### 4.2.2 Temporal covariates

The age of the building at a given time of sale reflects property depreciation over time and should decrease house prices, notwithstanding a vintage effect, having an opposite effect (Can, 1998). Stevenson (2004) proposes an additional quadratic age term to capture non-linearities (see also Helbich et al., 2013b). The year of the house purchase can be considered as the remaining

unexplained temporal heterogeneity and is a measure for the quality adjusted development of prices over time.

Tab. 1: Description of the variables.

Variable	Description
Inp	Log purchase price
lnarea_total	Log of total floor area (m <sup>2</sup> ; except basement)
lnarea_plot	Log of plot space (m <sup>2</sup> )
cond_house	Condition of the house (0=good; 1=poor)
heat	Quality of the heating system (0=high; 1=poor)
Bath	Quality of the bathroom/toilet (0=high; 1=poor)
Attic	Existence of an attic (0=no; 1=yes)
Basement	Existence of a basement (0=no; 1=yes)
Garage	Quality of the garage (0=high; 1=poor)
Terr	Existence of a terrace (0=no; 1=yes)
Age	Age of building at time of sale (year)
Time	Year of purchase 1998-2009 (date)
muni_pp_ind	Purchase power (index, federal state = 100, 2009)
muni_acad	Proportion of academics (%), 2001
muni_unempl	Unemployment rate (%), 2009
muni_age_ind	Age (index calc. from 5-years age cohorts, 2001)
muni_comm_ind	Commuter (index, 2001)
ln_muni_popd	Log population density (2001)
muni_liv_ind	Quality of living (index, 2001)
enumd_pp_ind_dmuni	Purchase power index (dev. from municip. mean, 2009)
enumd_acad_dmuni	Proportion of academics (dev. from municip. mean, 2001)
enumd_unempl_dmuni	Unemployment rate (dev. from municip. mean, 2009)
enumd_age_ind_dmuni	Age index (dev. from municip. mean, 2001)
ln_enumd_popd_dmuni	Log population density (dev. from municip. mean, 2001)
fed_states	9 Federal states

#### 4.2.3 Neighborhood covariates

Both the purchase power index and the proportion of academics reflect disposable income and the economic status. These covariates should therefore affect house prices positively. In contrast, the population age index reflects excess of age and serves as a proxy for structural weakness. This index should have a negative effect on house prices. Also, the unemployment rate reflects economically weakness and should have a negative impact. Furthermore, urban economic theory (e.g., McDonald, 1997) states that shorter commuting distances to centers of economic activity should raise property prices, which is why a high commuter index should tend to affect prices positively. Although close proximity to these centers can provide certain amenities (e.g., functional services) there might occur disamenities, such as environmental pollution. Therefore, the effect of commuting distances is unclear. Logged population density measures urbanity. As land becomes more valuable in densely populated areas, a positive effect on house prices is expected. Finally, the quality of living within a municipality is quantified by a weighted average of the dwelling category according to the Austrian rental law. It is assumed that in areas with largely high-quality dwellings prices are higher.

#### **4.2.4 Compositional effects**

Neighborhood effects can be used at the level of the enumeration district or the municipality. In the latter case, compositional effects (Goldstein, 2011) allow neighborhood variables to be integrated at both hierarchical levels in which they have been measured. The difference between a hierarchy level (e.g., enumeration district) from the average of the higher level (e.g., municipality level) is considered. The resulting effect is then interpretable as a level-specific differential effect of the higher level.

#### **4.2.5 Regional indicators**

Each house is nested within one of the nine Austrian federal state. However, due to a sparse sample in the western federal states of Vorarlberg and Tyrol both regions are considered as one spatial unit in the global models. Such ad-hoc stratifications are commonly used to model small-scale spatial heterogeneity (e.g., Bischoff and Maennig, 2011). This kind of unexplained spatial heterogeneity is accounted for by using a regional indicator specification.

### **5 Results**

#### **5.1 Descriptive and exploratory analysis**

The mean transaction price of a house is approximately 167,200 Euro, with a minimum of 30,000 Euro (in Haugschlag next to Czech Republic border), and a maximum of 550,000 Euro in Mödling, in the south of Vienna. Initially, for exploratory spatial analysis the weight matrix, reflecting the houses' spatial configuration, is defined. Distance-based weights are rejected, as a distance of 21.6 km would have been necessary to avoid houses without any neighbors, which seems unrealistic. Instead, a  $k$ -nearest neighbor definition with row standardization is used, accounting for heterogeneously distributed dwellings (Bivand et al., 2008). To be consistent with the global models below, a commonly used value of 5 is chosen (e.g., Wilhelmsson, 2002). The impression of spatially autocorrelated dwelling prices is supported by a significant Moran's  $I$  statistic ( $I = 0.286, p < 0.001$ ). Due to a sharing of similar socio-economic characteristics of the residents and quality of services, Dubin (1998) claims that such patterns are common in real estate analysis and may affect subsequent regression analysis.

#### **5.2 Global models**

This section reports the estimation results of the OLS, SAR, and S2SLS models. To realize a more parsimonious OLS model, a stepwise variable selection procedure minimizing the Akaike Information Criterion (AIC) is applied. With a highest value around two, the Generalized Variance Inflation Factors gives no indication of multicollinearity. With an adjusted  $R^2$ , the OLS model (Table 1) explains around 47% of the variation in logged prices. Excluding the regional dummies lowers the adjusted  $R^2$  and increases the AIC about 204. Further model inspection shows inconsistency with the Gauss-Markov assumptions. The Breusch-Pagan test indicates heteroskedasticity ( $BP = 166.519, p < 0.001$ ). This is a recognized problem in econometrics and, hence, White's heteroskedasticity consistent coefficient covariance matrix is calculated (Kleiber and Zeileis, 2008). However, the adjusted standard errors show only marginal differences. Furthermore, these results are based on the incorrect assumption of spatially uncorrelated model residuals ( $I = 0.080, p < 0.001$ ), leading to inefficient parameter

estimations, with the resulting confidence intervals being incorrect. This misspecification is a result of existing spatial effects and must be taken into account.

Autoregressive models adjust for SAC. On the basis of the robust LM test ( $LM = 28.506, p < 0.001$ ) it is concluded that a re-specification in terms of a SAR model is the proper alternative, which is in line with Yu et al. (2007). Thus, the OLS model results are not only inefficient, but also inconsistent, and an additional spatial lag based on a  $k = 5$  nearest neighbors is considered. Tests show that the parameter estimations are robust against changes in the weight matrix (e.g., different  $k$ -values, distance decay functions) and lead to marginal changes in the coefficients. Again, variables are selected based on the full model. Predictors not significant on a 0.05 level are omitted stepwise. If the main effect shows no significance the corresponding compositional effect is also dropped. Overall, the SAR model (Table 1) performs significantly better than the OLS model, confirmed by the AIC reduction. A comparison of the model predictive quality, applying the root mean square error (RMSE), also supports the SAR model. Dropping the federal state dummy increases both the RMSE and AIC.  $\rho$  is highly significant. Thus, the model captures SAC effectively and points to relevant spillover effects. Consequently, the error term is not plagued by SAC. However, the BP test rejects the homoscedasticity assumption ( $BP = 171.308, p < 0.001$ ). Hence, heteroskedasticity and autocorrelation consistent standard errors, applying the S2SLS model, are required. As above, non-significant predictors are removed in a stepwise manner, starting with the full model.

Both the SAR and S2SLS coefficients of the neighborhood covariates are smaller compared with the OLS model, showing the bias induced by the neglection of SAC in the OLS model. A comparison between SAR and S2SLS with a triangular kernel and  $k = 5$  nearest neighbors indicates that the coefficients remain virtually unchanged. The S2SLS results in slightly smaller standard errors compared with the SAR model. The final S2SLS consists of 17 significant covariates. Both covariates quantifying the size of a house are positively related to the house price. A 1% increase of total floor area results in a 0.4% increase of the price, and the plot space shows a positive elasticity of 0.1%. A poor overall house condition leads to a loss in value of approximately 4%, whereas a poor heating system reduces house prices by more than 10% compared with a high quality heating system. Furthermore, a poor quality bath leads to a price reduction of 7% and a traditional attic reduces dwelling prices by up to 2%. A poor quality garage reduces prices by 8%. Other properties, such as the existence of a basement (+13%) and terrace (+7%), have the expected positive effects on housing prices. Both temporal effects are statistically significant. The house's age has a marginally negative impact on the price. Contradicting original expectations, the quadratic age term is insignificant. The year of purchase, modeled as linear effect, suggests that there has been a price increase over time.

However, compositional effects are only used, if both levels are significant. In the S2SLS model, the main and the differential effects of the proportion of academics show significance. A one percentage point increase in the proportion of academics in a municipality results in a 0.8% increase in house prices. The corresponding differential effect shows that within a municipality, a one percent point increase of this proportion results in a further 0.5% increase in house price value. Also an increase in the purchase power index has a positive effect (+0.5%). Furthermore, an increase in the average population age of one year, reflecting structural weaknesses, reduces expected house prices by some 2%. Densely populated areas show a marginally positive elasticity. Although the compositional effect for the rate of unemployment is significant, it is removed because its main corresponding effect is not significant. Finally, with the exception of Upper Austria, the federal state dummies are highly significant and indicate SH. For instance, properties located in the structurally weaker state of

Burgenland have a significantly lower price compared with houses situated in Lower Austria, serving as reference category.

Tab. 2: Estimation results of the global models.

	<u>OLS</u>		<u>SAR</u>		<u>S2SLS</u>		SHAC Std. err.	<i>t</i> -val.
	Coeff.	HC Std. err.	<i>t</i> -val.	Coeff.	Std. err.	<i>z</i> -val.	Coeff.	
Intercept	9.391	0.195	48.278***	6.872	0.328	20.969***	6.542	0.548
Inarea_total	0.410	0.017	23.988***	0.409	0.018	22.825***	0.411	0.017
Inarea_plot	0.111	0.012	8.945***	0.111	0.013	8.855***	0.111	0.012
cond_house_1	0.052	0.014	-3.742***	0.045	0.014	-3.330**	0.044	0.014
heat_1	0.102	0.020	-5.018***	0.107	0.023	-4.602***	0.105	0.021
bath_1	0.076	0.019	-3.910***	0.074	0.022	-3.339***	0.075	0.019
attic_1	0.024	0.012	-2.018*	0.024	0.011	-2.077*	0.024	0.012
basement_1	0.132	0.013	9.936***	0.126	0.013	9.597***	0.124	0.014
garage_1	0.082	0.012	-6.899***	0.084	0.012	-7.162***	0.084	0.012
terr_1	0.070	0.012	5.893***	0.066	0.012	5.673***	0.066	0.012
age	0.006	0.000	16.976***	0.006	0.000	17.617***	0.006	0.000
time	0.009	0.003	3.550***	0.010	0.002	4.100***	0.010	0.003
muni_pp_ind	0.005	0.001	5.800***	0.005	0.001	5.890***	0.005	0.001
muni_acad	0.011	0.002	5.695***	0.009	0.002	4.384***	0.008	0.002
muni_unempl	1.454	0.649	-2.241*					
muni_age_ind	0.020	0.004	-5.538***	0.017	0.004	-4.788***	0.016	0.004
ln_muni_popd	0.033	0.007	4.518***	0.020	0.006	3.153**	0.020	0.006
enumd_acad_dmuni	0.008	0.001	6.072***	0.006	0.001	4.997***	0.005	0.001
enumd_unempl_dmuni	1.758	0.607	2.898**					
enumd_age_ind_dmu ni	0.004	0.003	1.568***					
Burgenland	0.129	0.023	-5.644***	0.096	0.025	-3.905***	0.087	0.024
Carinthia	0.082	0.024	-3.397***	0.093	0.024	-3.912***	0.092	0.024
Upper Austria	0.018	0.021	0.870	0.016	0.019	0.832	0.012	0.019
Vorarlberg & Tyrol	0.241	0.042	5.791***	0.182	0.036	5.023***	0.179	0.041
Salzburg	0.231	0.030	7.748***	0.190	0.028	6.796***	0.185	0.031
Styria	0.055	0.020	-2.760**	0.056	0.020	-2.805**	0.055	0.020
Vienna	0.266	0.034	7.759***	0.220	0.031	7.011***	0.215	0.034
$\rho$				0.205		$p < 0.001$	0.229	$p < 0.001$
Adj. $R^2$	0.473			0.486				
AIC	2783			2707				
AIC without regions	2988			2850				
RMSE	0.344			0.339			0.339	
RMSE without regions	0.353			0.345			0.345	

Signif.: \*\*\*\* 0.001, \*\*\* 0.01, \*\* 0.05, \* 0.10

### 5.3 Local and semi-local model

Next, GWR and MGWR are estimated. A data-driven approach to explore functional instability replaces the federal states dummies. For both models the analysis is restricted exclusively to the main effects. Otherwise, compositional effects would result in complex and difficult interpretational local effects. Because of a significant compositional effect of the unemployment rate and an insignificant corresponding main effect in the S2SLS model, the unemployment rate is additionally considered at the enumeration district level. To achieve a trade-off between computational burden and estimation accuracy, (a) a random sample of 50% of the overall data set is used and (b) only those variables are chosen that are significant in the global S2SLS model (Table 1).

First, it is necessary to distinguish between global and local effects using non-parametric GWR with a Gaussian kernel. CV determines an optimal adaptive bandwidth that includes 227 neighboring houses. The test statistic after Leung et al. (2000a) does not reject the  $H_0$  of stationary parameters for six out of 16 variables. In a MGWR specification, those six variables are thus held spatially constant, while the remaining 10 variables are allowed to vary. Again, applying a Gaussian kernel function, CV results in an optimal bandwidth of 177 neighboring houses for the MGWR. Compared to the GWR the AIC score is noticeably reduced from 1,396 (GWR) to 1,381 (MGWR), while the RMSE for both models is quasi similar (0.338). Statistically the MGWR is favored. Sensitivity analysis with alternative kernels (i.e. bi-square kernel) shows no significant differences.

The spatial variation of the model fit ranges from 0.24 to 0.49. The lowest values are located in Vienna plus its surroundings and in the south of Austria. The best fit is achieved in the northern parts of Austria. The reason for the low fit in Vienna might be that flats dominate the market in the capital. Figure 2 visualizes the parameter surfaces of the local  $R^2$  and the predicted dwelling prices in Euro, achieved through ordinary kriging (Pebesma, 2004). The  $F(3)$ -test (Leung et al., 2000a) confirms that now all spatially varying covariates are statistically significant. No significant residual dependence on a 0.01 level is detected by the Moran's  $I$  (Leung et al., 2000b).

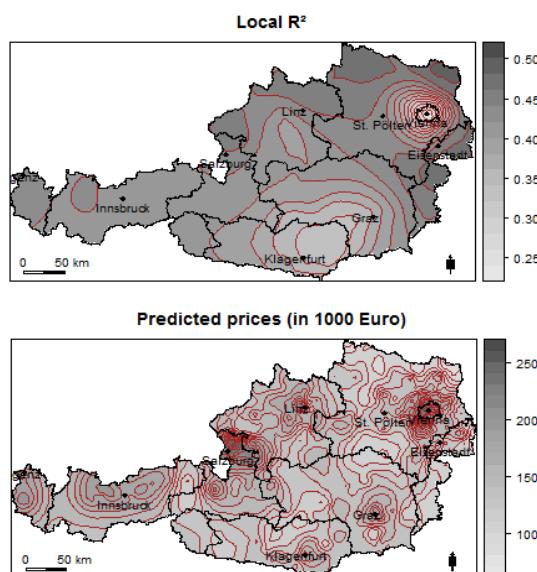


Fig.2: Local model performance and the predicted house prices.

Table 3 lists the MGWR results and its global counterparts. Additionally, Figures 3 through 5 visualize the parameter surfaces. Based on pseudo-*t*-values, only significant coefficients on  $\alpha = 0.1$  have gray tones.

Tab. 3: Estimation results of the semi-local model.

	Global coeff.			Local coeff.				F(3)-test
	Estim.	Std.err.	t-val.	1. quantile	Median	3. quantile		
Inarea_total	0.422	0.024	17.453***	0.367	0.398	0.458	***	
Inarea_plot	0.077	0.018	4.268***	0.061	0.087	0.111	***	
heat_1	-0.111	0.031	-3.544***	-0.163	-0.125	-0.072	*	
Age	-0.005	0.000	-11.441***	-0.006	-0.005	-0.004	***	
Time	0.006	0.003	1.990*	0.007	0.011	0.013	***	
enumd_unempl	-2.103	0.587	3.581***	-2.335	-1.520	-2.236	***	
muni_pp_ind	0.004	0.001	4.442***	-0.001	0.006	0.007	***	
muni_acad	0.012	0.003	4.567***	0.007	0.015	0.022	***	
muni_age_ind	-0.038	0.005	-8.260***	-0.037	-0.028	-0.022	***	
ln_muni_popd	0.067	0.008	8.623***	0.036	0.065	0.079	***	
cond_house_1	-0.038	0.020	-1.953'					
bath_1	-0.064	0.031	-2.029*					
attic_1	-0.028	0.016	-1.674'					
basement_1	0.144	0.019	7.665***					
garage_1	-0.068	0.017	-4.060***					
terr_1	0.076	0.017	4.587***					

Signif.: \*\*\* 0.001, \*\* 0.01, \* 0.05, . 0.10

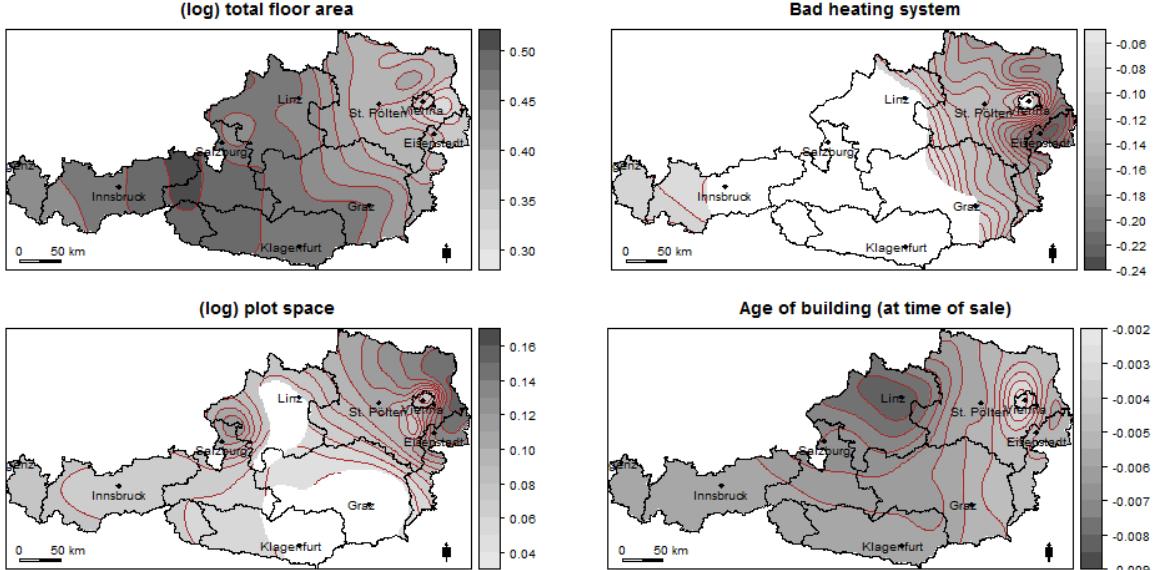


Fig. 3: Physical covariates and temporal effect.

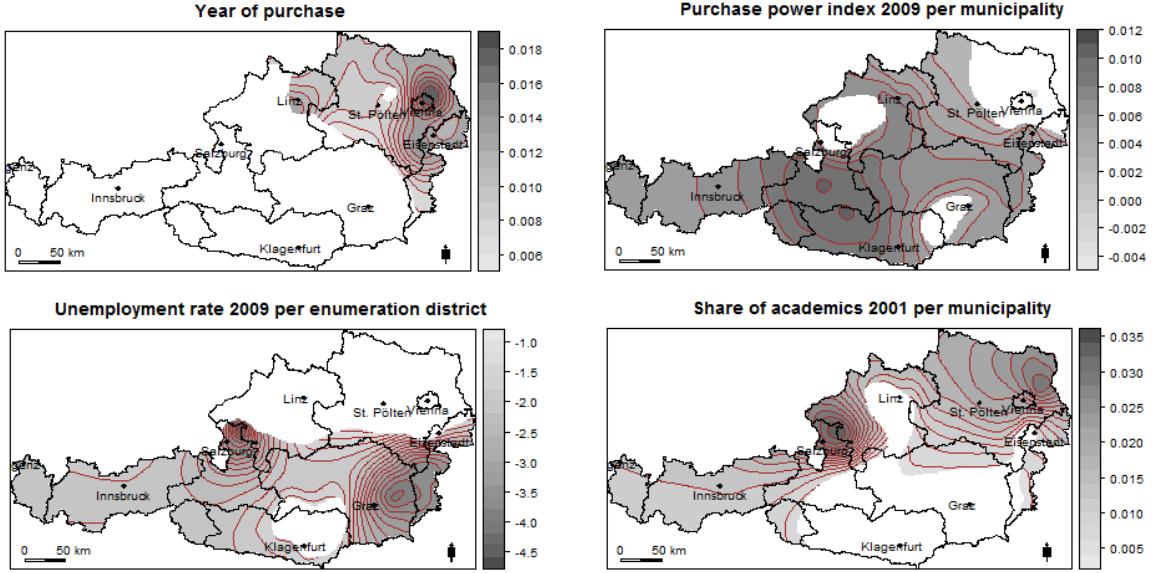


Fig. 4: Temporal effect and neighborhood covariates.

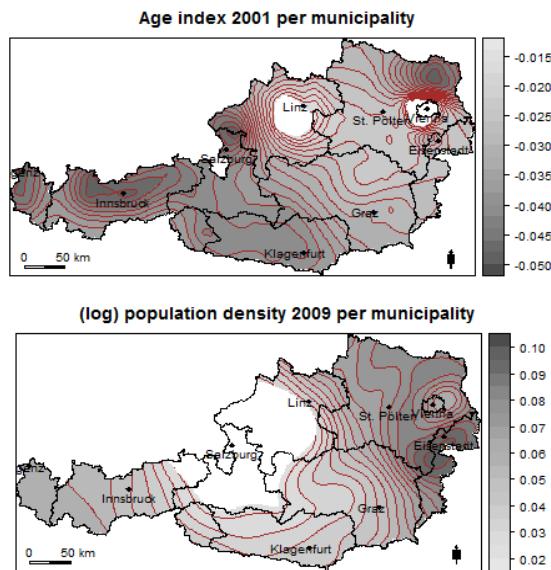


Fig. 5: Neighborhood covariates.

The negative MGWR intercept is not, by itself, meaningful. The spatially fixed coefficients consist solely of physical covariates, meaning buyers value these housing characteristics equal across space. Furthermore, these stationary coefficients illustrate that real estate markets are economically connected. Compared to the S2SLS estimates, the MGWR coefficients do not considerably deviate, which is an indication of model suitability.

The total floor area is statistically significant across the entire study site and shows strong spatial variation (Figure 3). Its positive effect on house prices is less pronounced closer to Vienna and reaches its largest positive effects in parts of the federal states of Salzburg and Tyrol. These areas include well-established winter sports sites (e.g., Kitzbühel) where tourism and vacation homes, secondary residences etc. increase house prices. Particularly, the scarcity of space leads to higher prices for additional floor area in these regions. In contrast, a different pattern can be found for

logged plot space. The marginal values are more distinct in the northeast of Vienna's suburbs (Figure 3). A 10% increase in plot space results in a price increase of 1.5% and prices decrease the more removed from Vienna. Additional plot area has little effect on house prices in southern and central Austria, meaning that in these regions, the proportion of land value in the total value of the house is relatively low. This leads to the conclusion that plot space is sufficiently available and lowers demand. Inferior heating systems have a negative effect (Figure 3), yet this effect is not significant in central Austria and has its most negative impact in the surroundings of Vienna (up to -21%). Hence, in these suburban areas the appreciation of a high technical standard is relatively high. It turns out that in these regions there is a large share of new constructions, so the spatial variation of this effect is likely to be due to an unobserved interaction effect between the age of the house and the technical standard of the heating system. Both temporal effects possess a significant non-stationary pattern (Figure 3 and 4). Most notably, the age effect is rather weak in Vienna (-0.2%), while in Linz and its surroundings the marginal effect is close to -1%. Lowest age effects are observed in central Upper Austria. In the Vienna region, there seems to be a tendency towards renovation of older buildings. This can also be interpreted with respect to the effect of an inferior heating system, as this effect is very pronounced where the effect of the age is weak. The year of purchase is most significant in the western federal states, whereas the northern surroundings of Vienna show the highest house price values increase over time (2% per year). This is a consequence of the on-going suburbanization, provoking increasingly high demand.

The unemployment rate<sup>1</sup> (Figure 4) is not significant for the northern regions of Austria but, as expected, are negatively related to house prices in the southern federal states. Unemployment yields the strongest negative effect close to the cities of Salzburg and Graz. The purchase power index (Figure 4) shows both negative and positive effects, but only the latter are statistically significant. Highest positive effects are achieved in the federal state of Salzburg (+1%). The proportion of academics, in Figure 4, shows a slight northwest to southeast trend, where maximum values can be found in the northern parts of Salzburg. In this area, an increase of the proportion of academics of a one percentage point results in a 3% increase of dwelling prices, holding all other effects constant. This compares to just 0.5% in the southwestern parts of Austria (e.g., Innsbruck). Beside this, Figure 5 illustrates a limitation of MGWR. For example, the area around the city of Salzburg shows a high impact of the proportion of academics on price. Against our expectation that high purchase power effects house price positively in this area, the coefficient is partly negative, but not significant at all. Local sign reversals are also reported in Wheeler and Tiefelsdorf (2005) and Wheeler (2007). Possibly, this is caused by local correlations between these covariates. The age index in Figure 5 shows a slight east-west divide. In the western federal states of Salzburg, Tyrol, and Vorarlberg an increase in average population age of one year results in a price reduction of approximately 5%. Outside of the cities of Linz and Vienna this covariate shows no significance at all. Finally, the logged covariate population density (Figure 5) has the highest elasticity in central Burgenland, while in the federal state of Salzburg and most parts of Upper Austria this covariate has no explanatory power.

Finally, by visualizing the predicted house prices in Figure 2, a distinct pattern of house prices becomes evident. Prices decline smoothly and relatively symmetrically with distance to the urban cores and peripheral areas yield unsurprisingly lower prices. These results confirm some agreement with monocentric urban models. Comparing the price gradients between federal state capitals, substantial slope differences of the price gradients are noticeable. Significantly higher prices are especially achieved in Vienna, Salzburg, and Innsbruck and their surroundings. Additionally, localized

---

<sup>1</sup> Note, the unemployment rate is scaled between 0 and 1, which leads to relatively high effects.

peaks with high house prices primarily in scenic and tourist-oriented areas are noticeable. In contrast, significant lower prices are predicted within the structurally and economically weaker Burgenland as well as the northern parts of the Weinviertel next to the Czech Republic and Slovakia.

## 6 Conclusions

Hedonic modeling is frequently used to access the price of real estate. Although SH in the price function is obvious, it is *a priori* unclear how to model it and thus constitutes a major challenge. Analyzing data of single-family-homes throughout Austria from 1998-2009, the prime intention of this research is an empirically comparison and evaluation of global and locally weighted state-of-the-art hedonic models in order to model SH.

In the global specifications - comprising an OLS, SAR, and S2SLS model - SH is considered by means of federal states dummies, representing homogeneous regions. The empirical model comparison reveals that, independent of the model, ignoring SH always consequence a lower model fit and worse prediction accuracies. The OLS estimates confirm findings reported in Yu et al. (2007), that the non-spatial model systematically tends to overestimate the importance of covariates. Besides correcting for SAC, the SAR and S2SLS model also fit substantially better than OLS. Moreover, the S2SLS coefficients show only minor changes compared with the SAR model, which is similar to Bivand (2010). The results clearly confirm that heterogeneity must be always taken into account, even if dummy variables are obviously less suitable to explain micro-geographic relationships. Exogenously defined regional indicators are only partially capable to account for SH, predominantly when SH beyond the selected dummies is present. Due to remaining and unmodeld SH, this requires a technical amendments in terms of a S2SLS considering SHAC to achieve a well-specified hedonic model.

Having provided evidence that global specifications are not fully capable to model SH, functional instability is further explored by LWRs, namely GWR and MGWR, which prevent the observed weaknesses of fixed effects. Both GWR and MGWR, are data-driven approaches to model smoothly varying marginal prices without adhering to regional dummies to model SH and are consequently being less prone to the MAUP effects. Compared with previous studies by Bitter et al. (2007), Yu et al. (2007), Páez et al. (2008), and Hanink et al. (2010) which assume without data support that *all* covariates have non-stationary effects on house prices, this research proposes MGWR dealing with both stationary and non-stationary effects simultaneously. Compared to the global models, MGWR is evidently more flexible, while being more parsimonious than GWR, which improves model efficiency (Wei and Qi, 2012). In contrast to the global models, the prediction error is reduced by both LWRs. It is assumed that the spatial heteroskedasticity in the SAR model originate from an inappropriate selection of fixed effects, which is also resolved by LWRs. Moreover, it is shown by the MGWR that significant spatial variation in *some* of the estimated parameters is present, while the global effects provide evidence for policy-based linkages and an economically connected housing market across Austria, which would be neglected by the traditional GWR.

There are, however, some limitations to this research. The variable selection is based on the global S2SLS model. Therefore, it is possible that different features in different parts of the study area may cancel each other out, incorrectly suggesting a lack of significance in the global model and, hence, omitted in the LWRs, which can cause, among others, omitted variable bias, although the non-significant residual SAC refutes this argument. Furthermore, computational burdens limit MGWR for

large datasets. Hence, more computation-efficient methods are promising such as tensor product smoothing (Wood, 2006).

Finally, this study results in some general implications for policy-makers, mortgage lenders, appraisers etc. LWRs produce meaningful and reliable estimates which are highly suitable to inform about local housing conditions. The spatial diversity of the coefficients are of utmost importance for locally acting decision-makers, requiring explicit knowledge of the local or regional housing market. This serves to refine policies and gain deeper understanding about local house price anomalies. Thus, it recommends that LWR should occupy a more central role in practitioners' toolboxes and methodologies.

## References

- Ahlfeldt, G. (2011) If Alonso was right: Modeling accessibility and explaining the residential land gradient. *Journal of Regional Science*, 51, pp. 318-338.
- Anas, A., Arnott R., Small, K.A. (1998) Urban spatial structure. *Journal of Economic Literature*, 36, pp. 1426-1464.
- Anselin, L. (1988) Spatial econometrics. Methods and models. Dordrecht: Kluwer.
- Bischoff, O., Maennig, W. (2011) Rental housing market segmentation in Germany according to ownership, *Journal of Property Research*, 28, pp. 133-149.
- Bitter, C., Mulligan, G., Dall'erba, S. (2007) Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method, *Journal of Geographical Systems*, 9, pp. 7-27.
- Bivand, R. (2010) Comparing estimation methods for spatial econometrics techniques using R. Norwegian School of Economics and Business Administration.
- Bivand, R., Pebesma, E., Gómez-Rubio, V. (2008) Applied spatial data analysis with R. Berlin: Springer.
- Bourassa, S., Hamelink, F., Hoesli, M., MacGregor, B. (1999) Defining housing submarkets, *Journal of Housing Economics*, 8, pp. 160-183.
- Can, A. (1998) GIS and spatial analysis of housing and mortgage markets, *Journal of Housing Research*, 9, pp. 61-86.
- Can, A., Megbolugbe, I. (1997) Spatial dependence and house price index construction, *The Journal of Real Estate Finance and Economics*, 14, pp. 203-222.
- Cassetti, E. (1972) Generating models by the expansion method: Applications to geographical research, *Geographical Analysis*, 4, pp. 81-91.
- Cho, S., Lambert, D.M., Kim, S.G., Jung, S. (2009) Extreme coefficients in geographically weighted regression and their effects on mapping, *GIScience & Remote Sensing*, 46, 273-288.

Dubin, R. (1992) Spatial autocorrelation and neighborhood quality, *Regional Science and Urban Economics*, 22, pp. 433-452.

Dubin, R. (1998) Spatial autocorrelation: A primer, *Journal of Housing Economics*, 7, pp. 304-327.

Farber, S., Páez, A. (2007) A systematic investigation of cross-validation in GWR model estimation: Empirical analysis and monte carlo simulations, *Journal of Geographical Systems*, 9, 371-396.

Farber, S., Yeates, M. (2006) A comparison of localized regression models in a hedonic house price context, *Canadian Journal of Regional Science*, 29, 405-420.

Fischer, M. M., Wang, J. (2011) *Spatial data analysis: Models, methods and techniques*. Berlin: Springer.

Florax, R., Rey, S. (1995) The impacts of misspecified spatial interaction in linear regression models, in: L. Anselin, R. Florax (Eds.) *New directions in spatial econometrics*, pp. 21-74. Berlin: Springer.

Fotheringham, S., Charlton, M., Brunsdon, C. (2002) *Geographically weighted regression. The analysis of spatially varying relationships*. Chichester: Wiley.

Gelman, A., Hill, J. (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Goldstein, H. (2011) *Multilevel statistical models*. Chichester, Wiley.

Goodman, A., Thibodeau, T. (2003) Housing market segmentation and hedonic prediction accuracy, *Journal of Housing Economics*, 12, pp. 181-201.

Hanink, D.M, Cromley, R.G., Ebenstein A.Y. (2010) Spatial variation in the determinants of house prices and apartment rents in China, *Journal of Real Estate Finance and Economics*, 45, 347-363.

Helbich, M., Brunauer, W., Hagenauer, J., Leitner, M. (2013a) Data-driven regionalization of housing markets, *Annals of the Association of American Geographers* (online first, DOI: 10.1080/00045608.2012.707587).

Helbich, M., Jochem, A., Mücke, W., Höfle, B. (2013b) Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning, *Computers, Environment and Urban Systems* (in press, <http://dx.doi.org/10.1016/j.comenvurbssys.2013.01.001>).

Kelejian, H., Prucha, I. (2007) HAC estimation in a spatial framework, *Journal of Econometrics*, 140, pp. 131-154.

Kestens, Y., Theriault, M., Des Rosiers, F. (2004) The impact of surrounding land use and vegetation on single-family house prices, *Environment and Planning B*, 31, pp. 539-567.

Kleiber, C., Zeileis, A. (2008) *Applied econometrics with R*. Berlin: Springer.

Lancaster, K. (1966) A new approach to consumer theory, *Journal of Political Economy*, 74, pp. 132-157.

LeSage, J., Pace, K. (2009) *Introduction to spatial econometrics*. Boca Raton: CRC Press.

Leung, Y., Mei, C.-L., Zhang, W.-X. (2000a) Statistical tests for spatial non-stationarity based on the geographically weighted regression model, *Environment and Planning A*, 32, pp. 9-32.

Leung, Y., Mei, C.-L., Zhang, W.-X. (2000b) Testing for spatial autocorrelation among the residuals of the geographically weighted regression, *Environment and Planning A*, 32, pp. 871-890.

Malpezzi, S. (2003) Hedonic pricing models: A selective and applied review, in: T. O'Sullivan, K. Gibb (Eds.) *Housing Economics and Public Policy*, pp. 67-89. Oxford: Blackwell

McDonald, J. (1997) *Fundamentals of urban economics*. Upper Saddle River, Prentice Hall.

McMillen, D.P. (2003) Spatial autocorrelation or misspecification? *International Regional Science Review*, 26, pp. 591-612.

McMillen, D.P. (2010) Issues in spatial data analysis, *Journal of Regional Science*, 50, pp. 119-141.

McMillen, D.P., Redfearn, C.L. (2010) Estimation, interpretation, and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science*, 50, pp. 712-733.

Mei, C.-L., Wang, N., Zhang, W.-X. (2006) Testing the importance of the explanatory variables in a mixed geographically weighted regression model, *Environment and Planning A*, 38, pp. 587-598.

Orford, S. (2000) Modelling spatial structures in local housing market dynamics: A multilevel perspective, *Urban Studies*, 37, pp. 1643-1671.

Pace, K., Barry, R., Sirmans, C. (1998) Spatial statistics and real estate, *The Journal of Real Estate Finance and Economics*, 17, pp. 5-13.

Pebesma, E. (2004) Multivariable geostatistics in S: The gstat package, *Computers & Geosciences*, 30, pp. 683-691.

Páez, A., Fei, L., Farber, S. (2008) Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques, *Urban Studies*, 45, pp. 1565-1581.

Páez, A., Farber, S., Wheeler, D. (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships, *Environment and Planning A*, 43, pp. 2992-3010.

Piras, G. (2010) Sphet: Spatial models with heteroskedastic innovations, *Journal of Statistical Software*, 35, pp. 1-21.

Redfearn, C.L. (2009) How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets, *Regional Science and Urban Economics*, 39, pp. 297-306.

Robinson, P.M. (1988) Root-N-Consistent Semiparametric Regression, *Econometrica*, 56, 931-954.

Rosen, S. (1974) Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy*, 82, pp. 34-55.

Sheppard, S. (1997) Hedonic analysis of housing markets, in: P. Chesire, E. Mills (Eds.) *Handbook of regional and urban economics*, Vol. 3, pp. 1595-1635. Amsterdam: Elsevier.

Stevenson, S. (2004) New empirical evidence on heteroscedasticity in hedonic housing models, *Journal of Housing Economics*, 13, pp. 136-153.

Ugarte, M., Goicoa, T., Milinto, A. (2004) Searching for housing submarkets using mixtures of linear models, in: J. LeSage, K. Pace (Eds.) *Spatial and spatiotemporal econometrics*, pp. 259-276. Oxford: Elsevier.

Wei, C.H., Qi, F. (2012) On the estimation and testing of mixed geographically weighted regression models, *Economic Modelling*, 29, 2615-2620.

Wheeler, D., Tiefelsdorf, M. (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression, *Journal of Geographical Systems*, 7, pp. 161-187.

Wheeler, D. (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A*, 39, 2464-2481.

Wilhelsson, M. (2002) Spatial models in real estate economics, *Housing, Theory and Society*, 19, pp. 92-101.

Wood, S. (2006) Low-rank scale-invariant tensor product smooths for generalized additive mixed models, *Biometrics*, 62, pp. 1025-1036.

Yu, D., Wei, Y., Wu, C. (2007) Modeling spatial dimensions of housing prices in Milwaukee, WI, *Environment and Planning B*, 34, pp. 1085-1102.

