

Modeling hedonic residential rents for land use and transport simulation while considering spatial effects

Michael Löchl
ETH Zürich^a

Kay W. Axhausen
ETH Zürich^b

Abstract: The application of UrbanSim requires land or real estate price data for the study area. These can be difficult to obtain, particularly when tax assessor data and data from commercial sources are unavailable. The article discusses an alternative method of data acquisition and applies hedonic modeling techniques in order to generate the required data. Many studies have highlighted that ordinary least square (OLS) regression approaches lack the ability to consider spatial dependency and spatial heterogeneity, consequently leading to biased and inefficient estimations. Therefore, a comprehensive data set is used for modeling residential asking rents by applying and comparing OLS, spatial autoregressive, and geographically weighted regression (GWR) techniques. The latter technique performed best with regard to model fit, but the issue of correlated coefficients favored a spatial simultaneous autoregressive model. Overall, the article reveals that when housing markets are a particular concern in UrbanSim applications, significant efforts are needed for the price data generation and modeling. The study concludes with further development potentials for UrbanSim.

Keywords: Hedonic technique; UrbanSim; asking rents; spatial effects; Switzerland

1 Introduction

One of the key components of UrbanSim is the use of land or real estate price data. These are applied in the model system as an indicator of the relative market valuations for attributes of housing, non-residential space, and location (Waddell and Ulfarsson 2003, 63). However, finding suitable data sources for real estate transaction prices and rents can be a challenge while setting up an UrbanSim application when transaction and rent price data from data suppliers are unavailable to researchers. As a minimum prerequisite for modeling purposes, sufficient information about a decent amount of properties is needed, including price (transaction cost or rent) and some explanatory spatial variables to which the model system should be sensitive in the application. These typically include at least some kind of regional accessibility, proximities, and neighborhood characteristics. The hedonic approach is a suitable method to model price values for every cell or parcel in the UrbanSim application based on such a sample. Moreover, the analysis exposes the implicit prices of housing and location characteristics. In the case of the UrbanSim application for Zürich (Löchl *et al.* 2007), tax assessor data and data from commercial sources were unavailable. Therefore, an alternative method of data acquisition was needed.

^aloechl@ivt.baug.ethz.ch

^baxhausen@ivt.baug.ethz.ch

One pillar of those efforts included a household survey whose purpose was manifold. For the hedonic modeling of residential rents, its sampling strategy was found to be too clustered, resulting in insufficient variance of the spatial explanatory variables. Another pillar was the use of a publicly available web-based portal of residential asking rents. The data were collected for the area of Canton Zürich from the end of 2004 to fall of 2005, and used to generate a basic hedonic model for the first application of UrbanSim in the Greater Zürich area. However, further analysis of the data revealed the need to consider spatial effects and the introduction of additional explanatory variables.

The hedonic approach was not adopted in the field of housing and real estate before the work of Lancaster (1971, 1966), Griliches (1971), and Rosen (1974). Today, it is regularly used in the field of real estate for property taxation and mortgage underwriting, but it has also been used for property price generation in land use and transport models (e.g. Waddell and Ulfarsson 2003).

Location is essential for determining housing prices. Bitter *et al.* (2007, 7) note that controlling for location and the spatial structure of markets is essential to explaining price differentials and deriving accurate coefficient estimates in hedonic residential price models. One common way to incorporate information about the location in hedonic models is to introduce distance to the central business district (CBD) or sub-market indicators including regional, local, or neighborhood specific binary coded dummy variables or interaction terms into the regression equation. However, previous studies have revealed that inclusion does not necessarily take all of the spatial effect into account (Clark 2007, 189; Wilhelmsson 2002, 100). Two types of spatial effects have been identified: spatial dependence and spatial heterogeneity (Anselin 1988, 8), both of which have been major challenges in spatial data analysis (e.g. Du and Mulley 2006, 201). Anselin (1988, 12) defines spatial dependence (also called spatial autocorrelation) as the existence of a functional relationship between what happens at one point in space and what happens elsewhere. Spatial heterogeneity (or spatial non-stationarity) may be present when there is a lack of uniformity from the effects of space or the spatial units of observation are not homogeneous. For example, price contributions of housing attributes may not be constant throughout a study area and may vary over space. Therefore, there may be spatial heteroscedasticity or spatially varying parameters present. Páez *et al.* (2007, 1566) summarize dependency as a locational/adjacent effect and heterogeneity as market segmentation. In general, housing markets often involve both spatial dependence and spatial heterogeneity due to localized supply and demand imbalances (Bitter *et al.* 2007, 8). De Graaff *et al.* (2001, 259) list three reasons why spatial dependence and heterogeneity should be considered jointly. First, there may be no differences between heterogeneity and dependence in an observational sense. Second, spatial dependency induces a particular form of heteroscedasticity (see also Keljian and Robinson 2000). Finally, it may be empirically difficult to separate the two effects. Overall, failure to incorporate spatial effects will result in biased or misleading coefficients and a loss of explanatory power.

For more than two decades, researchers and practitioners have considered spatial effects in hedonic regression models. Several advanced methods have been proposed to incorporate spatial structural instability, spatial drift and spatial lag into models (Leung *et al.* 2000, 10). These models commonly propose to make use of the spatial characteristics of variables to improve the results (Gao *et al.* 2006, 1040). Some of the most popular approaches are the spatially adaptive filtering methods (Trigg and Leach 1967; Widrow and Hoff 1960), expansion methods

(Fotheringham and Pitts 1995), multi-level approaches (Goldstein 1987; Jones 1991), spatial simultaneous autoregressive approaches such as spatial lag and spatial error models (Anselin 1988), as well as geographically weighted regression (GWR) models (Fotheringham *et al.* 2002). Beyond that, there are, for example, the local kriging and co-kriging methodology of Haas (1995, 1996) and the Bayesian spatially varying coefficient process models of Gelfand *et al.* (2003). Each of these approaches has its benefits and drawbacks, but all emphasize that parameters identified in global models may not resemble parameters estimated in local models. Therefore, they are often non-stationary.

It is the aim of this article to comparatively analyze outcomes from spatial simultaneous autoregressive approaches with results from GWR models that use the same data set, and furthermore, to suggest one method for the housing price modeling for the next application of UrbanSim in the greater Zürich area. At the same time, the article seeks to add to the discussion about appropriate methods for the consideration of spatial effects and non-stationarity in hedonic rent models. Moreover, this article describes an innovative way to acquire publicly available data and highlights possible further development potentials of UrbanSim applications. The focus is on residential rents in this article, due to the fact that Switzerland has (as compared with the United States and other European countries) a very low ownership rate of 37 percent in 2006 (Wüest & Partner 2007, 76), which is even lower in the Swiss urban areas. The reasons for the low ownership rate are manifold, including low land supply, rigorous mortgage down payment constraints, and tax regulation and landlord-tenant laws. But foremost, prices are high relative to rents and relative to household incomes and wealth (Bourassa and Hoesli 2010).

While there are many international studies on hedonic house and apartment transaction prices (for a comprehensive review of around 125 studies, see Sirmans *et al.* 2005), only few studies have focused on residential rents. Those rent studies either use the net annual or monthly rent (Baranzini and Ramirez 2005; Baranzini and Schaerer 2007; Baranzini *et al.* 2006; Brunauer *et al.* 2009) or the monthly gross rent (Banfi *et al.* 2006). For other hedonic rent studies, it is not specified if gross or net rents have been used (e.g. Djurdjevic *et al.* 2008; Sirmans *et al.* 1989; Valente *et al.* 2005). Of all those studies, only a minority has focused on rents per square meter (Brunauer *et al.* 2008), while the vast majority have been using the absolute rent or its log-transformation.

The remainder of this article is organized as follows: Section 2 introduces the methods of spatial simultaneous autoregressive models and geographically weighted regression. Section 3 gives a short overview of earlier studies, which compared the results of various hedonic modeling techniques. The data are introduced and described in Section 4, followed by the estimation results in Section 5. In Section 6, the findings are summarized.

2 Spatial simultaneous autoregressive models and geographically weighted regression (GWR)

Spatial simultaneous autoregressive modelling is a popular approach to consider spatial effects. It assumes that the response variable at each location is a function not only of the explanatory variable at that location, but of the response at neighbouring locations as well. The models are based on maximum-likelihood estimations and commonly applied in the fields of regional science, sociology, political science and the various fields of economics (Anselin 2001, 310). A comprehensive introduction might be found in Anselin (1988). Various studies are also avail-

able in the field of real estate appraisal, such as [Kim *et al.* \(2003\)](#), [Shin *et al.* \(2007\)](#), and [Wilhelmsson \(2002\)](#).

Three different spatial simultaneous autoregressive models are usually distinguished, depending on where the spatial autoregressive process is believed to occur ([Kissling and Carl 2008, 61](#)). The spatial simultaneous autoregressive lag model (SARlag) assumes that the autoregressive process occurs in the response variable. A spatial lag hedonic rent model can be written as follows:

$$P = \rho WP + \beta X + \varepsilon \quad (1)$$

where P is a vector of rents, ρ is a spatial autocorrelation parameter, W is a $N \times N$ spatial weight matrix (where N is the number of observations), β is a vector of regression coefficients, X is a matrix with observations on structural and spatial explanatory characteristics and ε is assumed to be a vector of independent and identically distributed (iid) error terms. Typically, the definition of neighbors used in the weights matrix is based on a notion of distance decay or contiguity.

When spatial dependence is present in the error term, a spatial autoregressive specification for this dependence is usually assumed. This is called the spatial simultaneous autoregressive error model (SARerr) and can be formulated as follows:

$$P = \beta X + u \quad (2)$$

i.e., a linear regression with error vector u , and

$$u = \lambda W u + \varepsilon \quad (3)$$

where λ is the spatial autoregressive coefficient, W is the spatial weight matrix, and u is assumed to be a vector of independent and identically distributed errors. This model is a special case of a regression specification with a non-spherical error variance-covariance matrix. Therefore, W now pertains to shocks in the unobserved variables (the errors u) but not to the explanatory variables of the model (X). Consequently, the price at any location is a function of the local characteristics but also of the omitted variables at neighboring locations. This is the most popular spatial simultaneous autoregressive model and it is widely used in the literature ([Taylor 2008, 25](#); [Wilhelmsson 2002, 94](#)).

Finally, spatial autocorrelation can affect both response and explanatory variables, having both "inherent spatial autocorrelation" and "induced spatial dependence." In this case, an additional term ($WX\gamma$) must appear in the model, which describes the autoregression coefficient (γ) of the spatially lagged explanatory variables (WX) as [Kissling and Carl \(2008, 61\)](#) describe. The so-called spatial Durbin model (SARmix) takes the form

$$P = \rho WP + \beta X + WX\gamma + \varepsilon. \quad (4)$$

According to [Valente *et al.* \(2005, 110\)](#), one of the advantages of those models are a (nearest) neighbor-based smoothing of the means and convenient computation besides the often shown improvement over OLS models.

The technique of GWR has been developed by [Brunsdon *et al.* \(1998\)](#); a more comprehensive overview is found in [Fotheringham *et al.* \(2002\)](#). The method attempts to incorporate

geographical information into a regression model using a series of distance-related weights. Essentially, it consists of a series of locally linear regressions that use distance-weighted overlapping samples of the data (Farber and Yeates 2006, 412). The method explicitly allows parameter estimates to vary over space, which leads to independent spatial error terms. Rather than specifying a single global model to characterize the entire housing market, GWR estimates a separate model for each data point and weights observations by their distance to this point, thus allowing unique marginal price estimates at each location (Bitter *et al.* 2007, 10).

The typical output from a GWR model is a set of parameters that can be mapped in the geographic space to represent non-stationarity or parameter drift. Similarly, local measures of standard errors and goodness-of-fit statistics are obtained (Fotheringham *et al.* 2000, 113). Therefore, the additional benefit of the GWR approach is that it offers the potential of increased understanding of the nature of varying relationships between variables across space. Du and Mulley (2006, 201) describe GWR as an alternative to spatial simultaneous autoregressive models which “is perhaps more intuitive.” The GWR approach has recently found use in various applications. There are examples in the fields of climatology (Brunsdon *et al.* 2001), ecology (Kimsey *et al.* 2008; Zhang and Shi 2004), education (Fotheringham *et al.* 2001), marketing research (Mittal *et al.* 2004), regional science (Huang and Leung 2002), political science (Calvo and Escolar 2003), and transport research (Chow *et al.* 2006; Clark 2007; Hadayeghi *et al.* 2003; Lloyd and Shuttleworth 2005; Nakaya 2001). In the housing field, there are studies by Bitter *et al.* (2007), Farber and Yeates (2006), Fotheringham *et al.* (2002), Kestens *et al.* (2006), Páez *et al.* (2007), and Yu *et al.* (2007). To our knowledge, there have been no studies on residential rents employing GWR so far, as all of the authors mentioned above focus on apartments or single-family house transaction prices. Overall, GWR is considered a standard tool in exploratory spatial data analysis due to its effectiveness and wide applications (Wang *et al.* 2008, 987). Nevertheless, there are some limitations to the method, and authors have pointed out that GWR results should be interpreted with caution (Shearmur *et al.* 2007, 701). In particular, multicollinearity and correlation among local regression coefficients is a problem in GWR, even in the presence of uncorrelated exogenous variables in the data generating process. The issue was raised by Wheeler and Tiefelsdorf (2005), who emphasized that the effects of multicollinearity are substantially stronger in the GWR model than in global regression models. They suggested that this potentially invalidates any interpretation of individual GWR parameter estimates and reduces confidence in the method for more than exploratory purposes—for example, there may be inflated variances and, at times, counter-intuitive and contradictory in sign to the global regression estimates. Young *et al.* (2008, 4013) notes that local regression models are designed as exploratory smoothing methods and not as inferential statistical tools.

Therefore, several authors have compared the GWR models with alternatives in this regard, particularly with methodological improvements of the GWR approach (Griffith 2008; Wheeler 2007, 2009) and Bayesian spatially varying coefficient process models (Wheeler and Waller 2009). Waller *et al.* 2007, 585 note that more methodological research is required to determine the most flexible and robust structures. Additionally, the computational cost burden of alternative methods might be high (Wheeler and Waller 2009, 21). Overall, the approaches are still in an experimental phase and the debate about the most appropriate methodology is ongoing.

A weighting function is applied in GWR in order to give greater influence to close data points. A spatial kernel is usually used. Choosing the shape of the kernel and its bandwidth are both important issues and various options are available (Fotheringham *et al.* 2002, 56). The bandwidth choice is not a parameter relating to the model itself, but is essentially part of the calibration strategy for a given sample (Fotheringham *et al.* 2002, 63).

In the case of rent estimations, a GWR model can be written as follows:

$$P_i = \beta_{i0} + \sum_k \beta_{ik} X_{ik} + \varepsilon_i \quad (5)$$

where P_i is the i th observation of the rent, β_{i0} is the constant estimated for local regression i , β_{ik} is the regression coefficient of structural or spatial explanatory variable k estimated for local regression i , and ε_i is the i th value of a normally distributed error vector with mean equal to zero. This differs from OLS by using distinct constants and regression parameters for each data point. The estimation algorithm iterates through N OLS, each one modified by a unique distance-decay weight matrix. The estimation takes the form:

$$\beta_i = (X^T W_i X)^{-1} X^T W_i P \quad (6)$$

where β_i is the vector of estimated coefficients for observation i , X is the $N \times K$ matrix of explanatory variables, W_i is a diagonal distance-decay weight matrix customized for the location of i relative to the surrounding observations, and P is the vector of observed rents.

3 Comparative studies in the literature

Recently, hedonic literature has shifted from simple applications of local models to comparative studies, where several local model approaches are tested and evaluated. Because of the increasing popularity of the approaches, the need for assessing the relative merits of the different modeling techniques is obvious (Páez *et al.* 2007, 1566). Consequently, comparisons of the performance of different measures, which take into account spatial effects, have begun to emerge in the literature. Clark (2007) applied a spatial error model and GWR for estimating local car ownership in the United Kingdom, but did not focus on comparing the results in particular. Wall (2004) compared conditional autoregressive models (CAR) and simultaneously autoregressive models, but found possible counter-intuitive or impractical results.

Bitter *et al.* (2007) compared the spatial expansion method and the GWR approach for a dataset from Tuscon, Arizona, applying seven housing characteristics, including two from a principal components analysis. While they found spatial variation in both models, they could not separate the observed spatial heterogeneity by localized supply and demand characteristics and the impact of omitted variables. However, GWR outperformed the spatial expansion specification in terms of explanatory power and predictive accuracy in their study, although the differences were narrowed to some degree with the addition of the spatial lag term in the expansion specification. The authors concluded that when explanatory power and predictive accuracy are the primary objectives, GWR is the superior approach (Bitter *et al.* 2007, 23).

Farber and Yeates (2006) compared OLS with a spatial lag model, GWR, and a moving window regression (MWR) approach. They found that the GWR approach is superior to all others; therefore, it may be regarded as the one which accounts best for the spatial variation.

They concluded that by allowing parameters to vary spatially, estimation accuracy of the response variable improves dramatically while spatial biases diminish to nominal amounts (Farber and Yeates 2006, 417). However, they observed some extreme coefficients in their GWR model, which were clustered in specific neighborhoods of their study area. They noted that in light of the unavailability of a robust statistical framework for GWR and MWR, irrational coefficients pose a major threat to the adoption of GWR by assessment authorities (Farber and Yeates 2006, 418). The finding of extremely high coefficient variability is not unusual and often found in practice. This has raised some concern that GWR results may be misleading because of the suspicion that the variability observed is somehow built into the model by its own calibration and estimation mechanisms (Páez 2005, 163). By comparing GWR and the expansion method in a simulation exercise, Páez concluded that, on average, spatial variability in GWR is not an artifact of the calibration procedure, and that GWR is sufficiently flexible to reproduce the type of map patterns used in the simulation experiment. Moreover, he found that both approaches are able to provide reasonable representation of the spatial patterns inherent in the simulated data. Kestens *et al.* (2006) compared GWR and the expansion method as well, highlighting the individual merits of both methods. While GWR provides additional insight by measuring local regression statistics, the expansion term makes it possible to analyze and explain the cause of parameter heterogeneity, whether its structure is spatial or not. They concluded that both methods are complementary rather than substitutes for each other (Kestens *et al.* 2006, 94).

Páez *et al.* (2007) compared moving windows regression (MWR), GWR, and moving windows kriging. They found that the MWR approach leads to superior results relative to single market and global approaches and also to modeling spatial dependencies. Therefore, they concluded that market segmentation may be more important than spatial dependencies.

The authors of the above articles used various measures to compare and evaluate the model approaches. The most popular OLS goodness-of-fit measure is the adjusted R -square. This measure is useful and applied in GWR as well, but it is no longer applicable in spatial simultaneous autoregressive models (Yu *et al.* 2007, 1092). Therefore, other measures such as residual sum of squares (RSS), sum of squared errors (SSE), maximum likelihood value, Log likelihood (Loglik), and Akaike's Information Criterion (AIC) are applied. The null hypothesis that the contribution of a relationship is zero is often investigated based on t -test or F -test (Gao *et al.* 2006, 1041). Additionally, model prediction power verification is evaluated through out-of-sample testing (e.g. Bitter *et al.* 2007).

4 Data and variables

Butler (1980, 97) states that, in principle, all characteristics relevant to the determination of market price should be included in a hedonic function. In practice, this cannot be done because the number of such characteristics is unmanageably large, and data on many of these are either unavailable or of poor quality. In addition, even without data constraints, a relatively small number of explanatory variables lead to considerable multicollinearity. Therefore, the aim is to find a broad set of statistically significant variables with expected signs and moderate impact of multicollinearity, while the estimate should have a sufficient model fit. Butler (1980, 97) notes that for the aforementioned reasons, any estimate of the hedonic relationship must be misspecified because some of the relevant explanatory variables must be omitted. He concludes

that consequently all estimates are to some extent “incorrect” and differences among them must be attributed at least in part to differences in adaptation to the specification problems common to all. However, ample and accurate variable specification is essential to infer properly and to generate statistically significant results. Finding suitable data sources in the field of real estate prices when transaction and rent data are not open to researchers presents an additional challenge.

The underlying data for this study were taken from a publicly available Web site¹ between December 2004 and October 2005. The database included rent offers from various Swiss real estate online platforms. Duplicate entries, furnished and shared apartment offers, and apartments of unusual sizes (<20 and >500 m²) were dropped. The addresses for all dwelling units in the dataset were geocoded at building level and matched with a wide set of spatial variables. The generation of some variables included significant further work (see Löchl 2007), for details on the calculation of solar radiation and visibility variables based on a digital elevation model); others were simply matched with available data or layers by a geographic information system (GIS). Generally, the availability of additional non-spatial and spatial data are rather good in Switzerland—for example, both the population census and the business census are available at the hectare level. Overall, the dataset comprehended rents and additional information of 8,592 dwelling units in Canton Zürich, which consists of 171 municipalities.

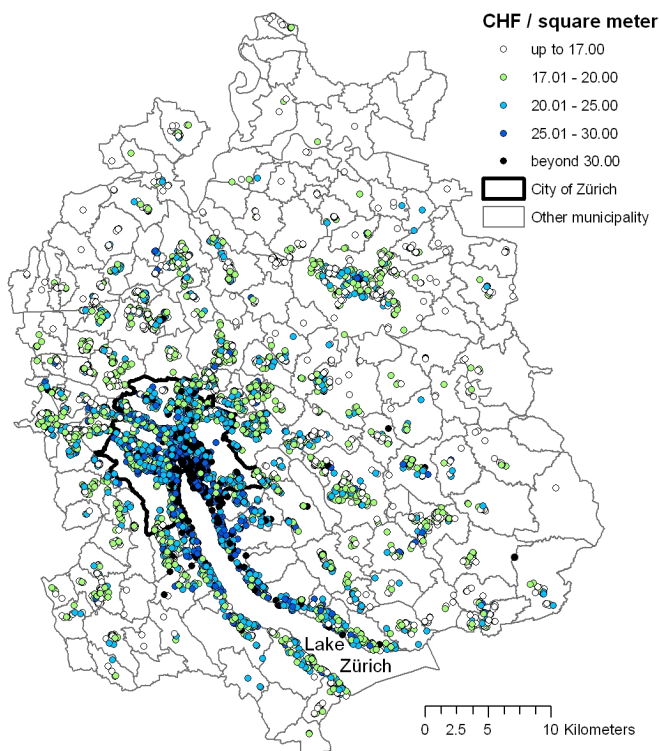
Because of the source and characteristics of the input price data, several restrictions apply. As the dataset includes asking rents, it does not necessarily reflect paid market prices. Moreover, the sample might be slightly biased because certain vacant dwelling units do not make their way to the Web site. However, an earlier data comparison showed that the differences in the structural variables to the complete inventory in the Federal Building and Apartment Register of the Federal Statistical Office are minor (Löchl 2006, 6). Additionally, the structural characteristics of the properties might include faulty information, as they are self-reported by the person placing the ad. Finally, only a limited set of structural dwelling unit characteristics were available for the estimations. It included the price information and the dwelling unit size in square meters, the number of rooms (a value which counts bedrooms, living room, and kitchen), and some information about available facilities, such as balcony, fireplace, terrace/garden, and the availability of a lift in the building. In cases where the age of the building was unavailable, the records were matched with the Federal Building and Apartment Register to generate the information. Some variables, which might have been important contributors, were unavailable and had to be left out of the equations, such as information about the last major renovation and garage availability. However, parking costs are usually not included in the net rent. Another known relevant factor is landlord identity since non-profit organizations such as cooperative building associations offer rents well below average market rents. They can be found in the whole of Canton Zürich and have a market share of almost 20 percent (Statistical Office of Canton Zürich 2008, 15). However, the landlord identification was not possible with the data, which leads potentially to some bias. In Figure 1, the observations are displayed as the monthly asking rent per square meter.

The selection of explanatory variables was the result of exploratory data analysis through OLS stepwise regression in SPSS Version 16.² The final dataset included two response and

¹ www.comparis.ch

² The GWR software does not provide functions for variable selection such as forward, backward, and stepwise methods.

Figure 1: Net asking rent per square meter in CHF per month in Canton Zurich.



31 explanatory variables. The descriptive statistics are found in Table 1. Basically, all available structural variables in the source data were used except for the number of rooms, as there is high multicollinearity with the dwelling unit floor area. Additionally, a broad set of spatial variables were found to be significant without considerable correlation among each other and with the structural variables. One was public transport accessibility (PTACC) besides car accessibility (CARACC), as the study area has a well-developed public transport system. The last micro-census in 2005 ([Statistical Office of Canton Zürich 2008](#)) revealed that public transport is used by the inhabitants for almost 30 percent of all kilometers traveled. A remarkable feature of the sample is that the mean distance to the next railway station is 910 meters on average. The whole region of Zürich is connected by a dense network of suburban railways, and all stations are served at least once per hour, with most served twice or even four or more times per hour, assuring a very high quality of service. The seven explanatory time dummy variables not only reflect rent changes over time, but also include the known process that landlords tended to reassign specific costs from the net rent to the additional costs in that period.

Table 1: Descriptive statistics of the final variable selection.

Variable	Description	Type ¹	Min	Max	Mean	S.D.
Dependent variables						
RENT	Monthly net asking rent in CHF	C	476.00	15 000.00	1845.26	908.60
SQRENT	Monthly net asking rent per square meter in CHF	C	10.60	53.57	20.62	5.35
Structural explanatory variables						
SQM	Floor area in square meter	C	20.00	400.00	91.58	36.92
LIFT	Building has a lift	D	0.00	1.00		
FIREPLACE	Dwelling unit has a fireplace	D	0.00	1.00		
BALCONY	Dwelling unit has one or more balconies	D	0.00	1.00		
G'TERRACE	Dwelling unit has a garden terrace	D	0.00	1.00		
ISHOUSE	Single family house	D	0.00	1.00		
BUILBEF20	House built before 1921	D	0.00	1.00		
BUIL21TO30	House built 1921–1930	D	0.00	1.00		
BUIL31TO80	House built 1931–1980 ²	D	0.00	1.00		
BUIL81TO90	House built 1981–1990	D	0.00	1.00		
BUIL91TO05	House built 1991–2005	D	0.00	1.00		
Spatial explanatory variables						
CARTT_CBD	Average travel time to Zürich CBD by car in min	C	8.00	58.4	29.92	9.22
CARACC	Regional car accessibility to employment	C	7.18	10.30	9.22	0.57
PTACC	Regional public transport accessibility to employment	C	-19.46	12.39	10.69	1.60
RAILSTATION	Euclidean distance to next rail station in km	C	0.01	5.73	0.91	0.66
AUTOBAHN	Autobahn within 100 m	D	0.00	1.00		
AIRNOISE	Daily avg. air noise above 52 dB	D	0.00	1.00		
HOTREST_JOBS	Number of jobs in hotel and restaurant industry within 1 km	C	1.27	7187.78	316.33	782.43
POP_DENS	Number of inhabitants in hectare	C	0.00	2004.00	92.95	71.43
FOREIGNERS	Fraction of foreigners ³ in hectare	C	0.00	0.50	0.05	0.04

Continued

Variable	Description	Type ¹	Min	Max	Mean	S.D.
TAXLEVEL	Local income tax level for individuals	C	69.00	122.00	110.36	14.12
SLOPE	Slope by 25 m raster	C	0.00	0.26	0.36	0.03
VIEW_LAKE	Visibility ⁴ of lake surface (> 1 km ²) in hectare	C	0.00	8887.81	439.68	1050.07
VIEW_ALL	Total visibility ⁴ of terrain surface in hectare	C	90.75	78 421.00	8840.15	5638.71
SOLAR_EVE	Evening solar exposure index	C	0.25	607.11	238.22	80.85

¹ C = Continuous; D = Dummy
² Reference case.
³ Foreigners are defined as inhabitants with nationalities outside of Northwestern Europe, North America, and Australia.
⁴ Visibility is calculated based on a 25 m DEM, not considering buildings, trees, etc.

5 Estimation

In general, hedonic theory does not strictly specify a functional form (Cropper *et al.* 1988; Halvorsen and Pollakowski 1981). Sirmans *et al.* (2005, 6) explain that studies have wrestled with the problem of the correct functional form and no consensus has been found of which is most appropriate.³ The semi-log specification has some advantages, such as the coefficients can be easily interpreted as the percentage change in the price given a one-unit change in the characteristic. Moreover, it helps minimize the problem of heteroscedasticity and mitigates the impact of nonlinear relationships between market price and the explanatory variables (Malpezzi 2003). Therefore, hedonic pricing equations are typically estimated using either linear or semi-logarithmic models (Sirmans *et al.* 2005, 4). A second challenge is the adequacy of parametric specification. Some authors (e.g. Anglin and Gençay 1996, 633; Martins-Filho and Bin 2005, 93ff) indicate that this problem arises from the inability of economic theory to provide guidance on how characteristics of similar products relate functionally to their market price. Consequently, there have been attempts to use semi- or nonparametric methods (Clapp 2003, 2004; Fahrländer 2005; Pace 1993, 1998; Parmeter *et al.* 2007), which allow for the possibility of nonlinearity in the hedonic price functions and flexible modeling of the influence of continuous covariates on the dependent variable.

In this study, after variable selection, further data modeling and diagnostics were performed in the open-source statistical software R Version 2.8⁴ (R Development Core Team 2008). Fox (2002) provides a good introduction into regression analysis with R. For a technical introduction into the R package *spdep* and available functions for spatial data analysis, see Bivand (2008). Krström (2008) provides an example of the estimation of a hedonic model with the statistical software R. Both the dependent and the explanatory variables were transformed as suggested by Mosteller and Tukey (1977, 109), whereas a logarithmic transformation was taken for an amount or count and a logit transformation was used for fractions.⁵ Due to memory capacity issues of the GWR methods in R (the computational overhead of this method is considerable, especially for a large dataset), the GWR 3.0 software by Martin Charlton and Stewart Fotheringham was applied for the GWR approach. A documentation of the software is available in Fotheringham *et al.* (2002).⁶

Two OLS models (Models 1 and 2) were selected for the analysis including both structural and spatial explanatory variables, taking the logarithmic transformation of RENT as the dependent variable (see Table 2). Similar models, considering rent per square meter as the dependent variable, did not show any significant difference of the coefficients, besides the one referring to the floor area variable, but had a much lower model fit and are therefore not presented here.

³ Functional forms for the hedonic price function that have been proposed and used in the literature include the linear, the quadratic, the log-log, the semi-log, the inverse semi-log, the exponential, and the Box-Cox transformation (Freeman 1993, 379).

⁴ <http://www.r-project.org>

⁵ Tests with Box-Cox transformations did not succeed since there are some independent dummy variables where more than two thirds of the sample is coded zero, making it impossible to automatically compute a sufficiently large constant in R (Fox 2002, 107). As it is, such specifications are not readily implemented in the presence of spatial dependence (Kim *et al.* 2003, 31).

⁶ The fact that some variables do not show local variability (e.g. AIRNOISE) actually requires a mixed GWR approach, in which the referring coefficients are global and the others are local as in the basic GWR model (Fotheringham *et al.* 2002, 65ff). However, this procedure is not implemented in the GWR 3.0 software; it is planned to be included in the upcoming 4.0 release.

The OLS results of two additional models (Models 3 and 4), which are only considering spatial explanatory spatial variables, are found in the appendix. While models with solely spatial explanatory variables can be used in UrbanSim currently, it is hoped that in the future the simulation system becomes sensitive to structural variables of the building stock, since this presents the opportunity to better reflect the local situation and to make the simulation more realistic.

The explanatory variable selection is essentially the same in Models 1 and 2 with one exception: Model 1 uses car travel time to Zürich Central Business District (CARTT_CBD), while Model 2 considers regional car accessibility to employment (CARACC). Both explanatory variables could not be used in one model due to multicollinearity issues. Model 1 using CARTT_CBD reveals the best overall model fit, whereas Model 2 including CARACC has more policy relevance, as it is sensitive to any improvement in the street network beyond those affecting the travel time to Zürich CBD.

Outliers were identified as suggested by [Fotheringham *et al.* \(2002, 78\)](#) and [Chatfield \(1995, 265\)](#) where a data point with the absolute value of the studentised residual exceeding 3 is a potential outlier. This threshold has been applied for both models. Efforts of modeling interaction of the selected explanatory variables, particularly those with slope and the total view variable, did not improve the models. SLOPE might indeed have an intrinsic value, as it usually comes with a certain potential view (at least of immediate surroundings), while VIEW_ALL and VIEW_LAKE include the total visible area in hectares, which is rather coarse and considers, in particular, the long-range view.

The adjusted R -square values indicate good model fits, while Model 1 performs slightly better than Model 2. Researchers often apply out-of-sample testing for the purpose of model accuracy and model prediction power verification, where a proportion of the available data records are left out from the estimations and used to be compared to estimated outcomes. However, in this study, no such analysis was performed, as data fitting was the priority, and there was less focus on the predictive power of the models.

When detecting and incorporating spatial effects, it is necessary to produce a weight matrix based on some kind of contiguity. Several approaches are available (e.g. [Anselin 2002, 258](#); [Bivand 2008, 251ff](#)). Because of a relatively high heterogeneity of spatial distribution of the data points (see Figure 1), a k -nearest-neighbors approach by Euclidean distance was chosen, whereas $k = 9$ produced the best results in terms of model fit measured by log likelihood (LogLik) and Akaike Information Criterion (AIC) for the spatial simultaneous autoregressive models. Therefore, each observation has the nine next observations in terms of linear distance defined as contiguity in the weight matrix.

A set of diagnostic tests for spatial autocorrelation were performed based on the contiguity weight matrix for both models, which clearly indicated the need to consider spatial autocorrelation in the models. This is unsurprising, given that most variables in the estimation actually have a spatial relation. The global Moran's I measure clearly indicated spatial autocorrelation in both cases, while the Lagrange Multiplier tests ([Anselin *et al.* 1996](#)) pointed to both spatial lags and spatial errors in the models, as can be seen by the high significance level even in their robust versions, although the test values are much higher for the robust Lagrange Multiplier test of the spatial error (Robust LMerr).

Consequently, the SARerr and the SARmix approach were considered besides the GWR approach in the following analysis, all applying the variable selection from Model 2 due to its higher policy relevance as described above. Visually inspecting the resulting residual scatter

Table 2: Estimated OLS parameters ($N = 8592$).

Y = Ln(RENT) Explanatory Variable	Model 1				Model 2			
	Estimate	SE	sig.	VIF	Estimate	SE	sig.	VIF
(Constant)	5.214				3.638			
Ln(SQM)	0.777	0.005	***	1.392	0.776	0.005	***	1.394
LIFT	0.026	0.004	***	1.160	0.024	0.005	***	1.159
FIREPLACE	0.108	0.010	***	1.167	0.118	0.011	***	1.167
BALCONY	0.033	0.004	***	1.186	0.033	0.004	***	1.186
GTERRACE	0.105	0.015	***	1.013	0.107	0.016	***	1.013
ISHOUSE	0.133	0.013	***	1.153	0.125	0.013	***	1.153
BUILTBEF21	0.077	0.006	***	1.230	0.109	0.006	***	1.209
BUILT21TO30	0.086	0.011	***	1.062	0.094	0.011	***	1.061
BUILT81TO90	0.018	0.005	***	1.199	0.018	0.005	***	1.199
BUILT91TO05	0.067	0.005	***	1.305	0.067	0.005	***	1.306
Ln(CARTT_CBD)	-0.294	0.007	***	2.089				
CARACC					0.119	0.005	***	2.471
PTACC	0.006	0.001	***	1.472	0.011	0.001	***	1.493
Ln(RAILSTATION)	-0.012	0.002	***	1.184	-0.012	0.003	***	1.232
AUTOBAHN	-0.060	0.013	***	1.033	-0.067	0.013	***	1.033
AIRNOISE	-0.039	0.006	***	1.281	-0.096	0.006	***	1.276
Ln(HOTREST_JOBS)	0.020	0.002	***	2.339	0.032	0.002	***	2.447
Ln(POP_DENS)	-0.028	0.002	***	1.242	-0.026	0.002	***	1.243
FOREIGNERS (logit)	-0.018	0.002	***	1.340	-0.023	0.002	***	1.338
Ln(TAXLEVEL)	-0.130	0.014	***	1.326	-0.223	0.015	***	1.292
SLOPE (logit)	0.016	0.002	***	1.285	0.026	0.002	***	1.255
Ln(VIEW_LAKE)	0.008	0.001	***	1.759	0.012	0.001	***	1.944
Ln(VIEW_ALL)	0.005	0.002	**	1.450	-0.001	0.002		1.528
Ln(SOLAR_EVE)	0.037	0.003	***	1.107	0.018	0.003	***	1.089
1Q_04	0.080	0.013	***	1.195	0.079	0.014	***	1.195
2Q_04	0.038	0.012	***	1.228	0.032	0.013	**	1.228
3Q_04	0.059	0.009	***	1.618	0.057	0.009	***	1.618
4Q_05	0.032	0.007	***	2.612	0.033	0.007	***	2.613
1Q_05	0.045	0.007	***	2.616	0.044	0.007	***	2.616
2Q_05	0.026	0.007	***	2.035	0.024	0.008	***	2.035
3Q_05	0.037	0.007	***	2.526	0.034	0.007	***	2.526
Adjusted R-square	0.854				0.836			
F-test	1677.000		***		1460.000		***	

Probability of rejecting H_0 : *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 3: OLS diagnostics ($N = 8592$).

Measure	Model 1		Model 2	
	Value	sig.	Value	sig.
LogLik	3685.653		3183.261	
AIC	-7307.306		-6302.522	
Global Moran's I	0.213	***	0.287	***
LMerr	2164.867	***	3918.393	***
LMlag	454.444	***	977.235	***
Robust LMerr	1722.903	***	2981.065	***
Robust LMlag	12.480	***	39.907	***

Probability of rejecting H_0 : *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

plots of the model indicated a problem with heteroscedasticity, as very low and very high observed asking rents tended to be overestimated. The problem may have been caused by missing and unavailable crucial structural variables, such as information about building conditions or the landlord. However, the standard error of the OLS model was compared to the results based on the heteroscedasticity consistent covariance matrix and did not show any significant difference (for the procedure in R, see [Bivand 2008](#), 290). Therefore, heteroscedasticity was not considered to be a serious problem in this case.

Since the data points were not randomly located and sometimes spaced widely about (see Figure 1), an adaptive kernel was used for the GWR approach. As there was no prior justification for supplying a particular bandwidth, the two available options for finding the most appropriate bandwidth were tested: 1) minimization of the cross-validation (CV) score and 2) minimization of the AIC measure ([Fotheringham et al. 2002](#), 212). Both methods suggested the same bandwidth. In Table 4, the results of the OLS, SARerr, SARmix, and GWR estimations, based on the variable selection in Model 2, are shown. For the GWR model, the mean and the standard deviations of the parameter values are presented. The parameter estimates do not vary dramatically among the models (exceptions are CARACC, AIRNOISE, and VIEW_ALL, for which the GWR model indicates a large spatial variation of those variables indicated by multiple standard deviation values compared to the mean of the estimates). For model evaluation purposes, both the sum of squares of the errors (SSE) and the AIC have been calculated. In both measures, smaller values indicate better model fit. In this case, both measures identify the best model fit for the GWR estimates; second best, the SARmix model. However, four explanatory variables (PTACC, AIRNOISE, TAXLEVEL, and SOLAR_EVE) are not significant in the SARmix (as indicated by the values marked with a * in Table 4), whereas AIRNOISE even has a counter-intuitive sign. Particularly with regard to the envisaged application of the hedonic model in UrbanSim, the insignificance of the public transport accessibility coefficient creates a major limitation of the model.

In order to further assess and compare the predictive accuracy of spatial autoregressive approaches and the GWR model, the predicted rents were compared with the observed values. The results of this analysis (presented in Table 5), show that the OLS approach has the lowest accuracy. Less than 66 percent of the predictions were within the range of ± 2 percent of the actual rent price and it showed the lowest accuracy within ± 5 percent and ± 10 percent as

Table 4: Comparison of coefficients from various modeling approaches.

Y = Ln(RENT)	OLS Estimate	SARerr Estimate	SARmix Estimate	GWR	
				Estimate mean	Std.dev. ¹
(Constant)	3.638	3.611	1.329	3.954	
Ln(SQM)	0.776	0.773	0.773	0.772	0.055
LIFT	0.024	0.026	0.025	0.022	0.017
FIREPLACE	0.118	0.085	0.088	0.096	0.038
BALCONY	0.033	0.030	0.030	0.031	0.011
GTERRACE	0.107	0.094	0.097	0.098	0.073
ISHOUSE	0.125	0.137	0.141	0.138	0.047
BUILTBEF21	0.109	0.082	0.075	0.091	0.026
BUILT21TO30	0.094	0.075	0.070	0.084	0.042
BUILT81TO90	0.018	0.019	0.020	0.026	0.024
BUILT91TO05	0.067	0.080	0.085	0.078	0.023
CARACC	0.119	0.113	0.068	0.011	0.098
PTACC	0.011	0.006	0.002*	0.039	0.034
Ln(RAILSTATION)	-0.012	-0.013	-0.016	-0.009	0.016
AUTOBAHN	-0.067	-0.076	-0.079	-0.071	0.066
AIRNOISE	-0.096	-0.083	0.032*	-0.014	0.032
Ln(HOTREST_JOBS)	0.032	0.029	0.014	0.025	0.017
Ln(POP_DENS)	-0.026	-0.028	-0.028	-0.029	0.014
FOREIGNERS (logit)	-0.023	-0.014	-0.013	-0.018	0.007
Ln(TAXLEVEL)	-0.223	-0.201	-0.002*	-0.160	0.268
SLOPE (logit)	0.026	0.016	0.007	0.010	0.009
Ln(VIEW_LAKE)	0.012	0.010	0.003	0.008	0.007
Ln(VIEW_ALL)	-0.001*	0.008*	0.017	0.006	0.024
Ln(SOLAR_EVE)	0.018	0.012	0.006*	0.024	0.027
1Q_04	0.079	0.078	0.079	0.072	0.034
2Q_04	0.032	0.046	0.044	0.040	0.036
3Q_04	0.057	0.067	0.066	0.060	0.021
4Q_05	0.033	0.038	0.036	0.032	0.020
1Q_05	0.044	0.050	0.049	0.048	0.020
2Q_05	0.024	0.027	0.026	0.024	0.018
3Q_05	0.034	0.040	0.039	0.037	0.022
Lambda		0.637			
Rho			0.599		
SSE	239.781	182.655	179.700	174.202	
AIC	-6302.522	-8169.980	-8319.190	-8392.987	

* Not significant at 0.05 level

¹ Standard deviation of coefficient values across study area

well. The other three models performed considerably better, whereas around 73 percent are in the ± 2 percent range and slightly above 98 percent are in the ± 5 percent range. The GWR approach had the best predictive accuracy in the two tighter ranges but its performance was slightly inferior in the ± 8 percent range, although the differences were only marginal among the models. The SARerr model is more accurate in the ± 2 percent range than the SARmix model, but slightly lower in the ± 5 percent and ± 8 percent range. Finally, the residuals of all four models were tested for autocorrelation, measured by the Moran's I test. Besides the OLS, the GWR approach seems not to solve the autocorrelation, while there is essentially no autocorrelation observable anymore in the residuals of the SARerr and the SARmix model.

Table 5: Predictive accuracy of models: Percentage of predicted rents within specified range of actual asking rents ($\text{Ln}(\text{RENT})$) and Moran's I of the residuals.

	2% range	5% range	8% range	Moran's I
OLS	65.77	96.81	99.91	0.287
SARerr	72.65	98.02	99.93	-0.018
SARmix	72.59	98.14	99.94	-0.016
GWR	73.79	98.46	99.92	0.119

The fact that the GWR approach does not solve the autocorrelation in the residuals is alarming. Wheeler and Tiefelsdorf (2005, 186) emphasize that spatially auto-correlated residuals can produce severely correlated local regression coefficients. This is proven by a further investigation of the correlation matrix of the regression coefficients. In the correlation matrix, significant values beyond ± 0.5 can be found in almost 8 percent of the cases and values beyond ± 0.6 in 4 percent. Therefore, the issue of multicollinearity among the coefficients clearly exists in the GWR results. This raised strong doubts as to whether the GWR method is the most suitable and reliable approach here despite its good model fit. The SARmix model has some limitations as well, namely that four explanatory variables, including public transport accessibility, are insignificant. Given that the SARerr model accuracy is almost as good as the SARmix model, it is preferred here and suggested for the next update of the UrbanSim application for Zürich.

6 Conclusions

Setting up an UrbanSim application for a metropolitan area is a major task and includes, in particular, a significant data collection effort. This article focused on hedonic residential rent modeling in order to establish a housing price surface for the UrbanSim application in Zürich. The task was particularly difficult since there was no tax assessor data or data from commercial sources available for the study area at the required spatial resolution. Eventually, publicly available residential asking rents from a web-based portal were used. The variable selection was found by considering significant explanatory variables while strictly controlling for multicollinearity. Spatial autocorrelation later proved to be a problem and several spatial simultaneous autoregressive models and the GWR approach were therefore tested. Although all goodness-of-fit measures indicated slightly better performance for the GWR model compared to the spatial simultaneous autoregressive models, its residuals were still auto-correlated. Ad-

ditionally, the resulting GWR coefficients were correlated, which severely reduced confidence in the method in this study. The SARerr model has been preferred against the SARmix model because the latter had several insignificant variables, including public transport accessibility, which is a crucial measure in land use and transport modeling. Additionally, the SARerr model showed good accuracy of the predicted values compared to observed values.

Overall, the analysis highlighted the complex spatial structure of housing markets. The need to explicitly address spatial effects is obvious since a failure to do so may result in loss of explanatory power and erroneous estimates. The GWR method may not always be the best choice, although an additional benefit of GWR is to provide a means to visualize the spatial structure of housing markets which was also emphasized in earlier studies. In this study, sufficient spatial explanatory variables were available but other studies have highlighted that the GWR method is also helpful in situations where locational information is difficult to obtain or when knowledge of local submarkets is unavailable (Bitter *et al.* 2007, 24). However, locally correlated GWR coefficient estimates are a remaining problem which was present in this study as well. Methodological improvements of the GWR approach have been suggested in the literature, but are subject to an ongoing debate among econometricians. Spatial simultaneous autoregressive approaches proved to be a reasonable alternative in the analysis, which can be implemented in UrbanSim more easily because of its structure of a single set of resulting parameters.

Technically, UrbanSim is somewhat flexible concerning variable selection for the various needed models, as long as the variables are available and constantly updated. Therefore, the results of this article open discussion about further development of the model system, particularly with regard to real estate price data used in the hedonic modeling. It is of particular concern since real estate prices are also a major determinant of location choice. The hedonic modeling efforts for the Zürich application of UrbanSim are based on residential rents due to the predominant role of renting in Switzerland. Moreover, UrbanSim may not become sensitive for some explanatory variables which have been used in the models for this article. But it can be stated that besides using spatial variables, more information about the building stock, such as size, type and age (presuming data availability), would greatly improve the hedonic estimations in addition to advancing the household location choice models, perhaps even the developer model. Moreover, local rent and buying prices are not perfectly correlated. Therefore, combining rent models with transaction price models would additionally incorporate the rent/purchasing decision of households. With regard to model methodology, it again became obvious that hedonic housing models considering spatial effects are more reliable and are therefore suggested for further exploration in future applications of UrbanSim.

References

- Anglin, P. M. and R. Gençay. 1996. Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, 11(6):633–648. URL <http://www.jstor.org/stable/2285156>.
- Anselin, L. 1988. *Spatial econometrics: Methods and models*. Number 4 in Studies in Operational Regional Science. Dordrecht; Boston: Kluwer Academic Publishers.
- Anselin, L. 2001. Spatial econometrics. In B. Baltagi, ed., *A Companion to Theoretical Econometrics*, pp. 310–330. Malden: Blackwell.

- Anselin, L. 2002. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267. doi: 10.1016/S0169-5150(02)00077-4.
- Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon. 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104. doi: 10.1016/0166-0462(95)02111-6.
- Banfi, S., A. Horehejaova, and M. Filippini. 2006. Hedonic price functions for Zürich and Lugano with special focus on electrosmog. In *3rd World Congress of Environmental and Resource Economists*. Kyoto.
- Baranzini, A. and J. Ramirez. 2005. Paying for quietness: The impact of noise on Geneva rents. *Urban Studies*, 42(4):633–646.
- Baranzini, A. and C. Schaerer. 2007. A sight for sore eyes: Assessing the value of view and landscape use on the housing market. Technical report, Geneva School of Business Administration, University of Applied Sciences Western Switzerland, Geneva. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=981189.
- Baranzini, A., C. Schaerer, J. V. Ramirez, and P. Thalmann. 2006. Feel it or measure it – perceived vs. measured noise in hedonic models. Technical report, Geneva School of Business Administration, University of Applied Sciences Western Switzerland, Geneva. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=937259.
- Bitter, C., G. F. Mulligan, and S. Dall’erba. 2007. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, 9(1):7–27.
- Bivand, R. S. 2008. *Applied Spatial Data Analysis with R*. New York, NY: Springer Science+Business Media, LLC.
- Bourassa, S. and M. Hoesli. 2010. Why do the Swiss rent? *The Journal of Real Estate Finance and Economics*, 40(3):286–309. doi: 10.1007/s11146-008-9140-4.
- Brunauer, W., S. Lang, P. Wechselberger, and S. Bienert. 2008. Additive hedonic regression models with spatial scaling factors: An application for rents in Vienna. Working Papers 2008-17, Faculty of Economics and Statistics, University of Innsbruck. URL <http://ideas.repec.org/p/inn/wpaper/2008-17.html>.
- Brunauer, W., S. Lang, P. Wechselberger, and S. Bienert. 2009. Additive hedonic regression models with spatial scaling factors: An application for rents in Vienna. *Journal of Real Estate Finance and Economics*. doi: 10.1007/s11146-009-9177-z.
- Brunsdon, C., S. Fotheringham, and M. Charlton. 1998. Geographically weighted regression – modelling spatial non-stationarity. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431. ISSN 00390526.
- Brunsdon, C., J. McClatchey, and D. Unwin. 2001. Spatial variations in the average rainfall-altitude relationship in Great Britain: an approach using geographically weighted regression. *International Journal of Climatology*, 21(4):455–466. doi: 10.1002/joc.614.
- Butler, R. V. 1980. The specification of hedonic indexes for urban housing. *Land Economics*, 58(1):96.
- Calvo, E. and M. Escobar. 2003. The local voter: A geographically weighted approach to ecological inference. *American Journal of Political Science*, 47(1):189.
- Chatfield, C. 1995. *Problem Solving: A Statistician’s Guide*. Chapman & Hall texts in statistical science. London; New York: Chapman & Hall.

- Chow, L.-F., F. Zhao, X. Liu, M.-T. Li, and I. Ubaka. 2006. Transit ridership model based on geographically weighted regression. *Transportation Research Record*, 1972(1):105–114. doi: 10.3141/1972-15.
- Clapp, J. M. 2003. A semiparametric method for valuing residential locations: Application to automated valuation. *The Journal of Real Estate Finance and Economics*, 27(3):303–320. doi: 10.1023/A:1025838007297.
- Clapp, J. M. 2004. A semiparametric method for estimating local house price indices. *Real Estate Economics*, 32(1):127–160.
- Clark, S. D. 2007. Estimating local car ownership models. *Journal of Transport Geography*, 15(3):184–197. doi: 10.1016/j.jtrangeo.2006.02.014.
- Cropper, M. L., L. B. Deck, and K. E. McConnell. 1988. On the choice of functional form for hedonic price functions. *Review of Economics & Statistics*, 70(4):668.
- de Graaff, T., R. J. Florax, P. Nijkamp, and A. Reggiani. 2001. A general misspecification test for spatial regression models: Dependence, heterogeneity, and nonlinearity. *Journal of Regional Science*, 41(2).
- Djurdjevic, D., C. Eugster, and R. Haase. 2008. Estimation of hedonic models using a multilevel approach: An application for the Swiss rental market. *Swiss Journal of Economics and Statistics (SJES)*, 144(4):679–701. URL <http://econpapers.repec.org/RePEc:ses:arsjes:2008-iv-7>.
- Du, H. and C. Mulley. 2006. Relationship between transport accessibility and land value: Local model approach with geographically weighted regression. *Transportation Research Record*, 1977(1):197–205. doi: 10.3141/1977-25.
- Fahrländer, S. S. 2005. Semiparametric construction of spatial generalized hedonic models for private properties. Diskussionschriften dp0507, Universität Bern, Departement Volkswirtschaft. URL <http://ideas.repec.org/p/ube/dpvwib/dp0507.html>.
- Farber, S. and M. Yeates. 2006. A comparison of localized regression models in a hedonic house price context.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2000. *Quantitative geography: Perspectives on spatial data analysis*. London; Thousand Oaks, Calif.: Sage Publications.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression the analysis of spatially varying relationships*. Chichester, West Sussex; Hoboken, NJ: Wiley.
- Fotheringham, A. S., M. E. Charlton, and C. Brunsdon. 2001. Spatial variations in school performance: A local analysis using geographically weighted regression. *Geographical & Environmental Modelling*, 5(1):43–66.
- Fotheringham, A. S. and T. C. Pitts. 1995. Directional variation in distance decay. *Environment and Planning A*, 27(5):715–729. doi: 10.1068/a270715.
- Fox, J. 2002. *An R and S-Plus companion to applied regression*. Thousand Oaks, Calif.: Sage Publications.
- Freeman, A. M. 1993. *The measurement of environmental and resource values: Theory and methods*. Washington, D.C.: Resources for the Future.
- Gao, X., Y. Asami, and C.-J. F. Chung. 2006. An empirical evaluation of spatial regression models. *Computers & Geosciences*, 32(8):1040–1051. doi: 10.1016/j.cageo.2006.02.010. Spatial Modeling for Environmental and Hazard Management.
- Gelfand, A. E., H.-J. Kim, C. F. Sirmans, and S. Banerjee. 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*,

- 98(462):387–396.
- Goldstein, H. 1987. *Multilevel models in educational and social research*. London; New York: Oxford University Press.
- Griffith, D. A. 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, 40(11):2751–2769. doi: [10.1068/a38218](https://doi.org/10.1068/a38218).
- Griliches, Z. 1971. *Price indexes and quality change: Studies in new methods of measurement*. Cambridge, Mass.: Harvard University Press.
- Haas, T. C. 1995. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association*, 90(432).
- Haas, T. C. 1996. Multivariate spatial prediction in the presence of non-linear trend and covariance non-stationarity. *Environmetrics*, 7(2):145–165. doi: [10.1002/\(SICI\)1099-095X\(199603\)7:2<145::AID-ENV200>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-095X(199603)7:2<145::AID-ENV200>3.0.CO;2-T).
- Hadayeghi, A., A. Shalaby, and B. Persaud. 2003. Macrolevel accident prediction models for evaluating safety of urban transportation systems. *Transportation Research Record*, 1840(1):87–95. doi: [10.3141/1840-10](https://doi.org/10.3141/1840-10).
- Halvorsen, R. and H. Pollakowski. 1981. Choice of functional form for hedonic price equations. *Journal of Urban Economics*, 10(1):37–49.
- Huang, Y. and Y. Leung. 2002. Analysing regional industrialisation in Jiangsu province using geographically weighted regression. *Journal of Geographical Systems*, 4(2):233. ISSN 14355930.
- Jones, K. 1991. Specifying and estimating multi-level models for geographical research. *Transactions of the Institute of British Geographers*, 16(2):148–159. ISSN 00202754. URL <http://www.jstor.org/stable/622610>.
- Keljian, H. and D. Robinson. 2000. The influence of spatially correlated heteroskedasticity on tests for spatial correlation. In L. Anselin and R. Florax, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*, pp. 79–97. Berlin: Springer.
- Kestens, Y., M. Thériault, and F. Des Rosiers. 2006. Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles. *Journal of Geographical Systems*, 8(1):61–96.
- Kim, C. W., T. T. Phipps, and L. Anselin. 2003. Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management*, 45(1):24–39. doi: [10.1016/S0095-0696\(02\)00013-X](https://doi.org/10.1016/S0095-0696(02)00013-X).
- Kimsey, M. J., J. Moore, and P. McDaniel. 2008. A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. *Forest Science*, 54(3):356–366.
- Kissling, W. D. and G. Carl. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology & Biogeography*, 17(1):59–71.
- Kriström, B. 2008. Applying hedonics in the housing market: An illustration. In A. Baranzini, J. Ramirez, C. Schaerer, and P. Thalmann, eds., *Hedonic Methods in Housing Markets*, pp. 247–263. New York: Springer.
- Lancaster, K. 1971. *Consumer demand: A new approach*. New York: Columbia University Press.
- Lancaster, K. J. 1966. A new approach to consumer theory. *Journal of Political Economy*, 74(2):132.

- Leung, Y., C.-L. Mei, and W.-X. Zhang. 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, 32(1):9–32. doi: 10.1068/a3162.
- Lloyd, C. and I. Shuttleworth. 2005. Analysing commuting using local regression techniques: Scale, sensitivity, and geographical patterning. *Environment and Planning A*, 37(1):81–103. doi: 10.1068/a36116.
- Löchl, M. 2006. Real estate and land price models for the greater Zürich application of UrbanSim. *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, 6, NSL, ETH Zürich. URL <http://www.ivt.ethz.ch/vpl/publications/reports/ab403.pdf>.
- Löchl, M. 2007. Considering spatial dependence in hedonic rent price regression. Paper presented at the Seventh Swiss Transport Research Conference, Ascona, September 2007. URL <http://www.ivt.ethz.ch/vpl/publications/reports/ab469.pdf>.
- Löchl, M., M. Bürgle, and K. Axhausen. 2007. Implementierung des integrierten Flächen-nutzungsmodells UrbanSim für den Grossraum Zürich – ein Erfahrungsbericht. *DISP*, 168:13–25. URL <http://www.ivt.ethz.ch/vpl/publications/reports/ab414.pdf>.
- Malpezzi, S. 2003. Hedonic pricing models: A selective and applied review. In T. O’Sullivan, K. Gibb, D. MacLennan, and RICS Foundation, eds., *Housing economics and public policy: Essays in honour of Duncan MacLennan*, pp. 67–74. Oxford, UK; Malden, MA, USA: Blackwell Science.
- Martins-Filho, C. and O. Bin. 2005. Estimation of hedonic price functions via additive non-parametric regression. *Empirical Economics*, 30(1):93–114. doi: 10.1007/s00181-004-0224-6.
- Mittal, V., W. A. Kamakura, and R. Govind. 2004. Geographic patterns in customer service and satisfaction: An empirical investigation. *Journal of Marketing*, 68(3):48–62.
- Mosteller, F. and J. Tukey. 1977. *Data Analysis and Regression – A Second Course in Statistics*. Reading: Addison-Wesley.
- Nakaya, T. 2001. Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal*, 53(4):347–358. doi: 10.1023/A:1020149315435.
- Pace, R. K. 1993. Nonparametric methods with applications to hedonic models. *The Journal of Real Estate Finance and Economics*, 7(3):185–204. doi: 10.1007/BF01096965.
- Pace, R. K. 1998. Appraisal using generalized additive models. *Journal of Real Estate Research*, 15(1/2):77.
- Parmeter, C. F., D. J. Henderson, and S. C. Kumbhakar. 2007. Nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, 22(3):695–699. URL <http://www.jstor.org/stable/25146538>.
- Páez, A. 2005. Local analysis of spatial relationships: A comparison of GWR and the expansion method. In *Computational Science and Its Applications – ICCSA 2005*, pp. 162–172. Springer. doi: 10.1007/11424857_18.
- Páez, A., F. Long, and S. Farber. 2007. Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Studies*, 45(8):1565–1581. doi: 10.1177/0042098008091491.
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Rosen, S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34.

- Shearmur, R., P. Apparicio, P. Lizion, and M. Polèse. 2007. Space, time, and local employment growth: An application of spatial regression analysis. *Growth and Change*, 38(4):696–722. doi: 10.1111/j.1468-2257.2007.00393.x.
- Shin, K., S. Washington, and K. Choi. 2007. Effects of transportation accessibility on residential property values: Application of spatial hedonic price model in Seoul, South Korea metropolitan area. *Transportation Research Record*, 1994(1):66–73. doi: 10.3141/1994-09.
- Sirmans, G. S., D. A. Macpherson, and E. N. Zietz. 2005. The composition of hedonic pricing models. *Journal of Real Estate Literature*, 13(1):3–43.
- Sirmans, G. S., C. F. Sirmans, and J. D. Benjamin. 1989. Determining apartment rent: The value of amenities, services and external factors. *Journal of Real Estate Research*, 4(2):33.
- Statistical Office of Canton Zürich. 2008. Mikrozensus 2005: Kilometerleistung. URL http://www.statistik.zh.ch/themenportal/themen/daten_detail.php?id=722&tb=&mt=2.
- Taylor, L. 2008. Theoretical foundations and empirical developments in hedonic modeling. In A. Baranzini, J. Ramirez, C. Schaerer, and P. Thalmann, eds., *Hedonic methods in housing markets pricing environmental amenities and segregation*, pp. 15–37. New York: Springer.
- Trigg, D. W. and A. G. Leach. 1967. Exponential smoothing with an adaptive response rate. *Operational Research Quarterly*, 18(1):53–59. URL <http://www.jstor.org/stable/3010768>.
- Valente, J., W. ShanShan, A. Gelfand, and C. F. Sirmans. 2005. Apartment rent prediction using spatial modeling. *Journal of Real Estate Research*, 27(1):105–136.
- Waddell, P. and G. F. Ulfarsson. 2003. Dynamic simulation of real estate development and land prices within an integrated land use and transportation system. Paper presented at the 83th Annual Meeting of the Transportation Research Board, Washington, D.C., January.
- Wall, M. M. 2004. A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference*, 121(2):311–324. doi: 10.1016/S0378-3758(03)00111-3.
- Waller, L., L. Zhu, C. Gotway, D. Gorman, and P. Gruenwald. 2007. Quantifying geographic variations in association between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment*, 21(5):573–588. doi: 10.1007/s00477-007-0139-9.
- Wang, N., C.-L. Mei, and X.-D. Yan. 2008. Local linear estimation of spatially varying coefficient models: An improvement on the geographically weighted regression technique. *Environment and Planning A*, 40(4):986–1005. doi: 10.1068/a3941.
- Wheeler, D. and M. Tiefelsdorf. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2):161–187.
- Wheeler, D. C. 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A*, 39(10):2464–2481. doi: 10.1068/a38325.
- Wheeler, D. C. 2009. Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environment and Planning A*, 41(3):722–742. doi: 10.1068/a40256.
- Wheeler, D. C. and L. A. Waller. 2009. Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems*, 11(1):1–22.

- Widrow, G. and M. Hoff. 1960. Adaptive switching circuits. Institute of Radio Engineers Western Electric Show and Convention Record, Part 4.
- Wilhelmsson, M. 2002. Spatial models in real estate economics. *Housing, Theory & Society*, 19(2):92–101.
- Wüest & Partner. 2007. *Immo-Monitoring 2008*. Wüest & Partner.
- Young, L., C. A. Gotway, J. Yang, G. Kearny, and C. DuClos. 2008. Assessing the association between environmental impacts and health outcomes: A case study from Florida. *Statistics in Medicine*, 27(20):3998–4015. doi: [10.1002/sim.3249](https://doi.org/10.1002/sim.3249).
- Yu, D., Y. D. Wei, and C. Wu. 2007. Modeling spatial dimensions of housing prices in Milwaukee, WI. *Environment and Planning B: Planning and Design*, 34(6):1085–1102. doi: [10.1068/b32119](https://doi.org/10.1068/b32119).
- Zhang, L. and H. Shi. 2004. Local modeling of tree growth by geographically weighted regression. *Forest Science*, 50(2):225–244.

Appendix: Additional variable data

Table 6: Estimated OLS parameters for models with spatial explanatory variables only.

Measure	Model 3		Model 4	
	Value	Sig.	Value	Sig.
(Constant)	9.515	***	8.366	***
Ln(CARTT_CBD)	-0.258	***		
CARACC05			0.064	***
PTACC05	-0.012	***	-0.004	
Ln(RAILSTATION)	-0.019	***	-0.015	**
AUTOBAHN	-0.061	**	-0.064	**
AIRNOISE	-0.009		-0.053	***
Ln(HOTREST_JOBS)	0.000		0.020	***
Ln(POP_DENS)	-0.093	***	-0.091	***
FOREIGNERS (logit)	-0.087	***	-0.091	***
Ln(TAXLEVEL)	-0.289	***	-0.368	***
SLOPE (logit)	0.018	***	0.027	***
Ln(VIEW_LAKE)	0.015	***	0.018	***
Ln(VIEW_ALL)	0.007		0.005	
Ln(SOLAR_EVE)	0.071	***	0.056	***
Adjusted <i>R</i> -square	0.235		0.215	
<i>F</i> -test	208.800	***	182.200	***
Moran's <i>I</i>	0.165	***	0.185	***

Probability of rejecting H_0 : *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$