Veie, Kathrine Lausted; Panduro, Toke Emil

Working Paper

An alternative to the standard spatial econometric approaches in hedonic house price models

IFRO Working Paper, No. 2013/18

Provided in Cooperation with:
Department of Food and Resource Economics (IFRO), University of Copenhagen

Suggested Citation: Veie, Kathrine Lausted; Panduro, Toke Emil (2013) : An alternative to the standard spatial econometric approaches in hedonic house price models, IFRO Working Paper, No. 2013/18, University of Copenhagen, Department of Food and Resource Economics (IFRO), Copenhagen

This Version is available at:
https://hdl.handle.net/10419/204365

DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

An alternative to the standard spatial econometric approaches in hedonic house price models

Kathrine Lausted Veie
Toke Emil Panduro

2013 / 18

IFRO Working Paper 2013 / 18

An alternative to the standard spatial econometric approaches in hedonic house price models

Authors: Kathrine Lausted Veie, Toke Emil Panduro

www.ifro.ku.dk/english/publications/foi_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk

An alternative to the standard spatial econometric approaches in hedonic house price models

Kathrine Lausted Veie[*] *and Toke Emil Panduro*[†]

Abstract

Hedonic models are subject to spatially correlated errors which are a symptom of omitted spatial variables, misspecification or mismeasurement. Methods have been developed to address this problem through the use of spatial econometrics or spatial fixed effects. However, often spatial correlation is modeled without much consideration of the theoretical implications of the chosen model or treated as a nuisance to be dealt with holding little interest of its own. We discuss the limitations of current standard spatial approaches and demonstrate, both empirically and theoretically the generalized additive model as an alternative. The generalized additive model is compared with the spatial error model and the fixed effects model. We find the generalized additive model to be a solid alternative to the standard approaches, having less restrictive assumptions about the omitted spatial processes while still being able to reduce the problem of spatial autocorrelation and provide trustworthy estimates of spatial variables. However, challenges connected with spatially varying data remain. The choice of flexibility in the spatial structure of the model affects estimated parameters of some spatially varying characteristics markedly. This suggests that omitted variable bias may remain an important problem. We advocate for an increased use of sensitivity analysis to determine robustness of estimates to different models of the (omitted) spatial processes.

1 Introduction

The hedonic model as described by Rosen (1974) remains popular in the environmental valuation literature. One major concern in hedonic estimation is the issue of omitted variable bias and spatially correlated errors. Spatial correlation in the error term is a common finding in applied hedonic analysis and can be caused by misspecification of spatially delineated variables, systematic mismeasurement of the spatial regressors or spatial covariates omitted from the model. The existence of spatially correlated errors can result in inconsistent and inefficient parameter estimates depending on the underlying cause (Anselin, 2010). Recent years have seen many improvements in the way such omitted variables are addressed. However, in a special issue of the Journal of Regional Science, Gibbons and Overman (2012) and McMillen (2012) criticize the "automatic" use of spatial lag and spatial error components in econometric modeling. Brady and Irwin (2011) discuss developments in land use modeling in the light of this recent critique. We focus our attention on hedonic house price models and the challenge of handling the spatial dimension in these models. In addition we present an alternative approach to modeling the spatial dimension of the housing market in the form of a spatial generalized additive model.

In the current hedonic literature, attempts to reduce spatial correlation in the error term are made using different methods with different interpretations of the character spatial correlation. To the best of

[*]kave@life.ku.dk, University of Copenhagen
[†]tepp@life.ku.dk, University of Copenhagen

our knowledge there are no surveys which explicitly contain a critical review of the use of spatial methods in hedonic house price modeling. However, Kuminoff et al. (2010) discuss the choice of functional form for hedonic estimation in the light of the increased use of fixed effects or spatial econometrics which aims to reduce omitted variable bias. Looking at more than 60 published papers, they find that more than half of the hedonic studies apply either spatial fixed effects or spatial econometric models to address omitted spatial variables. We looked at 21 hedonic studies published since 2010 in the Journal of Environmental Economics and Management, Land Economics, Ecological Economics and Environmental and Resource Economics. Approximately half of these studies used either a spatial error term or a spatial lag term to control for spatial correlation while the other half used fixed effects and differences-in-differences. As such it is clear that the use of these methods is extensive in current hedonic research. The specification of spatial econometric models varies across studies both with regard to the chosen model (spatial lag, spatial error or both) and the design of the spatial weight matrix (inverse distance weighting, contiguity etc.). Similarly, fixed effects are generally based on available spatial entities such as provinces, census blocks etc. Only two of the papers (Heintzelman and Tuttle (2012), Chamblee et al. (2011)) contained in-depth discussion of the choice of spatial controls and explicitly discuss sensitivity analysis of different spatial specifications.

The aim of this paper is to address the strengths and weaknesses of the standard econometric strategies to handle spatial autocorrelation applied in the literature on hedonic house price valuation. On the basis of this discussion, we suggest an alternative spatial model, a spatial Generalized Additive Model (GAM), which handles omitted spatial processes non-parametrically. To our knowledge the use of a GAM for explicit spatial modeling is novel to the peer reviewed hedonic house price literature. The only similar application of a GAM that we are aware of is in a book chapter by Geniaux and Napoleone (2008). However, their emphasis is different as they focus on comparison with a geographically weighted regression. They do not consider parametric spatial models or fixed effects both of which are more commonly used in the hedonic literature. The present paper is divided into two parts. In the first part we provide a general discussion of the standard econometric approaches and introduce the GAM model as a strong alternative. In the second part we illustrate the discussion using an empirical application. We estimate the hedonic price function using a simple linear model with no spatial corrections, a Spatial Error Model (SEM), a spatial fixed effect model and a GAM model. We then vary the choice of weight matrix and the number of basis functions to evaluate the sensitivity of our results to the scale of the spatial control.

Our main critique of the fixed effect and spatial econometric approaches is that the nature of the omitted spatial processes is assumed to be known when in fact it is not. In practice there are likely to be several omitted spatial processes at different spatial scales in the housing market. In the fixed effect model a geographical entity such as a school district or another spatial subdivision is assumed to capture the unknown omitted spatial processes. In the standard spatial econometric models, e.g. spatial error and spatial lag models, the spatial processes are assumed to be captured by a spatial weight matrix which is often based on the 10 or 20 nearest neighbors. We argue that it is inappropriate to place such restrictions on the omitted spatial processes. In contrast, the GAM does not require any assumptions about the structure of the omitted spatial processes in the housing market. The spatial processes are handled by letting the geographical coordinates of each property enter into the model through a smooth function based on thin plate splines. In practice, the spatial GAM can be understood as a smooth fixed effect as opposed to the standard discrete version of the fixed effect model.

The most commonly used parametric spatial econometric approaches have some additional drawbacks. The SEM is based on the assumption that the correlation in the error term is a result of processes which are

not correlated with the variables in the model. In that setting, the correlation in residuals would only lead to inefficient estimates and biased standard errors, which the SEM corrects for. In contrast, the GAM and the fixed effect model treat the omitted spatial processes as an additional regressor. The model specification in a spatial lag model implies the existence of a spatial multiplier effect of the marginal values of the model (Lesage and Pace, 2009). Such an effect seems unjustified in hedonic house price models when interpreting the marginal change in prices as a measure of an individual household's willingness to pay.

When comparing the spatial GAM with alternative approaches and varying the dimension of spatial modeling, we find that estimated coefficients of spatially varying regressors can be highly sensitive to the way in which the spatial dimension is modeled. The results in our example are largely robust for the structural housing characteristics. However, parameter estimates of several spatially varying regressors turn out to be sensitive to the level of spatial correction and the model estimated. We hypothesize that this sensitivity of parameters to different approaches to handling omitted spatial processes indicates that substantial risk of omitted variable bias in the hedonic model remains despite the use of spatial models. The severity of potential bias can be assessed by carrying out sensitivity analysis with different spatial models. We conclude that spatial sensitivity analysis should be a part of every hedonic study with spatially varying regressors.

2 Modeling the value of a residence

Housing is a composite good and its price reflects its composition. A house in a more attractive location with many amenities tends to be more expensive than a house with fewer amenities all else equal. For this reason, the transactions in the housing market can be used to place a value on many amenities not traded independently in the market, see Palmquist (2004) for an introduction to the literature. The hedonic method is based on analyzing the equilibrium price in the housing market under the assumption that a continuum of housing types exists. The equilibrium price schedule is determined by the structure of preferences and technologies in the market and can be expressed as a function of the attributes of the house[1]:

$$P = f(X) \tag{2.1}$$

Housing attributes, X, usually include structural variables such as the number of rooms, the size of the living area and the time of construction, as well as accessibility measures such as distance to the central business district or distance to train stations or highway access. The housing good further has different environmental attributes such as traffic noise exposure (Day et al. (2007)), air quality (Chay and Greenstone (2005)), green space (Abbott and Klaiber (2010a)) as well as neighborhood characteristics such as school quality, crime levels etc. Household utility is a function of consumption of housing, $H(X)$ and other goods, C:

$$U = g(H(X), C) \tag{2.2}$$

The household chooses a quantity of housing attribute, X, to maximize utility subject to a budget constraint: $M = P(X) + C$, where non-housing consumption, C, is the numeraire. Maximizing utility delivers the following first order condition:

[1] This is the first stage of the hedonic method, where an equilibrium house price schedule is estimated to reveal a household's marginal willingness to pay for a characteristic. The second stage of the hedonic analysis, where household preferences are recovered, is not discussed in this paper.

$$\frac{\frac{\partial U}{\partial X_j}}{\frac{\partial U}{\partial C}} = \frac{\partial P}{\partial X_j} \tag{2.3}$$

Utility maximization implies that the change in utility resulting from a marginal increase in X_j exactly equals the change in the house price following the same increase in X_j all else equal[2]. Based on equation 2.1, the marginal price of a housing attribute is defined as the partial derivative of the house price with respect to the characteristic:

$$\frac{\partial P}{\partial X_j} = f_j(x) \tag{2.4}$$

The main object of most published hedonic analyses is the recovery of (average) marginal prices sometimes also referred to as implicit prices. Only a few studies proceed to recover household preferences, as this is subject to several econometric challenges, see Epple (1987), Kahn and Lang (1988) and Ekeland et al. (2004) for further discussion. The marginal prices can be used to valuate marginal changes in e.g. environmental amenity levels. For larger changes in amenities, the marginal price can be a poor estimate of the change in welfare.

2.1 Applying the method

The unit of analysis in most hedonic studies is the individual transacted dwelling. With the use of Geographical Information Systems, researchers have access to ample data on the location and surroundings of dwellings. However, it remains close to impossible to measure every characteristic of a home and a neighborhood. Location is often described using proximity measures or similar proxies for accessibility to amenities. Such proxies may not always be an accurate reflection of the household's perception of the amenity. Further, it is not clear how different attributes should enter the hedonic price function. Theory gives little guidance as to the shape of the hedonic price schedule is determined by the preference parameters and technology parameters together, see Ekeland et al. (2004) for an excellent discussion. Environmental amenity access, general accessibility and neighborhood characteristics all vary with location. Variables measuring these housing attributes therefore tend to be spatially delineated, which in turn implies that they tend to be highly correlated across observations close to each other in space. Misspecification or mismeasurement of such spatially delineated variables can result in spatial autocorrelation in the model residuals, and potentially, biased parameter estimates (Anselin and Lozano-Gracia, 2008). The same results if a spatial regressor is omitted from the analysis or measured with (systematic) error. Take the hedonic model written below, where − for simplicity − two spatially delineated attributes X_1 and X_2 determine the price of the house and ϵ is a random i.i.d. error term:

$$P = X_1\beta_1 + X_2\beta_2 + \epsilon \tag{2.5}$$

Suppose that X_1 and X_2 are uncorrelated and linear related to P, and X_2 is wrongly specified with a log transformation. When estimating the hedonic equation, we will find $P = X_1\hat{\beta}_1 + ln(X_2)\hat{\beta}_2 + \hat{u}$, where $\hat{u} = \epsilon + X_2\beta_2 - ln(X_2)\hat{\beta}_2$, which will vary across space as X_2 varies across space. If X_2 is omitted from the analysis, the error will consist of $\hat{u} = X_2\beta_2 + \epsilon$. Finally, if X_2 has been mismeasured, so the model is estimated with $\tilde{X_2} = X_2 + e$, then the error term consists of $\hat{u} = \epsilon - e\hat{\beta}_2$. The estimated coefficient $\hat{\beta}_2$ will be

[2] It is an underlying assumption in hedonic theory that the housing market contains a continuum of houses, such that any combination of attributes can be found. When this is not the case, the equality may not hold as households may have preferred to have more or less of an attribute at the given "marginal price" than was available.

biased towards zero, if the error is uncorrelated with true X_2. In that case, there may be no resulting spatial autocorrelation. However, if the measurement error is also correlated with X_2, the bias will depend on the sign and size of the correlation, and in addition, the error term $\hat{u}$ will be spatially correlated. Additionally, measurement error can be inherently spatial, e.g. if it arises from interpolation of variables only measured at discrete points (e.g. air pollution data). If the data is interpolated without attention to, e.g. wind direction or barriers in the landscape, the result would be a spatially correlated measurement error. We refer to all three causes of spatial correlation as "omitted spatial processes".

The preceding discussion assumes that X_1 and X_2 are uncorrelated. However, spatially delineated regressors are often correlated. Homes near the central business district tend to be far from natural areas or agricultural fields, while industry is often located near waterways or other infrastructure. Through the development of the urban landscape, locational attributes will therefore tend to correlate with each other to varying degrees. When this is the case, it is not unlikely that mismeasurement, misspecification or omission of a spatially delineated variable can bias parameter estimates of other spatially varying regressors in the model as well as the parameter estimate of the affected variable. In practice, there are likely to be several omitted spatial processes in a hedonic data set which vary at different scales. As environmental amenities tend to be inherently spatial in nature, omitted spatial processes are particularly troubling in the hedonic valuation literature. Our main focus here lies on methods intended to aid in the recovery of robust marginal prices for spatially delineated attributes. In the following, we discuss the techniques most commonly used in the literature.

2.2 Spatial econometrics

In spatial econometrics spatial relationships are modeled parametrically through the use of weight matrices identifying the relevant neighboring observations. There are several different varieties with the most common being the SEM, the spatial lag model, and the combined model containing both spatial error and spatial lag processes. These spatial econometric models impose strong assumptions about the structure of spatial correlation in the data. A general spatial autoregressive model with spatial autoregressive errors (SARAR(1,1)) for the price of a house, P, is given by:

$$P = X\beta + \rho W_p P + u \tag{2.6}$$

where

$$u = \lambda W_u u + \epsilon \tag{2.7}$$

The matrices W_p and W_u are usually referred to as the spatial weight matrices and often coincide in applications. The diagonal elements are all zero and off-diagonal elements can be ones (contiguity indicators) or a function of distance between observations. In practice, the weight matrices are row standardized, so the term $\rho W_p P$ corresponds to a weighted average price of neighboring observations. The parameters λ and ρ are commonly known as the autocorrelation coefficients. Intuitively, the autocorrelation coefficients will be positive in most house price analyses reflecting clustering in high and low value (residential) areas. In the SEM, $\rho = 0$, and in the spatial lag model, $\lambda = 0$, so that the errors, u, are independent and identically distributed. The researcher usually chooses the relevant dimension of the weight matrix, i.e. how many neighbors to include or which distance boundary to set for neighborhood effects. This choice is made either based on tests, e.g. Moran's I test for spatial correlation, or justified by referring to the existing literature.

A notable exception is the SEM in Hoshino and Kuriyama (2010) where the relevant distance is estimated.

Autocorrelation in the error term is sometimes considered to affect inference through incorrect standard errors. When errors are positively correlated across space, failure to account for this correlation can lead to overestimating significance levels. Standard errors of estimated parameters tend to be underestimated in ordinary least squares estimation when errors are clustered. If spatial correlation derives from omitted spatial components correlated with regressors included in the model, the estimated parameters will additionally be biased. Spatially delineated amenities tend to be correlated with each other and hence also with omitted or misspecified spatial characteristics. This is due to the fact that location typically varies only in two dimensions (longitude, latitude). Therefore the conditions under which the SEM is valid are not likely to be satisfied in many housing market applications. Essentially, the SEM is similar to the use of random effects in panel data estimation or feasible generalized least squares. McMillen (2012) emphasizes that the SEM is also a form of spatial smoother, where the number of neighbors plays a role similar to the choice of bandwidth in terms of kernel smoothing or basis dimension in terms of splines.[3]

The spatial lag model in turn implies that there exist direct spillover effects between house prices of neighboring properties. Lesage and Pace (2009) give some interpretations of the spatial lag model, not all of which are consistent with the equilibrium assumption underlying the hedonic approach. In particular, such spillovers could describe the process of neighborhood gentrification in which wealthier households move in and in doing so change the composition of the neighborhood which leads to higher prices and so on.[4] It seems unlikely that the hedonic price function should remain the same in a new equilibrium if the composition of a neighborhood changes. Alternatively, the spillover can be interpreted as an information effect. If sellers and buyers are unsure of the appropriate value of a property given its characteristics, they may infer the appropriate price from looking at nearby properties with similar characteristics which have been sold recently. The information contained in previous transactions in the same area may also allow the household to form expectations about the future evolution of the prices in the area. For each of these interpretations, however, it is clear, there should be a subscript t indicating that the spillover effect occurs from recently sold properties to future sales and not vice versa. In most applications of the spatial lag model, that distinction is not made.

Gibbons and Overman (2012) emphasize the need to think of the theoretical context before specifying the model rather than choosing a model based on statistical tests. This becomes especially important as the spatial lag model implies the existence of a spatial multiplier on marginal effects which leads directly to higher marginal prices. Lesage and Pace (2009) distinguish between average direct, indirect and total impacts, depending on whether one looks solely at the estimated coefficient or accounts for neighboring observations. A similar interpretation is given in Won Kim et al. (2003), where the marginal price of a housing characteristic (total impact) becomes:

$$\frac{\partial P}{\partial x} = \beta \left(I - \rho W_p\right)^{-1} \tag{2.8}$$

When spatial correlation is positive and large, this multiplier can be quite significant. Small and Steimetz (2006) argue that the multiplier should only be applied when the spillover is technological, but not when it is purely informational. It is very hard empirically to distinguish between these interpretations as the model does not identify the source of the spillover. It seems unintuitive that the addition of e.g. an additional square meter of living area to a house should have value for all the neighbors. Nor does it seem intuitive

[3] A spline is a combination of a series of basis functions over covariate space. Basis functions can consist of e.g. polynomials of increasing order. A higher number of basis functions translates into more flexibility in the functional form.

[4] "We need to keep in mind that the scalar summary measures of impact reflect how these changes would work through the simultaneous dependence system over time to culminate in a new steady state equilibrium." (LeSage & Pace (2009), p. 37)

that this value including spillovers should correspond to the individual household's willingness to pay for the improvement, which is essentially what the model implies in a hedonic context. Several alternatives are available to address spatial correlation, which are more in tune with the underlying hedonic model.

A reason for the common use of the spatial lag model is that model estimations are likely to find the autoregressive parameter to be significantly different from zero. This finding may just as well be due to omitted variable bias and in any case does not imply that the chosen model correctly captures the spatial processes, see also McMillen (2012). From equation 2.6, it is clear that the elements in Wy and u will be correlated and an IV approach is used for consistent estimation, e.g. Kelejian and Prucha (2010). The instruments are constructed based on spatially lagged exogenous variables. For housing market applications the instruments would be a weighted combination of the characteristics of nearby properties. If prices are high in an area, the households living there tend to be wealthy, and the average wealth of a neighborhood will be correlated with crime rates, school quality and the neighborhoods general appearance as well as the size and style of a home. The lagged dependent variable is likely to be a proxy for these often unobserved characteristics. The instruments used to identify the spatial lag model may not be redundant in the model in the first place. In other words, estimating a significant spatial correlation coefficient does not imply that the model is a good approximation to the truth.

2.3 Fixed effects

Spatial fixed effects are quite popular as a control for omitted variables. Basically, fixed effect estimation corresponds to including a dummy variable for belonging to a geographical entity in the data. In that sense, fixed effects are similar to the use of spatial weight matrices with contiguity indicators, except the matrix is not centered on each observation but rather identifies observations belonging to the same spatial entities. With fixed effects, each observation belongs to only one neighborhood entity, whereas an observation can belong to several neighborhoods as these are defined by the weight matrix in the spatial econometric models. Remaining spatial correlation in the error term can be accounted for by clustering errors within entities to avoid overestimating significance levels. Fixed effects are easy to implement and can be more flexible than the parametric models in capturing the unknown urban structure depending on their level of aggregation. The flexibility comes at the cost of a loss of degrees of freedom when many fixed effects are included. Essentially, where the standard spatial econometric model estimates one or two spatial lag parameters, the fixed effect model estimates additional parameters corresponding to the number of entities in the data set. As data availability increases, sample sizes have also grown to make this constraint less binding.

Fixed effects imply discrete shifts in the level of house prices as one moves across the border of the entity used to establish the panel structure. The effect of omitted variables is constrained to be constant within the entity and only vary between entities. Just as the specification of the weight matrix determines the relevant neighborhood size, the fixed effect should coincide with the level at which the omitted variables vary in order to be effective. Unfortunately, the nature of most omitted processes is unobservable to the econometrician. In existing studies, fixed effects are usually created based on availability of, e.g. administrative units such as provinces in Brounen and Kok (2011), counties in Deaton and Vyn (2010) or municipalities as in Cavailhes et al. (2009), or they are created from the object of interest, i.e. beaches in Gopalakrishnan et al. (2011) or lakes in Walsh et al. (2011). The use of small entities is preferable to the use of larger ones with respect to controlling appropriately for omitted variable bias.

Fixed effects based on small entities demand a lot of variation in the data. There must be sufficient spatial or temporal variation in the data within the spatial entity to distinguish the variable of interest from

the fixed effect. In the extreme case of a repeat sales model, only the effect of variables which change over time can be identified. In cross-sections, using census block or school district fixed effects can make it very difficult to recover impacts of amenities such as air pollution or airport noise, which vary little across space. Any effect they might have on housing prices is likely to be "sucked up" in the fixed effect.[5] For proximity measures as well, the use of fixed effects can make estimation of parameters for such amenities as park access difficult. The variation in proximity to the nearest park within a spatial entity declines, the smaller the entity is. Proximity measures of spatially delineated amenities, which are scarce in the landscape, tend to vary less across observations than proximity to spatially delineated amenities found at several locations, which makes it more difficult to identify any effect they might have on housing prices.

2.4 Generalized additive modeling: A "flexible" fixed effect

Several non-parametric and semi-parametric approaches exist to account for spatially correlated data. All of these methods are based on the recognition that the researchers have limited knowledge of the spatial structure and processes of the market. Rather than impose structure on the data, the data is allowed to speak. One such alternative is the Locally Weighted Regression (LWR), see e.g. McMillen (2012). The LWR is characterized by locally estimating parameters leading to variation in parameters across space. This variation will reflect any omitted spatial processes in so far as they correlate locally with the variables included in the model.[6] The generalized additive model discussed below allows variation in the overall level of prices across space through a flexible fixed effect, but keeps parameters constant. Essentially the model can be written as:

$$P = g^{-1}\left(X, f_1(x_{lon}, y_{lat}); \beta\right) \tag{2.9}$$

where $g^{-1}(\bullet)$ is the inverse of the link function, and $f(x_{lon}, y_{lat}; k)$ is a smooth function of the spatial coordinates capturing the exact location of the property. The smooth function is made up of the sum of k thin plate regression spline bases $b_h(\bullet)$ multiplied by their coefficients to be estimated: $f = \sum_{h=1}^{k} \beta_h b_h\left(x_{lon}, y_{lat}\right)$. The non-parametric component of the model $f(x_{lon}, y_{lat}; k)$ is fitted using thin plate regression splines with a penalty on "wiggliness". The penalty, θ, is determined from the data using generalized cross validation or related techniques. The penalty enters the objective function directly through an additional term capturing wiggliness in the smooth function, i.e.:

$$\left\| P - \widehat{P} \right\|^2 + \theta \int\int \left[\frac{\partial^2 f}{\partial x_{lon}^2} + \frac{\partial^2 f}{\partial x_{lon} \partial y_{lat}} + \frac{\partial^2 f}{\partial y_{lat}^2} \right]^2 dx_{lon} dy_{lat} \tag{2.10}$$

Here $\widehat{P}$ is the fitted dependent variable and the second derivatives of the smooth function describe its wiggliness. The objective function explicitly contains the trade-off between bias and variance.[7] The researcher must choose the flexibility of the model by setting the number of basis functions k. This is a balancing act between accurately capturing the locational attribute without overfitting the model, although the penalty term also reduces the probability of overfitting.

The higher the choice of k, the less spatial variation remains in the data to be explained by other

[5] Abbott and Klaiber (2010a) have shown in the context of green space, that a spatial Hausman-Taylor model can recover components which do not vary within fixed effect entities. However, that solution requires that good instruments are available for the variable of interest.

[6] Geniaux and Napoleone (2008) compare LWR with a Generalized Additive Model similar to the one discussed here.

[7] More information on the fitting of GAM with thin plate regression splines and the use of GCV and alternative methods can be found in, e.g. Wood (2006) and in the vignette for the `mgcv` package in `R`.

variables. In this way, there is a clear parallel between the choice of k and the scale of the spatial fixed effect. Essentially, it is difficult to separate the influence of included covariates from that of omitted processes when both vary on a spatial scale. We require the included spatial covariates to vary on a finer spatial scale than the omitted spatial processes in order to identify them in the model. In comparison with the fixed effects estimator discussed above, there will be no discrete changes in the level of house prices across space in the GAM model. The location component instead acts as a sort of "flexible" fixed effect describing the landscape. Sensitivity to the scale of the fixed effect can then be carried out easily by varying the choice of k as we demonstrate below.

In the empirical example in this paper, the focus is on the flexible fixed effect, but it should be noted that the GAM has several other properties desirable for hedonic analysis. In addition to smoothing the spatial coordinates, the GAM can also include smooth terms for other regressors, e.g. to determine an appropriate functional form. Furthermore, the GAM is a semi-parametric version of a generalized linear model (GLM). As such it naturally incorporates log transformation of the dependent variable, which is common in hedonic analysis. For calculation of marginal prices, the predicted (expected) price in levels rather than logs must be calculated after estimation. With GLM and a log transformed dependent variable that calculation needs to take account of heteroscedastic errors using, e.g. Duan smearing, see Cameron and Trivedi (2009). For estimation of the GAM a logarithmic link function can be specified, and levels rather than logs can be predicted directly.

3 Empirical application

The purpose of the following empirical exercise is to illustrate the sensitivity of the results from hedonic modeling to different spatial specifications and modeling principles. To that end, we estimated the hedonic price function using four different models: a simple linear model with no spatial corrections, a SEM, a spatial fixed effect model and a GAM. For the SEM, we then vary the choice of weight matrix, and for the GAM, the number of basis functions to evaluate the sensitivity of our results to the level of the spatial correction.

In each of the models, we model spatial processes in the data in two ways: To capture the finer structure at a neighborhood level we include a vector of variables Z_i which describes the average visible characteristics of homes in the neighborhood of dwelling i at the level of the road for each house. These average characteristics are calculated based on *all* houses (including those not traded within our time frame) in the same street as house i, and are intended to proxy for unobservable neighborhood characteristics in close proximity to the individual dwelling. On a large spatial scale, our approach varies across the spatial models. For the parametric models we include a linear measure of the distance to the central business district. The fixed effect model additionally has fixed effects at the level of school districts. For the GAM our approach is based on the recognition that we do not know a priori how the land rent gradient declines as the distance from the center increases. For this reason, we model the location of the property through a smooth function of the spatial coordinates, rather than by including distance to the central business district. The smooth spatial component captures the shape of the land rent gradient. These approaches account for the spatial structure of the housing market at an aggregate level in our models.

To facilitate comparison across models, we have made a number of common assumptions: All models are estimated with maximum likelihood estimators and assume a Gaussian distribution. The choice of estimator is not perfect for either the SEM or the GAM. For the SEM, the GMM approach developed by Kelejian and Prucha (2010) would be a better option as it allows for unknown heteroskedasticity and is computationally fast. The GAM requires specification of an exponential family distribution for the dependent variable. As

the house price is always positive and the variance increases with the price, a Gamma distribution would be a more appropriate choice than the Gaussian distribution applied here. For each of the models, the dependent variable was log-transformed before estimation.

We estimate the generalized additive model:

$$ln\, P_i = X_i\beta + Z_i\gamma + f(x_{lon,i}, y_{lat,i}; k) + u_i \tag{3.1}$$

where $f(x_{lon}, y_{lat}; k)$ is a smooth function of the spatial coordinates capturing the exact location of the property. The SEM is specified with a simple linear term to account for the distance from the Central Business District (CBD):[8]

$$ln\, P_i = X_i\beta + CBD_i\delta + Z_i\gamma + u_i \tag{3.2}$$

$$u_i = \lambda W u_i + \eta_i \tag{3.3}$$

$$\eta_i \sim N\left(0, \sigma_\eta^2\right) \tag{3.4}$$

For the error model we used a row standardized spatial weight matrix, W, which captures the 10 nearest neighbors. The chosen spatial weight matrix corresponds to the standard choice of spatial weight matrix in the literature. The linear model is identical to (3.2)-(3.4) with $\lambda = 0$. Finally, the fixed effect specification is given by:

$$ln\, P_{ij} = a_j + X_i\beta + CBD_i\delta + Z_i\gamma + e_{ij} \tag{3.5}$$

where a_j is the fixed effect for school district j. For the fixed effect model, errors were clustered at the school district level to account for residual spatial correlation.

All models are estimated in `R(2012)`. The non-spatial linear model and the fixed effect model are estimated using software for generalized linear models (`glm`). The generalized additive model is estimated with the `mgcv` package developed for `R` by Simon Wood, see e.g. Wood and Augustin (2002), and the SEM is estimated using the `spdep` package (Bivand, 2012).

3.1 Data

The data set covers the transactions of single family houses in the city of Aalborg, Denmark, over the period from 2000 to 2007. The study area is depicted in figure 3.1 which shows the distribution of transacted properties on a map of the buildings in Aalborg. Aalborg is the fourth largest town in Denmark with approx. 125,000 inhabitants (2010). In terms of owner occupied dwellings approximately half of the available housing units consist of houses. In total 6,313 transactions were included in the analysis.

[8] The specification we have chosen for the spatial processes in equations (3.2) to (3.3) is a spatial autoregressive model, which implies that the spatial variation occurs over a larger area than would be the case with, e.g. a spatial moving average representation where correlations die out faster as we move further away from the single observation. The neighborhood variables in Z_i are intended to capture more local effects and were also included in our estimates of the SEM.

Fig. 3.1: The survey area, single family and terraced houses

The data set contains information about each transaction in terms of price, date and type of sale. Data also contains information on the structural characteristics of the property such as the number of rooms and size of the living area. A summary of the control variables in the data set is found in Table 1. The information was extracted from the Danish Registry of Buildings and Housing database which contains information on all dwellings in Denmark (Ministry of Housing, 2012). The data is a "snapshot" of the housing characteristics and is continuously updated. Our data therefore reflects the characteristics of a house in August 2011 when the data were collected. The register contains information on the date of the latest renovation, so it is possible to control for post-sale renovations. The exact coordinates of the location of each dwelling are also available. Based on this information and maps from the Danish National Survey and Cadastre, a number of measures of proximity have been calculated using ArcGIS desktop 10.1, e.g. proximity to large roads, industrial sites, and different types of green space.

Tab. 1: Control variables describing housing characteristics

Structural variables	Spatial Variables	Neighborhood variables (Z)
Size of living area	Highway	Average Garden
Room	Large Road	Share with Brick walls
Garden area	Railway track	Average Age
Basement	Industrial area	Share with Tile roof
Number of floors	Coastline	Share Renovated in 1970s
Number of apartments	City Center	Share Renovated in 1980s
Low basement	Park	Share Renovated in 1990s
Renovation 1970s	Nature	Share Renovated in 2000s
Renovation 1980s	Lake	
Renovation 1990s	Common area	
Renovation 2000s	Sport field	
Built before 1927	Agriculture field	
Built between 1927 and 1939	Scrapland	
Built between 1939 and 1955	Churchyard	
Built between 1955 and 1975		
Built between 1975 and 1999	*Other locational*	
Brick walls	Hasseris (High income area)	
Tile roof	Geographical coordinates	
Fiber board roof		

The neighborhood variables contained in our vector Z (see Equation 3.1, 3.2, 3.5) hold information about the appearance of surrounding properties in terms of the average age, average of dummies for renovation in the years preceding the sale, and the style of the building as captured by roof type and brick walls. Finally, the average size of gardens for houses in the same street was included as this gives an idea of the development density in the area. A description and a set of descriptive statistics of the data are found in appendix A. Given that the data set comprised 8 years of sales it was necessary to adjust for inflation in the house prices. We did this by fitting a fourth degree polynomial in the date of sale. In this way the inflationary movements are filtered out leaving the remainder of the variation in house prices to be explained by the housing characteristics in the model.

As we have several control variables describing each house, we limit our discussion of results to a few variables. Access to green space is inherently a spatial variable and has long been a topic for hedonic analysis. Recent surveys include McConnell and Walls (2005) and Waltert and Schlaepfer (2010). As green space has been so extensively studied (and such variation in results has been found) it is a useful example for our purposes, namely to demonstrate sensitivity to spatial modeling choices. We have identified several different types of green space which differ both in the services provided and in terms of their prevalence in the urban landscape. Our discussion of results will mainly focus on parks, natural areas, residential common areas, lakes, and scraplands for brevity. We also discuss the robustness of a few of the so-called "structural" characteristics of the houses for comparison. These variables are the size of the living area, lot size, and type of wall covering. Variation in these characteristics is not primarily spatial, and parameter estimates should therefore be less sensitive to the choice of spatial model for these variables.

3.2 Modeling spatial variables

The spatially delineated variables in the hedonic price function proxy for accessibility or exposure to an amenity. In general, the effect of such amenities have a limited spatial extent. Apart from air pollution

the majority of the hedonic literature has focused on amenities with a local impact (Palmquist, 2005). The spatial extent of the amenity should be given careful consideration. If no boundary is specified for the effect of the amenity in question, these proximity measures may end up capturing possible omitted spatial trends in the data, which are unconnected with the amenity. We therefore specify cut-off distances for these spatial variables beyond which the effect of the amenity is expected to be absent.

We describe accessibility to green space using proximity to the nearest property in a straight line with the exception of the common area category. Common area green space is attached to specific residential areas, which means that distance to nearest common area is generally small. However, the size of common areas varies and is included as our regressor. We work with two different proximity cut-offs (c_{cutoff}) for different types of green space to capture different scales of capitalization (Abbott and Klaiber, 2010b). Some types of public green space are used for outings and people would be willing to travel further to enjoy a stay in such a green space, whereas other types of green space are de facto a club good, e.g. because they are small and located out of the way in the middle of a residential area. This should be reflected by capitalization of the latter types at a more local scale. We set the high cut-off to 600 meters reflecting an 8-10 minute walking time for parks and natural areas. The lower cut-off for club goods was set at 300 meters for the remaining types of green space. The scale of proximity is calculated by $X_{prox} = c_{cutoff} - X_{dist}$, where X_{dist} is distance in a straight line from the house. Further, for homes beyond the cut-off distance the measure of amenity access is set to zero, $\{X_{prox}|X_{prox} < 0\} = 0$. The coefficients on the proximity variables are easy to interpret as amenities are expected to have positive coefficients and disamenities to have negative coefficient estimates. A quadratic specification has been applied to all green space proximity variables, as previous studies have found non-constant marginal effects (Panduro and Veie, 2013). The common space variable has not been transformed.

Tab. 2: Descriptive statistics - selected regressors

	spec	count	Min	Median	Mean
Parks	Prox. (meters)	18	3.7	943.8	1,277.1
Natural areas	Prox. (meters)	60	0	486.8	562.8
Lake	Prox. (meters)	6	0	1,739.6	1,779.6
Common area	Size (Ha)	113	0.2	0.7	1.5
Scraplands	Prox. (meters)	269	2.4	366.9	420.5
Size (log)	Size (log)(m^2)	-	4.0	4.9	4.9
Garden	Size(m^2)	-	0	736	681
Brick	Dummy	6,016	-	-	-

4 Results

4.1 Model estimates

Table 3 includes estimates for the GLM model, the fixed effect model, the SEM, and the GAM. The table contains parameter estimates of the selected regressors from table 2. The performance of each model is described by R^2, log likelihood, and AIC. Spatial autocorrelation of the residuals of each of the models are tested using a global Moran's I statistic based on a row standardized inverse distance weighted spatial weight matrix with a cut-off value of 500 meters. On average a 500 meters distance corresponds to 130 neighbors in the weight matrix, however, the number of neighbors varies with the density of the urban area. Parameter estimates of the full models can be found in appendix B.

The estimates of the structural variables across all four models are highly significant and vary only marginally. Among the spatial variables, proximity to parks, scraplands and the size of the nearest common area are associated with significantly higher and lower prices. Proximity to natural areas does not have a significant impact on the house price in the fixed effect model. For the GLM and the SEM an effect is significant at a 5 and 10 % level, respectively. The GAM provides highly significant estimates of the effect of proximity to natural areas, and proximity to lakes at the 10 %t level, but it should be noted, that the standard errors for the GAM and the GLM do not take account of the clustering of residuals in space. Hence significance levels for these two models are likely to be overestimated. The estimated coefficients for the spatial variables have the expected signs for all models except for the Common area variable, which has a negative effect. There is some variation in the size of the estimated coefficients. For proximity to Scraplands, the largest estimate (GAM) is 1.4 times as large as the smallest (SEM), although the value lies within the spatial error confidence interval. Looking across models, the span of coefficient estimates is rather wide and it is not without importance which model is used to calculate marginal prices.

The Moran's I statistic shows that the residuals of the models are significantly clustered in space within 500 meters of each dwelling. Although the Moran's I values for the residuals are significant for all models, they are relatively small, which may be due to the number of spatial covariates in the model. The GAM has the lowest Moran's I value followed by the SEM, the fixed effect model, and the GLM model. The GAM's focus on spatial processes at a broader level may explain why it performs better than the SEM with 10 neighbors in the weight matrix for the displayed Moran's I statistic.

In the SEM, the spatial autocorrelation coefficient, λ, is highly significant. The spatial smooth term in the GAM is highly significant and can be understood as the land rent gradient of the housing market. The spatial smooth term is mapped below in figure 4.1.

While the three spatial models cannot be compared directly as they are not nested, the fixed effect model and the error model are directly comparable to the GLM model and perform better on all the displayed model criteria in table 3. In the GAM, the proximity to the central business district has been replaced with the smooth "land rent gradient", which seems to increase the model's performance significantly, although, the loss of degrees of freedom is also larger than for the other spatial models.

Tab. 3: Model estimates

	GLM		Fixed effect		SEM		GAM	
Spatial variables								
Park^2	0.001831	***	0.003257	***	0.001881	*	0.002296	***
	(0.000458)		(0.000745)		(0.000737)		(0.000561)	
Nature^2	0.000739	*	0.000232		0.000994	+	0.001474	***
	(0.000339)		(0.000907)		(0.000527)		(0.000441)	
Lake^2	0.000132		0.000992		0.000539		0.001467	+
	(0.000496)		(0.001012)		(0.000807)		(0.000762)	
Scrapland^2	-0.006646	***	-0.006308	**	-0.005306	*	-0.007624	***
	(0.001703)		(0.002239)		(0.002466)		(0.001867)	
Common area	-0.004235	**	-0.0048543	+	-0.003597	*	-0.002167	+
	(0.001436)		(0.002598)		(0.001442)		(0.001237)	
Structural variables								
Size (log)	0.522822	***	0.51650	***	0.470979	***	0.490631	***
	(0.018421)		(0.02892)		(0.016441)		(0.016404)	
Garden	0.000177	***	0.000176	***	0.000190	***	0.000176	***
	(0.000017)		(0.000026)		(0.000013)		(0.000012)	
Brick	0.078805	***	0.078454	***	0.081458	***	0.080204	***
	(0.013950)		(0.01397)		(0.013908)		(0.013834)	
λ					0.49081	***		
k_{geo}							40	***
							t-statistic: 23.66	
							38.584 edf.	
GCV-score							0.043363	
df.	45		66		46		82	
$R^2(Adj.)$	0.6838		0.6909		0.7154		0.706	
Log likelihood	784.5938		868.9638		1038.626		1029.708	
AIC	-1479.2		-1605.928		-1985.3		-1894.722	
Moran's I	0.110042	***	0.08974707	***	0.05307406***		0.02356792	***

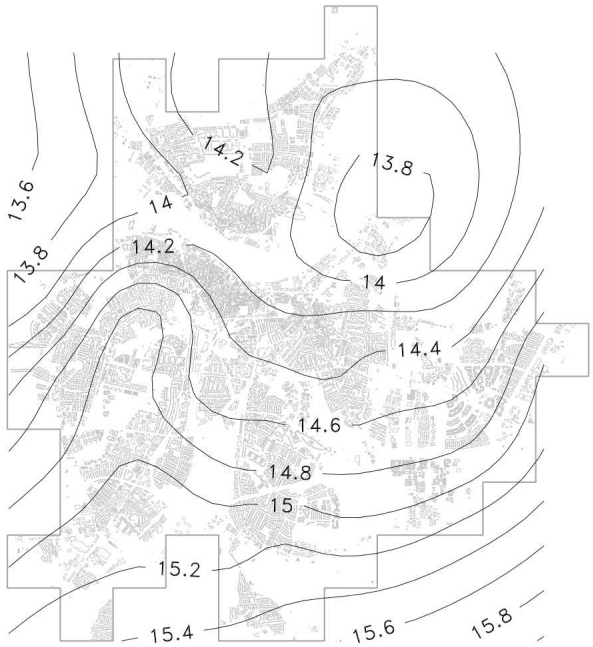
Note: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^+$: $p < 0.1$

4.2 Spatial smoothing

According to Wood (2006) the choice of basis dimensions is a part of model specification and the researcher should aim to ensure that sufficient flexibility is available for the individual application. The results presented in table 3 are for a basis dimension of $k = 40$ for the geographical coordinates, which was chosen based on a rule of thumb: $k = min\{\frac{n}{4},40\}$, see Ruppert (2002). The penalty term, θ, was determined through generalized cross-validation. The estimated spatial structure is depicted in the contour plots in figure 4.1. The plots show the spatial pattern of the log of the transactions price at the median of the covariates for properties built in the period between 1955 and 1975.

Generally speaking, the spatial price trend in Aalborg seems to conform to a wave of high prices rising in the southern part of Aalborg and falling near the industrial area near the harbor area. A local depression in prices is found northeast of the city. Note that the rate of decline in prices depends on the direction of movement away from the high price areas.

Fig. 4.1: Spatial Smoothing



4.3 Sensitivity analysis – spatial autocorrelation

Our models capture spatial processes at different levels with the SEM being the most localized and the GAM capturing broader trends in the data. As the spatial processes in the data are likely to occur at varying spatial frequencies, we test spatial autocorrelation for each of the four models within 9 distance bands. The nearest ring spans 0-200 meters around the observation, the next ring spans to 200-400 meters and so forth all the way up to 2500-3000 meters, see table 4. The tests are based on global Moran's I values for the residuals from each model against the null hypothesis of spatially random errors. Each of the distance bands is represented by a row standardized spatial weight matrix which assumes equal weights on the observations within the distance bands.

The global Moran's I values for the GLM model are significantly different from a random distribution across all distances bands. The Moran's I values decline rapidly from the nearest distance band at 0-200 meters. The residuals of the GLM model are significantly more clustered at all distance bands up to 800-1000 meters. Beyond this distance band, the residuals are significantly more dispersed than expected (negative Moran's I) if their distribution was random across space. The Moran's I values of the fixed effect model are generally lower than for the GLM model and only positive until the 800 meter band, after which they become negative. Both for the GAM and the SEM, Moran's I is lower than the GLM model at almost all distance bands. The SEM has fewer significant Moran's I values than the GAM. The GAM seems to result in overdispersed residuals at shorter distances than the other models.

To test sensitivity of the SEM and the GAM to the choice of weight matrix and basis function dimension, we varied these choices, re-estimated the models and re-calculated Moran's I for the different distance bands.

The SEM is re-estimated using a row standardized spatial weight matrix based on the 10, 20, 30, 40 and 50 nearest neighbors and then tested across the spatial distance bands (see table 5). For the SEM, residuals remain significantly clustered in those distance bands closest to the house as we increase the number of neighbors in the weight matrix. Spatial autocorrelation is reduced with an increase of neighbors in the error model based on 30, 40 and 50 nearest neighbors although significant global Moran's I values remain. The choice of 30 neighbors seems significantly better than 10 neighbors based on the calculated statistics. The GAM is re-estimated using a basis dimension of $k = \{20, 40, 60, 80, 100\}$ for the spatial smoothing splines (see table 6). The presence of spatial autocorrelation is reduced with an increase in the number of basis dimension especially for distance bands further away from the dwelling. However even for $k = 100$, significant clustering remains at a local level although the Moran's I values are small.

None of the spatial models completely removes the presence of spatial autocorrelation in the residuals, although there is less of a spatial pattern in the residuals as the spatial dimension is increased. There is no monotonic decline in the Moran's I values as we increase the dimension of the weight matrix or increase the choice of basis functions. Increasing the weight matrix corresponds to using a larger window in kernel estimation. Hence it is unsurprising that Moran's I for the nearest distance band should increase as the number of neighbors gets large. With a larger window size, the finer spatial processes will not be captured as precisely. Naturally, the results may differ if the spatial weights decline with distance, i.e. inverse distance weighting. For the GAM, different choices of k also seem to capture different spatial processes as the Moran's I statistic becomes insignificant for varying distance bands. In both cases, the findings are consistent with the existence of several distinct spatial processes. Note that variations in weight matrix and basis spline dimension in the GAM model and the SEM are not directly comparable. We have no way of determining what a given number of smoothing splines corresponds to in terms of spatial weight matrix.

Tab. 4: Moran's I for different distance bands

Distance band	GLM	Fixed effects	SEM[a]	GAM[a]
0-200	0.1234***	0.1032***	0.0197***	0.0660***
200-400	0.0597***	0.0373***	0.0230***	0.0049
400-600	0.0338***	0.0158***	0.0133***	-0.0113***
600-800	0.0252***	0.0066**	0.0121***	-0.0098***
800-1000	0.0100***	-0.0055*	0.0037	-0.0095***
1000-1500	-0.0054***	-0.0116***	-0.0018	-0.0082***
1500-2000	-0.0036**	-0.0028**	-0.0014	0.0062***
2000-2500	-0.0093***	-0.0022*	-0.0041***	0.0024**
2500-3000	-0.0111***	-0.0008	-0.0046***	-0.0018*

Note: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{+}$: $p < 0.1$

[a)] The SEM estimated with 10 neighbors and the GAM estimated with $k = 40$.

Tab. 5: Spatial autocorrelation - SEM

Dist.\Neigh.	10	20	30	40	50
0-200	0.0197***	0.0094**	0.0094**	0.0146***	0.0232***
200-400	0.0230***	0.0123***	0.0050	-0.0010	-0.0038
400-600	0.0133***	0.0073**	0.0051	0.0032	0.0023
600-800	0.0121***	0.0077**	0.0059*	0.0052*	0.0044
800-1000	0.0037	0.0022	0.0023	0.0021	0.0028
1000-1500	-0.0018	-0.0014	-0.0009	-0.0002	0.0002
1500-2000	-0.0014	-0.0000	0.0010	0.0015	0.0014
2000-2500	-0.0041***	-0.0023*	-0.0015	-0.0013	-0.0011
2500-3000	-0.0046***	-0.0032***	-0.0030***	-0.0022**	-0.0021*

Note: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{+}$: $p < 0.1$

Tab. 6: Spatial autocorrelation - Generalized additive model

Dist.\k	20	40	60	80	100
0-200	0.0820***	0.0660***	0.0453***	0.0365***	0.0310***
200-400	0.0197***	0.0049	-0.0116***	-0.0185***	-0.0208***
400-600	-0.0000	-0.0113***	-0.0172***	-0.0181***	-0.0166***
600-800	-0.0084***	-0.0098***	-0.0057*	-0.0042	-0.0023
800-1000	-0.0115***	-0.0095***	-0.0009	0.0023	0.0030
1000-1500	-0.0097***	-0.0082***	-0.0012	0.0005	0.0006
1500-2000	0.0033***	0.0062***	0.0033***	0.0010	0.0010
2000-2500	-0.0006	0.0024**	0.0001	0.0000	-0.0000
2500-3000	0.0015*	-0.0018*	-0.0011	0.0000	0.0004

Note: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^{+}$: $p < 0.1$

4.4 Sensitivity analysis – coefficient robustness

By increasing the dimension of the weight matrix in the SEM or the basis functions for the smooth component in the GAM we were able to remove much of the residual spatial autocorrelation. The next question is whether the estimated coefficients are robust to variations in the spatial dimension within models. The coefficient estimates for the selected regressors from table 2 are displayed in tables 7 and 8 for varying dimensions of the spatial weight matrix and geographical smoothing splines. The estimated parameters for structural (non-spatial) characteristics are robust across the spatial dimensions while the results for the spatially varying regressors are sensitive to the dimension of the spatial weight matrix in the error model and the number of spline basis functions. In the error model, parameter estimates for proximity to parks remain relatively stable and retain significance levels. For proximity to scraplands, the estimated parameter becomes larger in absolute terms as the number of neighbors increases. The significance level also increases. In the GAM, proximity to parks, natural areas and scraplands are significant up to 60 basis functions for the smooth land rent gradient. For larger k, the parameter estimates of parks become insignificant while nature and scrapland remain significant. The parameter estimates in the GAM for lake and common area are not robust. The lake variable is only significant at the 10 percent level for $k = 40$ and $k = 60$ and the size of the nearest common area is only associated with significantly different house prices with low values of k.

Scrapland is the only spatial variable which remains robust over both the different sizes of spatial weight matrices and varying levels of k basis functions. As mentioned, the models differ in the assumptions about

the correlations in the data. The error model assumes zero correlation between omitted spatial processes and included regressors, while the GAM captures the variation directly, as in the case of the fixed effects model. Of the two models then, the GAM seems the more prudent choice. This does not imply that access to, e.g. parks has no positive effect on house prices, it simply implies that these effects are hard to distinguish from omitted spatial processes using a high level of smoothing splines. Essentially, the smoothing splines compete with the spatial variables in explaining variations in the price levels across space. The variation in the variable of interest must be greater than the variation in the modeled spatial processes for an effect to be identified when the "flexible" fixed effect is included. In our case, the access to the (dis)amenity of interest is measured by proximity to the nearest object. For objects which are scarce in the urban landscape, there will be little variation in this measure across space. There are only 13 parks in the Aalborg area, whereas there are more than 200 scraplands spread out across the urban landscape.

Tab. 7: Spatial robustness - SEM

	10		20		30		40		50	
Park^2	0.001881	*	0.001820	*	0.002017	*	0.002009	*	0.001923	*
Nature^2	0.000994	+	0.001028	+	0.000980		0.001207	*	0.001395	*
Lake^2	0.000539		0.000341		0.000150		0.000326		0.000316	
Scrapland^2	-0.005306	*	-0.004996	*	-0.005829	*	-0.006688	**	-0.007285	***
Common area	-0.003597	*	-0.002290		-0.000798		-0.000566		-0.000489	
Size (log)	0.470979	***	0.463614	***	0.464899	***	0.465234	***	0.470670	***
Garden	0.000190	***	0.000187	***	0.000183	***	0.000180	***	0.000175	***
Brick	0.081458	***	0.078656	***	0.082338	***	0.081681	***	0.079803	***

Note: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^+$: $p < 0.1$

Tab. 8: Spatial robustness - generalized additive model

	20		40		60		80		100	
Park^2	0.001713	***	0.002296	***	0.001327	*	0.000461		0.000894	
Nature^2	0.001153	***	0.001474	***	0.001473	**	0.001949	***	0.001612	**
Lake^2	0.000007		0.001467	+	0.001587	+	0.001634		0.001701	
Scrapland^2	-0.005905	***	-0.007624	***	-0.008231	***	-0.007963	****	-0.006948	***
Common area	-0.002919	*	-0.002167	+	-0.000612		-0.000312		-0.000808	
Size (log)	0.494049	***	0.490631	***	0.484113	***	0.485431	***	0.482514	***
Garden	0.000178	***	0.000176	***	0.000171	***	0.000170	***	0.000171	***
Brick	0.081635	***	0.080204	***	0.077359	***	0.075998	***	0.076575	***

Note: ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, $^+$: $p < 0.1$

5 Concluding discussion

This paper builds on a criticism of the existing spatial (parametric) models, where the spatial structure is specified and treated as "known" either in terms of spatial weight matrices or through the use of spatial fixed effects. In practice, the spatial structure of omitted spatial processes is rarely known. We utilize an approach to account for spatial dependence which to our knowledge is novel to the literature on hedonic regressions. We model location as a semi-parametric function of the geographic coordinates which allows us to capture a large part of the spatial variation in the data using a GAM. We compare this model with alternatives commonly used in the literature: The SEM and the fixed effects model. We find the GAM to be

a sound alternative to the standard approaches in the hedonic house price literature, having less restrictive assumptions about the omitted spatial processes while still being able to reduce the problem of spatial autocorrelation and provide trustworthy estimates of spatial variables.

Spatial correlation in the error term can result from misspecification of the functional form or mismeasurement of spatial covariates or from omitted spatial covariates. The most obvious property of these omitted spatial processes is that the researcher does not know at which scale misspecification, mismeasurement and omitted variables operate. We suggest, therefore, that sensitivity analysis should be conducted using different levels of spatial corrections to determine which results are robust across models. In the existing literature, when parametric spatial econometrics is used, it is rarely the case that results are shown for different choices of weight matrices. The standard approaches rely on the researcher to specify how the omitted spatial processes vary in order to control for them in the model. Our findings suggest that omitted spatial processes are likely to play an important role in explaining the varying findings in hedonic models concerned with spatially delineated amenities.

It is a rather restrictive assumption that spatial autocorrelation can be corrected by a single spatial entity as in the fixed effect model. Omitted spatial processes are likely to be present on more than one spatial scale, which implies that the fixed effect model should use as small entities as possible to ensure that omitted spatial processes are accurately captured. Ideally a repeat sales model is capable of doing this, but requires time series variation in the amenity of interest. The critique of restrictive assumptions about the nature of omitted spatial processes can also be applied to the SEM. The spatial weight matrix typically applied in the SEM defines the spatial pattern of the omitted processes and imposes this restriction on the estimation. Given that spatial autocorrelation is limited to unfold in the (x,y)-geographical space, the spatial weight matrix is bound to pick up some of the spatial autocorrelation even though the weight matrix incorrectly models the omitted spatial processes. The appropriate dimension of the spatial weight matrix is bound to differ from application to application and should always be carefully tested, e.g. using distance bands as in our example. The GAM applies less restrictive assumptions about the structure of the spatial pattern of the omitted spatial processes. Essentially, the GAM is data driven in the sense that the specification of the spatial omitted spatial processes is determined by model fit using generalized cross validation. The choice of basis function dimension for the GAM remains a judgment call for the researcher, however, and sensitivity analysis should be carried out to check robustness of the model estimates.

In the empirical application of the hedonic house price model we estimated a GLM model, a spatial fixed effect model, a SEM, and a GAM. The structural variables that describe the properties are robust across model specifications. Variation in the dimension of spatial processes modeled in terms of spatial weight matrices in the error model and spatial smoothing splines in the GAM also seems to leave the structural variables unaffected. The spatial variables, represented by different types of green space, are more sensitive to the choice of model and dimension of the spatial model. This is not surprising given that the controls for omitted spatial processes will to some extent compete with the spatial covariates in explaining the data. In the empirical application this is most obvious in the fixed effect model where the parameter estimate of proximity to natural areas becomes insignificant. For identification there must be sufficient variation in the variable of interest independent of the variation in omitted spatial processes which the models attempt to correct for. In practice it is often difficult to say if the variable of interest varies on a sufficiently fine scale. In the empirical application, proximity to natural areas might not vary sufficiently across space to be identified in the model independently of the fixed effect. The same problem occurs for other spatial variables in the fixed effect model and the GAM as the spatial dimension is increased (see appendix B).

The SEM is able to reduce spatial autocorrelation better than the GAM at close distances. This does not apply at larger distances where the GAM outperforms the error model. The fixed effect model is not able to reduce spatial autocorrelation to the same extent as the two other spatial models. This is likely due to the relatively coarse scale of the fixed effect we employ. All three models do reduce residual spatial correlation in comparison with the linear model without spatial corrections. The error model assumes zero correlation between omitted spatial processes and included regressors. While zero correlation seems an appropriate assumption for non-spatial variables and a spatially correlated error term, it is less appropriate for the spatially varying regressors. We did not include a spatial econometric model with a spatial lag term in the empirical example. The interpretation of the spatial spillover in prices implied is hard to reconcile with Rosen's hedonic theory. Compared with the standard spatial econometric approaches the GAM model seems a prudent choice given its intuitive interpretation, the less restrictive assumptions and its ability to reduce the spatial correlation in the error term.

Although the potential for omitted variables bias is reduced through the increased use of geographical information systems to generate data, the inclusion of geographical covariates does not solve the omitted variable problem. Rather when spatially varying covariates are the main focal point of the analysis, extra care should be taken to ensure that results are robust to different spatial models as long as the true data generating process is unknown. In some cases spatial variation in the environmental variable in question can be increased through careful modeling of the services or sources of annoyance. For instance, in the empirical application we model accessibility by a proximity measure with a cut-off value which reflects the maximum distance that people are willing to travel in order to enjoy the amenity. We additionally construct neighborhood variables using the residential street to delineate the neighborhood. Despite these efforts spatial autocorrelation in the error term remains a problem. Careful attention to the nature of the environmental amenity and the way in which it is perceived by households improves the model's ability to measure household willingness to pay through reduction of measurement errors. However, for some environmental amenities cross-sectional hedonic analysis is unlikely to deliver reliable identification. Other methods are needed, such as instrumental variables with exogenous shifts in amenity levels, see Bayer et al. (2009), or quasi-experiments, see e.g. Pope (2008), though the latter are hard to interpret in terms of willingness to pay, see Kuminoff and Pope (2009).

References

Abbott, J. K. and Klaiber, H. A.: 2010a, An embarrassment of riches: Confronting omitted variable bias and multi-scale capitalization in hedonic price models, *Review of Economics and Statistics* **93**(4), 1331–1342.

Abbott, J. K. and Klaiber, H. A.: 2010b, Is all space created equal? uncovering the relationship between competing land uses in subdivisions, *Ecological Economics* **70**(2), 296 – 307.

Anselin, L.: 2010, Thirty years of spatial econometrics, *Papers in Regional Science* **89**(1), 3–25.

Anselin, L. and Lozano-Gracia, N.: 2008, Errors in variables and spatial effects in hedonic house price models of ambient air quality, *Empirical Economics* **34**(1), 5–34.

Bayer, P., Keohane, N. and Timmins, C.: 2009, Migration and hedonic valuation: The case of air quality, *Journal of Environmental Economics and Management* **58**(1), 1–14.

Bivand, R.: 2012, Spatial dependence: weighting schemes, statistics and models. R package version 0.4-56. **URL:** *http://cran.r-project.org/web/packages/spdep/index.html*

Brady, M. and Irwin, E.: 2011, Accounting for spatial effects in economic models of land use: Recent developments and challenges ahead, *Environmental & Resource Economics* **48**(3), 487–509.

Brounen, D. and Kok, N.: 2011, On the economics of energy labels in the housing market, *Journal of Environmental Economics and Management* **62**(2), 166 – 179.

Cameron, A. C. and Trivedi, P. K.: 2009, *Microeconometrics using Stata*, STATA Press.

Cavailhes, J., Brossard, T., Foltite, J.-C., Hilal, M., Joly, D., Tourneux, F.-P., Tritz, C. and Wavresky, P.: 2009, Gis-based hedonic pricing of landscape, *Environmental and Resource Economics* **44**, 571–590.

Chamblee, J. F., Colwell, P. F., Dehring, C. A. and Depken, C. A.: 2011, The effect of conservation activity on surrounding land prices, *Land Economics* **87**(3), 453–472.

Chay, K. Y. and Greenstone, M.: 2005, Does air quality matter? evidence from the housing market, *Journal of Political Economy* **113**(2), 376–424.

Day, B., Bateman, I. and Lake, I.: 2007, Beyond implicit prices: recovering theoretically consistent and transferable values for noise avoidance from a hedonic property price model, *Environmental & Resource Economics* **37**(1), 211–232.

Deaton, B. J. and Vyn, R. J.: 2010, The effect of strict agricultural zoning on agricultural land values: The case of ontario's greenbelt, *American Journal of Agricultural Economics* **92**(4), 941–955.

Ekeland, I., Heckman, J. J. and Nesheim, L.: 2004, Identification and estimation of hedonic models, *Journal of Political Economy* **112**(S1), 60–109.

Epple, D.: 1987, Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products, *Journal of Political Economy* **107**(August), 645–681.

Geniaux, G. and Napoleone, C.: 2008, *Hedonic Methods in Housing Markets*, number 5, Springer, chapter Semi-parametric tools for spatial hedonic models: An introduction to Mixed Geographically Weighted Regression and Geoadditive Models.

Gibbons, S. and Overman, H. G.: 2012, Mostly pointless spatial econometrics?, *Journal of Regional Science* **52**(2), 172–191.

Gopalakrishnan, S., Smith, M. D., Slott, J. M. and Murray, A. B.: 2011, The value of disappearing beaches: A hedonic pricing model with endogenous beach width, *Journal of Environmental Economics and Management* **61**(3), 297 – 310.

Heintzelman, M. D. and Tuttle, C. M.: 2012, Values in the wind: A hedonic analysis of wind power facilities, *Land Economics* **88**(3), 571–588.

Hoshino, T. and Kuriyama, K.: 2010, Measuring the benefits of neighborhood park amenities: Application and comparison of spatial hedonic approaches, *Environmental and Resource Economics* **45**(3), 429–444.

Kahn, S. and Lang, K.: 1988, Efficient estimation of structural hedonic systems, *International Economic Review* **29**(1), 157–66.

Kelejian, H. H. and Prucha, I. R.: 2010, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics* **157**(1), 53 – 67.

Kuminoff, N. and Pope, J. C.: 2009, Capitalization and welfare measurement in the hedonic model.

Kuminoff, N. V., Parmeter, C. F. and Pope, J. C.: 2010, Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?, *Journal of Environmental Economics and Management* **60**(3), 145–160.

Lesage, J. P. and Pace, R. K.: 2009, *Introduction to Spatial Econometrics*, Taylor and Francis Group, LLC.

McConnell, V. and Walls, M.: 2005, Assessing the non-market value of open space. Report, Ressources for the Future.

McMillen, D. P.: 2012, Perspectives on spatial econometrics: Linear smoothing with structured models, *Journal of Regional Science* **52**(2), 192–209.

of Housing, M.: 2012, http://www.boligejer.dk/om-ejendomsdata.

Palmquist, R. B.: 2004, Property value models, *Handbook of Environmental Economics*, Vol. 2, Elsevier North Holland.

Palmquist, R. B.: 2005, Chapter 16 property value models, *in* K.-G. Mler and J. R. Vincent (eds), *Valuing Environmental Changes*, Vol. 2 of *Handbook of Environmental Economics*, Elsevier, pp. 763 – 819.

Panduro, T. E. and Veie, K. L.: 2013, Classification and valuation of urban green spaces - a hedonic house price valuation, *Landscape and Urban Planning* **120**, 119 – 128.

Pope, J. C.: 2008, Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise, *Journal of Urban Economics* **63**(2), 498 – 516.

Rosen, S.: 1974, Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy* **82**(1), 34–55.

Ruppert, D.: 2002, Selecting the number of knots for penalized splines, *Journal of Computational and Graphical Statistics* **11**, 735–757.

Small, K. A. and Steimetz, S.: 2006, Spatial hedonics and the willingness to pay for residential amenities, *Working Papers 050631*, University of California-Irvine, Department of Economics.

Team, R. C.: 2012, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Walsh, P. J., Milon, J. W. and Scrogin, D. O.: 2011, The spatial extent of water quality benefits in urban housing markets, *Land Economics* **87**(4), 628–644.

Waltert, F. and Schlaepfer, F.: 2010, Landscape amenities and local development: A review of migration, regional economic and hedonic pricing studies, *Ecological Economics* **70**(2), 141 – 152.

Won Kim, C., Phipps, T. T. and Anselin, L.: 2003, Measuring the benefits of air quality improvement: a spatial hedonic approach, *Journal of Environmental Economics and Management* **45**(1), 24–39.

Wood, S. N.: 2006, *Generalized Additive Models: An Introduction with R*, Chapman and Hall.

Wood, S. N. and Augustin, N. H.: 2002, Gams with integrated model selection using penalized regression splines and applications to environmental modelling, *Ecological Modelling* **157**(2–3), 157 – 177.

A Descriptive statistic

A.1 Variable description

Name	Description
Price	The price of the property in DKK.
Time	The analysis period measured in days from January 1st 2000 to December 31st 2007.
Size (log)	The natural logarithm of the size of the living area measured in square meters.
Room (log)	The natural logarithm of the number of rooms in the house.
Basement	The size of the Basement in square meters.
Garden	The size of the Garden in square meters.
Number of floors	The number of floors in the house.
Highway	Proximity to the nearest highway described on a scale from 0 to 500 meters where 0 corresponds to being at distance of 500 meters and 500 corresponds to being located on the highway.
Large road	Proximity to the nearest Large road described on a scale from 0 to 100 meters where 0 corresponds to being at distance of 100 meters and 100 corresponds to being located on the Large road.
Railway track	Proximity to the nearest railway track described on a scale from 0 to 100 meters where 0 corresponds to being at distance of 100 meters and 100 corresponds to being located on the railway track.
Industrial area	Proximity to the nearest industrial area described on a scale from 0 to 500 meters where 0 corresponds to being at distance of 500 meters and 500 corresponds to being located on the industrial area.
Coastline	Proximity to the coastline described on a scale from 0 to 100 meters where 0 corresponds to being at distance of 100 meters and 100 corresponds to being located on the coastline.
Spatial lag: Garden	The average size of Gardens of all houses on the residential street.
Spatial lag: Brick	The average number of houses with the outer wall made of Brick on the residential street.
Spatial Lag: Age	The average age of all houses on the residential street.
Spatial Lag: Tile roof	The average number of houses with a Tile roof on the residential street.
Spatial Lag: Renovation in 1970s	The average number of houses which have undergone major renovation during the 1970s on the residential street.
Spatial Lag: Renovation in 1980s	The average number of houses which have undergone major renovation during the 1980s on the residential street.
Spatial Lag: Renovation in 1990s	The average number of houses which have undergone major renovation during the 1990s on the residential street.
Spatial Lag: Renovation in 2000s	The average number of houses which have undergone major renovation during the 2000s on the residential street.

Name	Description
Low Basement	Dummy variable that describes the presence of a Basement with a ceiling height less than 120 cm - 1 corresponds to the presence of low Basement and 0 corresponds to the absence of low Basement.
Renovation 1970s	Dummy variable that describes whether the house has undergone major renovation during the 1970s. 1 corresponds to major renovations and 0 corresponds to no major renovation.
Renovation 1980s	Dummy variable that describes whether the house has undergone major renovation during the 1980s. 1 corresponds to major renovations and 0 corresponds to no major renovation.
Renovation 1990s	Dummy variable that describes whether the house has undergone major renovation during the 1990s. 1 corresponds to major renovations and 0 corresponds to no major renovation.
Renovation 2000s	Dummy variable that describes whether the house has undergone major renovation during the 2000s before the house is sold. 1 corresponds to major renovations and 0 corresponds to no renovation.
Renovation after	Dummy variable that describes whether the house has undergone major renovation after it was sold. 1 corresponds to major renovations and 0 corresponds to no renovation.
Built before 1927	Dummy variable that describes whether the house was built before 1927. 1 corresponds to being built before 1927 and 0 corresponds to being built after 1927.
Built between 1927 and 1939	Dummy variable that describes whether the house was built between 1927 and 1939. 1 corresponds to being built between 1927 and 1939 and 0 corresponds to not being built between 1927 and 1939.
Built between 1939 and 1955	Dummy variable that describes whether the house was built between 1939 and 1955. 1 corresponds to being built between 1939 and 1955 and 0 corresponds to not being built between 1939 and 1955.
Built between 1955 and 1975	Dummy variable that describes whether the house was built between 1955 and 1975. 1 corresponds to being built between 1955 and 1975 and 0 corresponds to not being built between 1955 and 1975.
Built between 1975 and 1999	Dummy variable that describes whether the house was built between 1975 and 1999. 1 corresponds to being built between 1975 and 1999 and 0 corresponds to not being built between 1975 and 1999.
Brick	Dummy variable that describes whether the outer wall of the house is made of Bricks. 1 corresponds to the outer wall consists of Bricks and 0 corresponds to the outer wall consists of other materials.
Tile roof	Dummy variable that describes whether the roof is made of tile. 1 corresponds to being made of tile and 0 corresponds to not made of tile.
Fiber board roof	Dummy variable that describes whether the roof is made of fiber board. 1 corresponds to being made of fiber board and 0 corresponds to not made of fiber board.

Name	Description
Terraced house	Dummy variable that describes whether house is a single family house or a terraced house. 1 corresponds to terraced house and 0 corresponds to single family house.
Hasseris	Dummy variable that describes whether the house is located in Hasseris. 1 corresponds to being located in Hasseris and 0 corresponds to not being located in Hasseris. Hasseris is the high income area in Aalborg.
Park	Proximity to the nearest Park described on a scale from 0 to 600 meters where 0 corresponds to being at distance of 600 meters and 600 corresponds to being located on the boarder of the Park. Proximity is measured in 100 meters.
Nature	Proximity to the nearest nature areas described on a scale from 0 to 600 meters where 0 corresponds to being at distance of 600 meters and 600 corresponds to being located on the boarder of the nature area. Proximity is measured in steps of 100 meters.
Lake	Proximity to the nearest Lake area described on a scale from 0 to 600 meters where 0 corresponds to being at distance of 600 meters and 600 corresponds to being located on the boarder of the nature area. Proximity is measured in steps of 100 meters.
Common area size	The size of the nearest Common area measured in Hectares.
Sport field	Proximity to the nearest Sport field described on a scale from 0 to 300 meters where 0 corresponds to being at distance of 300 meters and 300 corresponds to being located on the boarder of the Sport field. Proximity is measured in 100 meters.
Agriculture field	Proximity to the nearest Agriculture field described on a scale from 0 to 300 meters where 0 corresponds to being at distance of 300 meters and 300 corresponds to being located on the boarder of the Agriculture field. Proximity is measured in steps of 100 meters.
Scrapland	Proximity to the nearest Scrapland described on a scale from 0 to 300 meters where 0 corresponds to being at distance of 300 meters and 300 corresponds to being located on the boarder of the Scrapland. Proximity is measured in 100 meters.
Churchyard	Proximity to the nearest churchyard described on a scale from 0 to 300 meters where 0 corresponds to being at distance of 300 meters and 300 corresponds to being located on the boarder of churchyard. Proximity is measured in steps of 100 meters.

A.2 Descriptive statistics - Continuous Variables

Variable	Min	X1st.Q.	Median	Mean	X3rd.Q.	Max
Price	250000	1020000	1300000	1449000	1680000	9000000
Time	1	774	1514	1474	2179	2918
Garden	0	496	736	681.400000	848	3538

Variable	Min	X1st.Q.	Median	Mean	X3rd.Q.	Max
Size (log)	4.025000	4.682000	4.875000	4.877000	5.063000	6.038000
Room (log)	0	1.386000	1.609000	1.508000	1.609000	2.708000
Basement	0	0	0	22.700000	44	210
Number of floors	1	1	1	1.100000	1	3
Highway	0	0	0	51.360000	13.520000	472.360000
Large road	0	0	0	0.190100	0	1
Railway track	0	0	0	1.098000	0	87.300000
Industrial area	0	0	130.800000	162.900000	308	500
Coastline	1.071000	10.869000	20.973000	24.273000	37.088000	66.036000
Spatial lag: Garden	0	583.100000	750	707.900000	858.700000	2200
Spatial lag: Brick	0.114300	0.937500	0.990200	0.951900	1	1
Spatial Lag: Age	1850	1942	1960	1957	1970	2009
Spatial Lag: Tile roof	0	0.037040	0.120000	0.218530	0.322030	1
Spatial Lag: Renovation in 1970s	0	0.016390	0.083330	0.102930	0.153850	1
Spatial Lag: Renovation in 1980s	0	0.034480	0.083330	0.093270	0.134020	1
Spatial Lag: Renovation in 1990s	0	0.015500	0.061860	0.072030	0.106670	1
Spatial Lag: Renovation in 2000s	0	0	0.043480	0.061740	0.086960	1
Park	0	0	0	0.916200	1.461000	5.963000
Nature	0	0	1.086000	1.754000	3.334000	6
Common area size	0	0	0.927300	1.263000	2.430000	5.473000
Lake	0	0	0	0.5908	0	6
Sport field	0	0	0	0.514300	0.878300	2.945000
Agriculture field	0	0	0	0.191200	0	2.921000
Scrapland	0	0	0	0.560400	1.050000	2.976000
Churchyard	0	0	0	0.157800	0	2.931000

A.3 Descriptive statistics - dummy variables

Variable	Variable.0	Variable.1
Low Basement	5618	695
Renovation 1970s	5734	579
Renovation 1980s	5774	539
Renovation 1990s	5948	365
Renovation 2000s	6226	87
Renovation after	5895	418
Built before 1927	5368	945

Variable	Variable.0	Variable.1
Built between 1927 and 1939	5605	708
Built between 1939 and 1955	5604	709
Built between 1955 and 1975	3761	2552
Built between 1975 and 1999	5193	1120
Brick	297	6016
Tile roof	4903	1410
Fiber board roof	2716	3597
Terraced house	5135	1178
Hasseris	4953	1360

B Full model estimation

B.1 GLM - Linear non-spatial model

Tab. 12: Generalized Linear model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	8.503690	0.597222	14.238733	0.000000
Garden	0.000177	0.000017	10.730968	0.000000
Size (log)	0.522822	0.018421	28.381871	0.000000
Room (log)	-0.000636	0.015259	-0.041711	0.966729
Basement	0.001143	0.000084	13.571994	0.000000
Low basement	0.088707	0.009977	8.890984	0.000000
Number of floors	-0.012809	0.011377	-1.125858	0.260226
Renovation 1970s	0.000210	0.008647	0.024293	0.980619
Renovation 1980s	0.025942	0.008942	2.901125	0.003718
Renovation 1990s	0.068039	0.010756	6.325475	0.000000
Renovation 2000s	0.131813	0.021528	6.122813	0.000000
Renovation after	-0.141583	0.013475	-10.507057	0.000000
Built before 1927	0.789417	0.032010	24.661863	0.000000
Built between 1927 and 1939	0.829979	0.030961	26.807278	0.000000
Built between 1939 and 1955	0.812930	0.030418	26.725557	0.000000
Built between 1955 and 1975	0.860642	0.028214	30.504192	0.000000
Built between 1975 and 1999	1.016630	0.026994	37.661832	0.000000
Brick	0.078805	0.013950	5.649174	0.000000
Tile roof	0.030999	0.010555	2.936875	0.003315
Fiber board roof	-0.057790	0.007311	-7.904519	0.000000
Highway	-0.000104	0.000028	-3.746664	0.000179
Large road	-0.032235	0.007372	-4.372372	0.000012
Railway tracks	-0.000217	0.000306	-0.710156	0.477607
Industrial area	-0.000061	0.000021	-2.956286	0.003114
Coastline	0.005506	0.000403	13.661463	0.000000
Terraced house	0.012938	0.012402	1.043246	0.296834
Hasseris	0.239948	0.009459	25.367792	0.000000
Spatial lag: Garden	-0.000009	0.000019	-0.481783	0.629960
Spatial lag: Brick	0.167709	0.035322	4.747998	0.000002
Spatial lag: Age	0.000860	0.000293	2.932759	0.003360
Spatial lag: Tile roof	0.068936	0.017107	4.029794	0.000056
Spatial lag: Renovation 1970s	0.064540	0.035700	1.807867	0.070627
Spatial lag: Renovation 1980s	0.106024	0.037897	2.797727	0.005146
Spatial lag: Renovation 1990s	0.141692	0.043758	3.238076	0.001203
Spatial lag: Renovation 2000s	0.123677	0.038286	3.230375	0.001236
City center	0.000099	0.000005	19.192674	0.000000
Park^2	0.001831	0.000458	4.000440	0.000063
Nature^2	0.000739	0.000339	2.182425	0.029078
Lake^2	0.000132	0.000496	0.266916	0.789534
Scrapland^2	-0.006646	0.001703	-3.903028	0.000095
Sport field^2	0.000654	0.001647	0.396990	0.691375
Churchyard^2	-0.000395	0.002764	-0.142838	0.886418
Agriculture field^2	-0.003342	0.002381	-1.403707	0.160406
Common area	-0.004235	0.001436	-2.948557	0.003193

B.2 Spatial Fixed effect model

Tab. 13: Spatial fixed effect model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	9.210188	1.378333	6.682123	0.000000
Garden	0.000176	0.000026	6.728654	0.000000
Size (log)	0.516506	0.028923	17.858248	0.000000
Room (log)	0.001003	0.017259	0.058094	0.953674
Basement	0.001118	0.000092	12.140897	0.000000
Low basement	0.088978	0.010552	8.432518	0.000000
Number of floors	-0.012378	0.021753	-0.569015	0.569346
Renovation 1970s	0.001294	0.009841	0.131444	0.895424
Renovation 1980s	0.027830	0.010154	2.740878	0.006128
Renovation 1990s	0.066238	0.013634	4.858306	0.000001
Renovation 2000s	0.135417	0.028856	4.692878	0.000003
Renovation after	-0.143725	0.012471	-11.524720	0.000000
Built before 1927	0.762143	0.065237	11.682602	0.000000
Built between 1927 and 1939	0.802979	0.062859	12.774212	0.000000
Built between 1939 and 1955	0.784160	0.070671	11.095945	0.000000
Built between 1955 and 1975	0.830159	0.067947	12.217670	0.000000
Built between 1975 and 1999	0.987368	0.062554	15.784365	0.000000
Brick	0.078455	0.013975	5.613848	0.000000
Tile roof	0.032752	0.011844	2.765370	0.005686
Fiber board roof	-0.056577	0.012822	-4.412523	0.000010
Highway	-0.000096	0.000080	-1.205104	0.228163
Large road	-0.038820	0.010864	-3.573279	0.000353
Railway tracks	-0.000316	0.000498	-0.633840	0.526185
Industrial area	-0.000076	0.000060	-1.252499	0.210388
Coastline	0.001630	0.003496	0.466245	0.641040
Terraced house	0.009013	0.021211	0.424904	0.670907
Hasseris	0.137266	0.051585	2.660976	0.007791
Spatial lag: Garden	-0.000025	0.000040	-0.633455	0.526437
Spatial lag: Brick	0.170004	0.066984	2.537965	0.011150
Spatial lag: Age	0.000670	0.000716	0.936696	0.348915
Spatial lag: Tile roof	0.060360	0.047890	1.260389	0.207529
Spatial lag: Renovation 1970s	0.072911	0.058342	1.249719	0.211402
Spatial lag: Renovation 1980s	0.093634	0.047051	1.990066	0.046584
Spatial lag: Renovation 1990s	0.107515	0.066020	1.628534	0.103412
Spatial lag: Renovation 2000s	0.085335	0.041058	2.078396	0.037673
City center	0.000064	0.000030	2.121874	0.033848
Park^2	0.003258	0.000746	4.367565	0.000013
Nature^2	0.000233	0.000907	0.256811	0.797325
Lake^2	0.000992	0.001012	0.979749	0.327210
Scrapland^2	-0.006308	0.002240	-2.816162	0.004860
Sport field^2	0.001808	0.002047	0.883376	0.377033
Churchyard^2	0.002706	0.004872	0.555486	0.578562
Agriculture field^2	-0.006475	0.005482	-1.181123	0.237554
Common area	-0.004854	0.002599	-1.868073	0.061752

B.3 SEM - spatial error model

Tab. 14: Spatial error model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	8.565638	0.600863	14.255557	0
Garden	0.000190	0.000013	14.961683	0
Size (log)	0.470979	0.016441	28.646423	0
Room (log)	0.007191	0.014577	0.493313	0.621791
Basement	0.001030	0.000088	11.770959	0
Low basement	0.082379	0.009382	8.780342	0
Number of floors	0.006436	0.012564	0.512248	0.608478
Renovation 1970s	0.001895	0.009533	0.198771	0.842442
Renovation 1980s	0.029662	0.009649	3.074133	0.002111
Renovation 1990s	0.068817	0.011429	6.021188	0
Renovation 2000s	0.137674	0.022252	6.186991	0
Renovation after	-0.142236	0.011334	-12.549990	0
Built before 1927	0.737451	0.022706	32.478629	0
Built between 1927 and 1939	0.780255	0.022522	34.643855	0
Built between 1939 and 1955	0.757792	0.022209	34.120598	0
Built between 1955 and 1975	0.799521	0.020256	39.471132	0
Built between 1975 and 1999	0.972320	0.020173	48.198109	0
Brick	0.081458	0.013908	5.856929	0
Tile roof	0.040255	0.009400	4.282475	0.000018
Fiber board roof	-0.045686	0.007501	-6.090953	0
Highway	-0.000126	0.000049	-2.575456	0.010011
Large road	-0.030382	0.009622	-3.157504	0.001591
Railway tracks	-0.000460	0.000470	-0.978550	0.327803
Industrial area	-0.000053	0.000036	-1.460393	0.144182
Coastline	0.006270	0.000680	9.215936	0
Terraced house	-0.000895	0.011838	-0.075590	0.939746
Hasseris	0.253914	0.015441	16.444185	0
Spatial lag: Garden	-0.000042	0.000022	-1.970207	0.048815
Spatial lag: Brick	0.085943	0.047776	1.798883	0.072037
Spatial lag: Age	0.000986	0.000296	3.329734	0.000869
Spatial lag: Tile roof	0.046096	0.020203	2.281570	0.022515
Spatial lag: Renovation 1970s	0.069076	0.040271	1.715284	0.086293
Spatial lag: Renovation 1980s	0.130640	0.044471	2.937637	0.003307
Spatial lag: Renovation 1990s	0.068966	0.049750	1.386259	0.165668
Spatial lag: Renovation 2000s	0.073930	0.045307	1.631762	0.102730
City center	0.000108	0.000008	13.133140	0
Park^2	0.001881	0.000737	2.552715	0.010689
Nature^2	0.000994	0.000527	1.884669	0.059475
Lake^2	0.000539	0.000807	0.668571	0.503769
Scrapland^2	-0.005306	0.002466	-2.151599	0.031429
Sport field^2	0.000088	0.002408	0.036402	0.970962
Churchyard^2	0.003795	0.004053	0.936345	0.349095
Agriculture field^2	-0.003970	0.003731	-1.064000	0.287329
Common area	-0.003597	0.001442	-2.494973	0.012597

B.4 GAM - Generalized additive model

Tab. 15: Generalized Additive Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.441177	0.545457	19.142058	0.000000
Garden	0.000176	0.000012	14.156827	0.000000
Size (log)	0.490631	0.016404	29.909151	0.000000
Room (log)	0.008223	0.014679	0.560183	0.575375
Basement	0.001047	0.000086	12.106932	0.000000
Low basement	0.085165	0.009387	9.072418	0.000000
Number of floors	-0.012972	0.011059	-1.172941	0.240864
Renovation 1970s	0.001449	0.009838	0.147279	0.882917
Renovation 1980s	0.025724	0.009869	2.606706	0.009164
Renovation 1990s	0.067982	0.011771	5.775617	0.000000
Renovation 2000s	0.127254	0.022827	5.574606	0.000000
Renovation after	-0.144840	0.011630	-12.454045	0.000000
Built before 1927	0.743429	0.021561	34.480999	0.000000
Built between 1927 and 1939	0.791304	0.021170	37.379191	0.000000
Built between 1939 and 1955	0.763521	0.020649	36.976943	0.000000
Built between 1955 and 1975	0.798661	0.018605	42.927178	0.000000
Built between 1975 and 1999	0.975576	0.017870	54.594239	0.000000
Brick	0.080204	0.013834	5.797618	0.000000
Tile roof	0.030475	0.009404	3.240631	0.001199
Fiber board roof	-0.051833	0.007290	-7.110271	0.000000
Highway	-0.000023	0.000039	-0.585887	0.557973
Large road	-0.040075	0.007526	-5.324475	0.000000
Railway tracks	0.000051	0.000386	0.131111	0.895692
Industrial area	-0.000068	0.000029	-2.366173	0.018003
Coastline	-0.020387	0.005779	-3.527591	0.000422
Terraced house	-0.004910	0.010804	-0.454473	0.649504
Hasseris	0.006180	0.032625	0.189436	0.849757
Spatial lag: Garden	-0.000045	0.000019	-2.399811	0.016433
Spatial lag: Brick	0.090979	0.037178	2.447140	0.014427
Spatial lag: Age	0.000565	0.000262	2.157091	0.031037
Spatial lag: Tile roof	-0.009284	0.018046	-0.514484	0.606932
Spatial lag: Renovation 1970s	0.024150	0.035625	0.677891	0.497866
Spatial lag: Renovation 1980s	0.082939	0.038211	2.170564	0.030002
Spatial lag: Renovation 1990s	0.073418	0.044315	1.656743	0.097622
Spatial lag: Renovation 2000s	0.065418	0.039060	1.674832	0.094017
Park^2	0.002296	0.000561	4.092079	0.000043
Nature^2	0.001474	0.000441	3.344189	0.000830
Lake^2	0.001467	0.000762	1.926020	0.054147
Scrapland^2	-0.007624	0.001867	-4.082655	0.000045
Sport field^2	0.001363	0.001632	0.835348	0.403554
Churchyard^2	0.005837	0.002665	2.190105	0.028554
$\text{Agriculture field}^2$	-0.003020	0.002784	-1.084734	0.278082
Common area	-0.002167	0.001237	-1.752392	0.079756