EUROPEAN COMMISSION
European Research Area

SEVENTH FRAMEWORK PROGRAMME

Funded under Socio-economic Sciences & Humanities

Deliverable 7.1

Spatial Issues on a Hedonic Estimation of Rents in Brussels.

Alain Pholo Bala
Dominique Peeters
Isabelle Thomas

Université Catholique de Louvain, Belgium

Revision: final
FP7-244557
15/08/2011

SustainCity

EUROPEAN COMMISSION
European Research Area
SEVENTH FRAMEWORK PROGRAMME
Funded under Socio-economic Sciences & Humanities

Contents

Spatial Issues on a Hedonic Estimation of Rents in Brussels.

Alain Pholo Bala	Dominique Peeters	Isabelle Thomas
Department of Economics,	CORE,Université Catholique	CORE,Université Catholique
University of Johannesburg.	de Louvain.	de Louvain.
phone: +27-11-559-3811		
fax: +27-11-559-3039	phone: +32-10-47-4343	phone: +32-10-47-2136
apholo@uj.ac.za	fax: +32-10-47-4301	fax: +32-10-47-4301
	dominique.peeters@uclouvain.be	

15/08/2011

Abstract

This paper estimates a hedonic regression on Brussels metropolitan area data. It assesses the impact of the distortions arising either from the choice of a specific zoning system, which is also known as the Modifiable Areal Unit Problem (MAUP), or from the choice of the delineation of the study area. We also evaluate the potential biases that may arise when spatial effects are not taken into account. To do so, we rely on a spatial econometrics model that captures spatial dependence. Given that in the official database, rent is a categorical variable whose different modalities represent distinct intervals of values, our estimation strategy implies the extension of the basic interval regression model. We achieve this extension through a Spatial Autoregressive Model. We find out that estimations results are sensitive to the MAUP as well as to the choice of the delineation of the study area. Moreover, through our Spatial Autoregressive Interval Regression model, we obtain a significant spatial dependence parameter which outlines that there is evidence of substantive spatial dependence.

Keywords

MAUP; Interval Regression; Spatial dependence; Brussels

Preferred citation style

Pholo Bala, A., Peeters, D. and Thomas, I. (2011) Spatial Issues on a Hedonic Estimation of Rents in Brussels., *SustainCity Deliverable*, **7.1**, Université Catholique de Louvain, Belgium.

1 Introduction

This paper is a contribution developed for the SustainCity Project whose goal is to advance the state-of-the-art in the field of micro-simulation of prospective integrated models of Land-Use and Transport (LUTI). The SustainCity Project requires a massive amount of geographical data, collected from several sources and often available at different spatial scales. Hence, choices have to be made about the relevant underlying basic spatial units (BSU), as well as the definition(s) of the studied area. These choices are likely to influence or even bias econometric results. Our objective in this paper is to analyze those biases and to conduct sensitivity analyses.

To address these goals we focus on the hedonic regression model. Before the seminal contribution of Rosen (1974), only few structural interpretations of the hedonic method were available. His model suggests a method that can identify the underlying structural parameters of interest. Econometrically, Rosen's hedonic model implies a two–step procedure: firstly, a hedonic price function is estimated by regressing product price on characteristics. In the second stage, these marginal prices and consumers' socioeconomic characteristics are used to estimate the parameters of the equation representing consumers' behaviour (Picard *et al.* (2010); Rosen (1974)).

A myriad of contributions have dealt with first–step hedonic regression. Four categories of concerns have arisen regarding this framework: (1) functional form, (2) identification, (3) statistical efficiency, and (4) benefit estimation (Kim *et al.* (2003)). The importance of spatial effects, spatial dependence and spatial heterogeneity, on the efficiency and consistency of hedonic model estimates has only recently started to receive some attention.

Yet, spatial dependence is one of the main methodological problems that has to be tackled in first–stage hedonic regression. In general terms, it may be "considered as the existence of a functional relationship between what happens at one point in space and what happens elsewhere" (Anselin (1988)). Many recent hedonic price analyses suggest that in a cross–sectional hedonic price analysis, the value of a property in one location may also be affected by the value of other properties located in its neighboring area (Yusuf (2004)).

Two broad causes may lead to spatial dependence: the nuisance and the substantive spatial dependence (Magrini (2004)). The nuisance spatial dependence refers to the byproduct of measurements errors for observations in contiguous spatial units. In several cases data are collected only at aggregate scale. Because it implies a poor correspondance between the spatial scope of the phenomenon under scrutiny and the delineation of the spatial units of observations, it may entail measurement errors. Those errors will tend to spill over across the frontiers of spatial entities as one may expect that errors for observations in one spatial unit are likely to be correlated with errors of neighboring geographical entities (Anselin (1988)).

Such measurement errors may be caused by bad choices of either the aggregation scale or the delineation of the study area. The aggregation of spatial data is not benign regarding statistical inference. The question of the sensitivity of statistical results to the choice of a particular zoning system is well known as the Modifiable Areal Unit Problem (hereafter MAUP). This issue has been raised by Gehlke and Biehl (1934). However, up to Briant *et al.* (2010) economists paid little attention to this problem. The choice of the frontier of the study area also have potential impact on statistical results. In this paper, we investigate both those issues and we give more details on them in the next section.

The substantive spatial dependence is a more fundamental cause of spatial dependence due to varieties of interdependencies across space. Location and distance do matter and formal frameworks proposed by spatial interaction theories, diffusion processes, and spatial hierarchies structure the dependence between phenomena at different locations in space (Anselin (1988)). Spatial heterogeneity is related to the lack of stability over space of the behavioral or other relationships under scrutiny. It implies that functional forms and parameters vary with location and are not homogenous across the dataset. Several factors, such as central place hierarchies, the existence of leading and lagging regions, vintage effects in urban growth, etc., suggest modeling strategies considering the particular characteristics of each location or spatial entity (Anselin (1988)).

It has been amply demonstrated that the neglect of spatial considerations in econometric models not only affects the magnitudes of the estimates and their significance, but may also lead to serious errors in the interpretation of standard regression diagnostics such as tests for heteroskedasticity (Kim *et al.* (2003)).

In this paper, we focus mostly on spatial dependence by considering one of the main components of the spatial econometrics toolbox: the Spatial AutoRegressive Model (SAR). Several contributions investigate on the spatial dependence issue through the estimation of SAR and Spatial Error Model (SEM). Kim *et al.* (2003), Löchl and Axhausen (2009) directly estimate SAR and SEM models. To avoid dealing with inversion of large matrices as required in maximum likelihood methods, Gawande and Jenkins-Smith (2001) estimate a SAR model by treating the spatial lag of the dependent variable as a regressor. Nevertheless, spatial lag is constructed only with chronologically previous values of the dependent variables to avoid endogeneity problems. Brasington and Hite (2005) estimate a Spatial Durbin Model, that is a model including spatial lags of the dependent variable and of the explanatory variables.

In most of these contributions, the dependent variable (house price or dwelling rent) is continuous. In this paper, we have to face with an extra problem: the information about the dependent variable (here: dwelling rent) is collected through a categorical variable. Each modality of this discrete variable refers to a unique interval of dwelling rents. Therefore, we have to resort on techniques designed to estimate spatially dependent discrete choice models.

LeSage and Pace (2009) provide a detailed overview of spatially dependent discrete choice models. From all those models, the ordered spatial probit model is the one that proposes the modelling strategy that is the closest to the one we have to implement. However, there are important differences between our "Spatial Interval Regression" model and the ordered spatial probit model. In the ordered probit model, the cut points separating interval of the latent variable are unknown. Therefore, there is an identification issue and the variance has to be normalized to one so that regression coefficients as well as cut points may be estimated. In our model the vector of boundaries of the dependent variable is known. Therefore, regression coefficients as well as the variance may be jointly estimated. Thus, this paper aims at developing an estimation strategy adapted to the specificity of our model and at empirically testing it on Brussels.

This paper is organized as follows. The next section will describe more deeply MAUP and urban area delineation issues. The third section is devoted to a detailed presentation of the estimation strategy. In the fourth section data used for estimation are presented. Section 5 presents the results of estimations of the different specifications of our Interval Regression model and section 6 concludes the paper.

2 Spatial issues.

Spatial dependence may occur because of measurement errors due to choices of either the aggregation scale (MAUP issue) or of the delineation of the study area. The MAUP outlines that changes in either the size (equivalently the number) of spatial units or their shape (equivalently the drawing of their boundaries) may alter the estimates of any statistical analysis based on spatial data.

Several contributions have assessed the impact of the MAUP on multivariate statistics. Gehlke and Biehl (1934) were the first to emphasize that simple statistics such as correlation coefficients could vary substantially with changing zoning systems. They outline the tendency for correlation coefficient to increase in size as the size of spatial units increases. Fotheringham and Wong (1991) demonstrated that the behavior of parameter estimates for a multiple linear regression model becomes much more complex and unpredictable corresponding with changes in both the scale at which data are collected and zone definition at a particular scale.

However, clear theoretical foundations are not obvious (Briant *et al.* (2010)). Amrhein (1995) is the first to suggest separating aggregation effects from other types of discrepancies, such as model mis–specification in multivariate settings. Consequently, he reached a less alarming conclusion than Fotheringham and Wong (1991), and suggested that, for well–specified models, such as Amrhein and Flowerdew (1992), aggregation does not imply too many distortions.

Briant *et al.* (2010) also perform a simulation exercise to analyze the behavior of simple regression coefficients. Concerning the size distortion, they find that if aggregation distortion on explanatory and the dependent variables are similar, the size effect of the MAUP will be small. Such a condition holds when both the explanatory and dependent variables are spatially autocorrelated and averaged. A contrario, the size issue is more disturbing when the dependent and the explanatory variables are not aggregated by the same process or do not display the same level of spatial autocorrelation.

Concerning the shape distortion, Briant *et al.* (2010) consider it as an error–in variables issue. Therefore, in case of a change in shape, aggregation yields a biased estimate of the regression coefficient. The larger are the variations in borders, the larger is the shape effect. Improving the specification or correcting the endogeneity of the regressor by instrumental variables techniques should mitigate shape distortions. However, those solutions do not alleviate bias induced by spatial autocorrelation. Indeed, in case of spatial correlation between the regressor and the regressand the bias increases.

Briant *et al.* (2010) build on previous results to further extend the MAUP literature in several ways. Firstly, they evaluate the relative importance of size and shape distortions comparatively to misspecification biases in the estimation of spatial concentration, agglomeration economies, and trade determinants. Secondly, they investigate different aggregation processes to test the sensitivity of economic inference to the MAUP. Finally, they complete the contribution of Fotheringham and Wong (1991) by comparing the estimates from six different administrative and grid zoning systems to those from a hundred equivalent random systems. They found that at large scales the size effect of MAUP might be important. However, at low scales they are pretty weak comparatively to misspecification issues. Concerning shape biases, they are weaker than both misspecification and size biases.

The choice of the delineation of the study area and its potential impact on statistical results is another geographical issue that desserves attention. Our contribution focuses on Brussels metropolitan area. But several other delineations may be considered for Brussels: administrative delineations, morphological delineations (Donnay and Lambinon (1997); Tannier *et al.* (2010); Van Hecke *et al.* (2009)), functional delineations (Cheshire (2010); Van Hecke *et al.* (2009); Vandermotten *et al.* (1999)), etc. While each way of defining Brussels agglomeration may be consistent according to a given standpoint, considering administrative definitions can be harmful since administrative borders do not capture the essence of economic phenomena and transportation issues that often spill over boundaries.

In a literature review on regional convergence, Magrini (2004) asserts that the use of administratively defined regions raises two fundamental problems: on the one hand, since output is measured at workplaces while population at residences, the measured levels of per capita income will be highly misleading. Moreover, processes of decentralisation or recentralisation of residences relative to workplaces is likely to affect per capita income growth rates for administratively defined regions. Those problems induce measurements errors that may be characterized as nuisance spatial dependence. Using functionally defined regions may mitigates reduces those biases.

In this paper, we further extend MAUP literature by investigating on the impacts of choices of the size of spatial units on the results of the basic Interval Regression Model. However, since estimating the “Spatial Interval Regression Model" on thousands of observations is too demanding, we do not assess the impact of the choice of those geographical issues on spatial econometrics results.

3 Estimation strategy.

In this paper, we estimate the hedonic model by means of interval regression and we analyze spatial effects. As in Kim *et al.* (2003), the spatial econometric aspects of the analysis are among our principal concerns.

The choice of the functional form is another critical issue. Indeed, economic theory does not suggest a specific functional form (Cassel and Mendelsohn (1985); Halvorsen and Pollakowski (1981); Goodman (1978); Picard *et al.* (2010)). Therefore, several functional forms have been proposed in the literature, including Box–Cox transformations, some of their special cases — like semi–log, log–log, translog, linear, quadratic, square root quadratic, generalized square root quadratic, generalized leontief (Halvorsen and Pollakowski (1981); Picard *et al.* (2010)). Box–Cox flexible functional form have been recommended on the ground that they have the best performance in terms of goodness of fit tests (Halvorsen and Pollakowski (1981)).

But such specifications are not readily implemented in the presence of spatial dependence (Kim *et al.* (2003)). Therefore, in this section we will present the methodological aspects of interval regression and of spatial econometrics models with a categorical variable whose different modalities correspond to different intervals of rent.

3.1 Benchmark model: interval regression.

In the Belgian Social and Economic Survey the information on rent prices has been collected through a categorical variable (cfr. 4.1). Therefore, that survey does not give the actual value of the rent price observation y_i^*; it just provides the value y_i of a categorical variable from which we can infer the interval where y_i^* lies:

$$y_i = j \text{ if } \alpha_{j-1} < y_i^* \leq \alpha_j$$

where $j \in \{1, ..., J\}$, with $J = 5$ the number of intervals, and $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_J)$ is a given vector of boundaries with $\alpha_0 \leq \alpha_1 \leq \cdots \leq \alpha_J$.

To estimate this model, without taking spatial effects into account, we rely on “interval regression". This model is close to the ordered probit model from a computational perspective, but it is conceptually different, since it may be interpreted as an extension of censored regression.[1]

[1] Extreme values of the categories on either end of the range are either left–censored or right–censored. The other categories are interval censored, that is, each interval is both left and right censored. Source: SAS Data Analysis Examples, Interval Regression. UCLA: Academic Technology Services, Statistical Consulting Group from `http://www.ats.ucla.edu/stat/sas/dae/intreg.htm` (last access June 17, 2011).

In such a framework, $y^* = (y_1^*, y_2^*, \cdots, y_N^*)'$ is a variable that has a quantitative meaning and not just a latent variable with only an ordinal signification, as in the ordered probit model (Wooldridge (2002)).

As Geoghegan *et al.* (1997), we opt for double log estimation. This functional form has the clear advantage to ease the interpretation of the estimated coefficients of continuous variables. Those coefficients are elasticities: the percent change of the regressand given a percent change in a continuous regressor. Therefore, in this model we are interested in estimating $E(ln(y_i^*)|x) = x_i'\beta$ where x_i denotes a vector of dummies and of logarithmic transformations of continuous variables. If we could know the actual value of y^*, we could use OLS to estimate β.

We can use the normality assumption to work out the probability that the regressand lies in any interval $[\alpha_{j-1}, \alpha_j]$ (Koop (2003)). With the regression model $\tilde{y}_i = x_i'\beta + \epsilon_i,$ where $\tilde{y}_i = ln(y_i^*)$, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ we have

$$\begin{aligned} Pr(y_i = j|\beta, \alpha) &= Pr(ln(\alpha_{j-1}) < \tilde{y}_i \leq ln(\alpha_j)|\beta, \alpha) \\ &= Pr(ln(\alpha_{j-1}) < x_i'\beta + \epsilon_i \leq ln(\alpha_j)|\beta, \alpha) \\ &= Pr(ln(\alpha_{j-1}) - x_i'\beta < \epsilon_i \leq ln(\alpha_j) - x_i'\beta|\beta, \alpha) \end{aligned}$$

Since $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, we have

$$Pr(y_i = j|\beta, \alpha) = \Phi(ln(\alpha_j) - x_i'\beta) - \Phi(ln(\alpha_{j-1}) - x_i'\beta)$$

with $\Phi(.)$ denoting the cumulative normal distribution function.

Therefore, we can estimate the parameters of β and σ^2 by maximizing likelihood after having defined the log–likelihood function for each observation i by

$$l_i(\beta, \sigma^2) = \sum_{j=1}^{J} I[y_i = j]\, ln[\Phi(ln(\alpha_j) - x_i'\beta) - \Phi(ln(\alpha_{j-1}) - x_i'\beta)] \qquad (1)$$

While this computational procedure is very similar to the one used in the classical ordered probit model, we may recall some important differences that desserve to be noticed. In the ordered probit model, the vector α is an ordered set of unknown cut points. Therefore, there is an identification issue in the ordered probit model and σ^2 is normalized to one so that the model can estimate β and σ^2. In interval regression α is rather a set of known interval boundaries, thus, β and σ^2 may be jointly estimated.

3.2 Spatial Interval Regression Models.

To capture spatial dependence, we may extend the basic Interval Regression model by the following specification:

$$\begin{aligned}\tilde{y} &= \rho W_1\,\tilde{y} + X\beta + u \\ u &= \lambda W_2\,u + \epsilon \\ \epsilon &\sim \mathcal{N}\left(0, \sigma^2 I_N\right)\end{aligned} \tag{2}$$

where $\tilde{y} = ln\,(y^*)$, N is the number of observations, X is a $N \times k$ matrix, ρ is the spatial dependence parameter, W_1 and W_2 are $N \times N$ standardized spatial weight matrices.

W_1 and W_2 tell us whether any pair of observations are neighbors. For example, supposing that $W_1 = W_2 = W$, if house i and house j are neighbors then, $w_{ij} = 1$ and zero otherwise. Whether or not any pair of houses are neighbors is based on whether or not they are located in the same geographical entity or in neighboring spatial units. We follow Yusuf (2004) and Kim *et al.* (2003) by considering two spatial units as neighbors when they share common borders.

We assume that there are S spatial entities. Any spatial entity i is populated by N_i individuals, with $\sum_{i=1}^{S} N_i = N$. Therefore, if $y^*_{N_i}$ denotes the $N_i \times 1$ vector of rents paid by the N_i households living in the ith spatial entity, the $N \times 1$ vector of rents paid by all the households of the sample is $y^* = \left({y^*_{N_1}}', \cdots, {y^*_{N_i}}', \cdots, {y^*_{N_S}}'\right)'$.

While much has been written on the techniques for dealing with spatial dependence in continuous econometric models, the study of spatial dependence in discrete choice models has received less attention in the literature. This is clearly due to the added complexity that spatial dependence introduces into discrete choice models and the subsequent need for more complex estimators.

There are several techniques to estimate this spatially dependent discrete choice model (Flemming (2004)). Those techniques have to solve two problems inherent to standard discrete choice model: the inconsistency due to the heteroskedasticity caused by spatial dependence and the efficiency implications of not using all the information in the non–spherical variance–covariance matrix. Some authors have attempted to address the heteroskedasticity issue through innovative specification of spatial dependence (Case (1992)). Others have used a Generalized Method of Moments (GMM) technique that uses the spatial structure to determine the heteroskedastic variance terms (Pinkse and Slade (1998)).

Correcting exclusively on heteroskedasticity may help to obtain consistent estimates. This has also the advantage to reduce the problem of estimating an N–dimensional integral to the estimation of the simpler product of independent density functions. However, those estimates are not efficient since this procedure does not use all the information in the off–diagonal elements of the variance–covariance matrix. In order to address the heteroskedasticity generated by spatial dependence while using the information in the off–diagonal terms of the variance–covariance matrix, one has to solve the problem of multidimensional integration. The Expectation Maximization (EM) algorithm, the Recursive Importance Sampling (RIS), and Bayesian techniques, like Gibbs Sampling, offer solutions to this problem.

An alternative to the aforementioned techniques is to describe the spatially dependent discrete choice problem as a weighted non–linear version of the linear probability model with a general variance–covariance matrix. This approach avoids the higher order integration problem and the computation of the N by N determinants.

In this paper we opt for the Gibbs Sampling approach. Indeed, while providing results that are similar to those of the RIS simulator, it is computationally and conceptually more simple (Bolduc *et al.* (1997)). Moreover, the Gibbs Sampler's method overcomes the problem encountered in the estimation of standard errors by the EM algorithm because estimates standard errors are derived directly from the posterior parameters distributions. Finally, Gibbs Sampler estimation do not face the drawback of weighted non–linear least squares estimators. Those estimators treat the spatial error autoregressive parameter as a nuisance parameter and are therefore unable to provide standard error estimates (Flemming (2004)).

Specification (2) is consistent with both the SAR and SEM models. If $W_1 = 0$, then (2) collapses to the SEM; if $W_2 = 0$ then (2) corresponds to the SAR model. While the SAR and the SEM models are quite similar mathematically, the logic underlying each model's structure is somewhat distinct.

The SAR model implicitly assumes that the spatially weighted average of housing prices in a neighborhood affects the price of each house (indirect effects) in addition to the standard explanatory variables of housing and neighborhood characteristics (direct effects). It is particularly appropriate when there is structural spatial interaction in the market and the modeler is interested in measuring the strength of that relationship. As the assumption of structural spatial interaction is peculiarly relevant in the hedonic regression, it is our favorite modelling strategy. It is also relevant when the modeler is interested in measuring the "true" effect of the explanatory variables, after the spatial autocorrelation has been removed. Indeed, the SAR model is probably the only way to obtain a consistent estimator for the parameter needed to carry out the spatial filtering (Anselin and Bera (1998)).

A contrario, in a SEM model spatial autocorrelation is assumed to arise from omitted variables that follow a spatial pattern (Kim *et al.* (2003)). The SEM is the most appropriate when there is no theoretical or apparent spatial interaction and the modeler is interested only in the correction of spatial autocorrelation (Anselin, 2001). Since, structural spatial interaction is strongly expected in the dwelling market, in this paper we will focus on the SAR model.

For the SAR, specification (2) may be rewritten as:

$$
\begin{aligned}
\tilde{y} &= \rho W \tilde{y} + X\beta + \epsilon \qquad (3) \\
\epsilon &\sim \mathcal{N}\left(0, \sigma^2 I_N\right)
\end{aligned}
$$

3.2.1 Estimation of Spatial Interval Regression Models.

We may express the likelihood for the SAR as

$$
L\left(\tilde{y}, W | \rho, \beta, \sigma^2\right) = \frac{1}{2\pi\sigma^N} \left| I_N - \rho W \right| exp\left\{ -\frac{1}{2\sigma^2} \left(\epsilon' \epsilon\right) \right\} \qquad (4)
$$

where

$$
\epsilon = \left(I_N - \rho W\right) \tilde{y} - X\beta.
$$

Using diffuse priors for (β, σ^2, ρ) results in the following expression of the joint posterior density:

$$
p\left(\beta, \sigma^2, \rho | \tilde{y}, X, W\right) \propto \left| I_N - \rho W \right| \sigma^{-(N+1)} exp\left\{ -\frac{1}{2\sigma^2} \left(\epsilon' \epsilon\right) \right\} \qquad (5)
$$

Estimates of this distribution should be sampled through a Gibbs–sampler with the following 4 steps:

1. Drawing β from $p\left(\beta|\sigma^2_{(0)}, \rho_{(0)}, \tilde{y}_{(0)}\right)$

$$\begin{aligned} \beta|\sigma^2_{(0)}, \rho_{(0)}, \tilde{y}_{(0)} &\sim \mathcal{N}\left(\tilde{\beta}, \sigma^2_{(0)} \left(X'X\right)^{-1}\right); \quad (6)\\ \tilde{\beta} &= \left(X'X\right)^{-1}\left(X'A\tilde{y}\right); \quad (7)\\ A &= I_N - \rho W. \quad (8) \end{aligned}$$

2. Drawing σ from $p\left(\sigma|\beta_{(1)}, \rho_{(0)}, \tilde{y}_{(0)}\right)$

$$\sigma|\beta_{(1)}, \rho_{(0)}, \tilde{y}_{(0)} \sim \sigma^{-(N-1)} exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\}. \quad (9)$$

3. Sample $p\left(\rho|\beta_{(1)}, \sigma^2_{(1)}, \tilde{y}_{(0)}\right)$ by inversion approach (LeSage and Pace (2009)), where

$$p\left(\rho|\beta, \sigma^2, \tilde{y}\right) \propto |A|\, exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\}. \quad (10)$$

4. Drawing $\tilde{y}$ from the $\mathcal{N}\left(\mu, \Omega\right)$ distribution

$$\begin{aligned} \tilde{y}|\beta_{(1)}, \sigma^2_{(1)}, \rho_{(1)} &\sim TMVN\left(\mu, \Omega\right); \quad (11)\\ \mu &= \left(I_N - \rho W\right)^{-1} X\beta; \quad (12)\\ \Omega &= \sigma^2\left[\left(I_N - \rho W\right)'\left(I_N - \rho W\right)\right]^{-1}; \quad (13) \end{aligned}$$

   where TMVN denotes a multivariate truncated normal distribution.

The conditional distribution of ρ does not take a known form as in the case of the conditionals for the parameters β and σ. Therefore, sampling for the parameter ρ must proceed using alternative approach, such as numerical integration and Metropolis-Hastings.

4 Data description.

In hedonic regression models, dwelling rent is characterized as a bundle of several kinds of characteristics (Kim *et al.* (2003); Brasington and Hite (2005)). The first attribute type refers to the structural characteristics of the dwelling i.e. its physical attributes. The second includes neighborhood characteristics such as median income by tax declaration, accessibility to the largest urban centers. The third type of characteristics relates to environmental quality, such as air pollution, proportion of agricultural areas or forests.

In principle, all the features pertinent for the characterization of market prices should be included in a hedonic regression. However, as Butler (1982) noticed, this can not be done in practice for two reasons. Firstly, the number of such characteristics is unmanageably large and data on many of these are either unavailable or of poor quality. Secondly, some explanatory variables may lead to considerable multicollinearity. For those reasons, Butler (1982) states that any estimate of the hedonic relationship is potentially misspecified because some of the relevant explanatory variables must be omitted. He concluded that all estimates are to some extent "incorrect" and differences among them must be attributed at least in part to differences in adaptation to the specification problems common to all. Therefore, the objective generally pursued in hedonic regression models is to find a broad set of statistically significant variables with expected signs, moderate impact of multicollinearity and estimations with a sufficient model fit (Löchl and Axhausen (2009)). The variables used here are selected in that spirit.

4.1 Dwelling structural characteristics.

The Belgian Socio–economic Survey of 2001 is a census that includes several different housing attributes that may be taken into account into a hedonic regression: type of dwelling, number of rooms of a specific kind (separated kitchens, fitted kitchens integrated in other rooms, separated lounges, bedrooms, toilets, bathrooms etc.), total surface of dwelling rooms, building period, renovation, energy or fuel used for heating, dwelling furniture, isolation (double glazing, wall or roof isolation), use of alternative energies, presence and size of a garage, presence of a garden. Most of the variables that can be constructed from those attributes are categorical and qualitative.

There is only one potential quantitative discrete variable: the number of rooms. Information about monthly rents (in euros without charges) has been collected into intervals corresponding to the following categories: 1 for rents inferior to 249.89; 2 for rents between 249.90 and 495.78; 3 for rents between 495.79 and 743.67; 4 for rents between 743.68 and 991,56; 5 for rents larger or equal to 991,57. As discussed previously, this way of coding the rent variable has an impact on the choice of the appropriate estimation strategy.

With some of those physical attributes, we may also contruct a quality index close to the index proposed by Vanneste *et al.* (2007).[2] This index has the following categories: 1 (insufficient quality) for dwellings without toilets or without bathrooms; 2 (basic quality) for dwellings with toilets and bathroom; 3 (good quality) for dwellings which have, in addition to the basic quality, a central heating, a kitchen, and a total surface of dwelling rooms between 35 m^2 and 84 m^2; 4 (good quality and spacious) similar with the preceding category but with a total surface of dwelling rooms between 85 m^2 and 105 m^2; 5 (very good quality) for dwellings fullfilling the requirements of the “good quality" category but with a total surface of dwelling rooms greater than 105 m^2, and with double glazing.

The following variables have been selected:

[2] The difference between our index and the one built by Vanneste *et al.* (2007) is that we do not consider the necessity of at least four important repairs.

Table 1: List of variables linked to dwelling physical attributes.

Variable	Description
Type of dwelling dummies	
APP	dummy for appartments
OTHER	dummy for other kinds of dwelling
SING[1]	dummy for single family dwelling
Dwelling surface dummies	
SURFA	dwelling with a surface lower than 35 m^2
SURFB	dwelling with a surface between 35 and 54 m^2
SURFC	dwelling with a surface between 55 and 84 m^2
SURFD	dwelling with a surface between 85 and 104 m^2
SURFE	dwelling with a surface between 105 and 124 m^2
SURFF[1]	dwelling with a surface higher or equal to 125 m^2
Heating installation dummies	
HEATA	Individual central heating installation
HEATB	Central heating installation common to several dwellings in one building
HEATC	Central heating installation common to several dwellings in several buildings
HEATD[1]	Other heating installations
Dummies related to the composite quality index	
QUALITY1	Insufficient quality
QUALITY2	Basic quality
QUALITY3	Good quality
QUALITY4	Good quality and spacious
QUALITY5[1]	Very good quality
Dummies related to the source of energy used for heating	
FUELA	Gasoil
FUELB	Coal
FUELC	Wood
FUELD	Heat pump
FUELE	Electricity
FUELF	Natural gas

[1] Reference case.

Continued on next page

Table 1 – concluded from previous page

Variable	Description
FUELG	Butane, propane
FUELH[1]	Other source of energy
Parking dummies	
PARKA	No parking available
PARKB	Parking for one car
PARKC[1]	Parking for more than one car
Other dummies	
LOFT	dummy for Studio or loft
RECENTBUILT	dummy for dwelling built after 1981
RENOVATION	dummy for dwelling renovated after 1991
FURNISH	dummy for dwelling furnished
DGLAZING	double glazing dummy
WALLISO	wall isolation dummy
BATHROOM	bathroom dummy
TOILET	toilet dummy
GARDEN	garden dummy
LNROOMS	ln of the number of rooms

[1] Reference case.

4.2 Study area and basic spatial unit.

4.2.1 Study area.

The 2001 Belgian Census includes 906,308 observations on private dwellings (we exclude collective households, households living in caravans, and in social housing) for all Belgium. But, we restrict the focus of our analysis to the private renting market of Brussels. Here comes the first spatial issue as there is no univocal definition of Brussels. Several delineations of the capital of Belgium have been proposed based on different criteria: administrative, morphological (Donnay and Lambinon (1997); Tannier *et al.* (2010); Van Hecke *et al.* (2009)) and functional (Cheshire (2010); Van Hecke *et al.* (2009); Vandermotten *et al.* (1999)). Table 2 and figure 1 present macrozones that are consistent with 8 delineations of Brussels agglomeration. One of those delineation, "Région Bruxelles–Capitale", corresponds to an administrative definition of Brussels agglomeration. Another delineation, Brussels "(operational) agglomeration" corresponds to a morphological definition of Brussels. It is one of the macrozone defined by Van Hecke *et al.* (2009) nomenclature of Belgian urban regions.

The other delineations represent merely functional definitions of Brussels agglomeration. They are macrozones that, because of the strong socio-economic ties of their peripheral rings with Brussels urban center, may serve for defining Brussels urban functional region. From Van Hecke *et al.* (2009) nomenclature, we may consider Brussels urban region on one hand and Brussels residential urban complex on another hand. We may also consider Brussels "extended residential commuting complex" which the union of the Brussels and Leuven residential commuting complexes. Stratec, an independent consultancy company involved in the Sustaincity Project, has proposed other macrozones: "Stratec RER Area" and "Stratec Extended RER Area" on the basis on commuting ties through the railway network transportation system. Those spatial entities are essentially based on the Official RER Area defined by ministerial decree and composed of 126 municipalities (Moniteur Belge, 2004).

The Belgian census includes the information from 177,721 dwellings pertaining to "Région Bruxelles–Capitale" and 330,147 observations in the set of municipalities pertaining to at least one of the most extensive delineations of Brussels. This set is labelled "Union" in table 2. Therefore, "Région Bruxelles–Capitale" concentrates more than half of the rented dwellings of the most extensive definitions of Brussels agglomeration. In figure 1, the "extended residential commuting complex" corresponds to the union of the "Extended residential commuting area" (in lightest blue), the Suburb (in medium blue), and the Agglomeration (in darkest blue). The hatched zone corresponds to "Stratec Extended RER Area" and the small and central area with a black border depicts the "Région Bruxelles–Capitale".

Table 2: Number of observations (dwellings) for eight different delineations of Brussels.

Spatial entity	Description	Nmun	Nobs
“Région Bruxelles–Capitale”	Administrative definition	19	177,721
Agglomeration	Morphological agglomeration adjusted to municipalities boundaries	36	208,371
Urban region	Agglomeration+Suburb	62	233,582
Residential urban complex (RUC)	Urban region+Residential Commuting area	122	283,079
Extended Residential urban complex (ERUC)	Brussels RUC+Leuven RUC	134	301,160
RER area	Area served by the RER	135	314,279
Extended RER area	RER area + 12 municipalities STRATEC study area	147	325,135
“Union”	Municipalities pertaining to at least one delineation	157	330,147

Nmun= number of municipalities, Nobs= number of observations

Source: Van Hecke *et al.* (2009) and STRATEC.

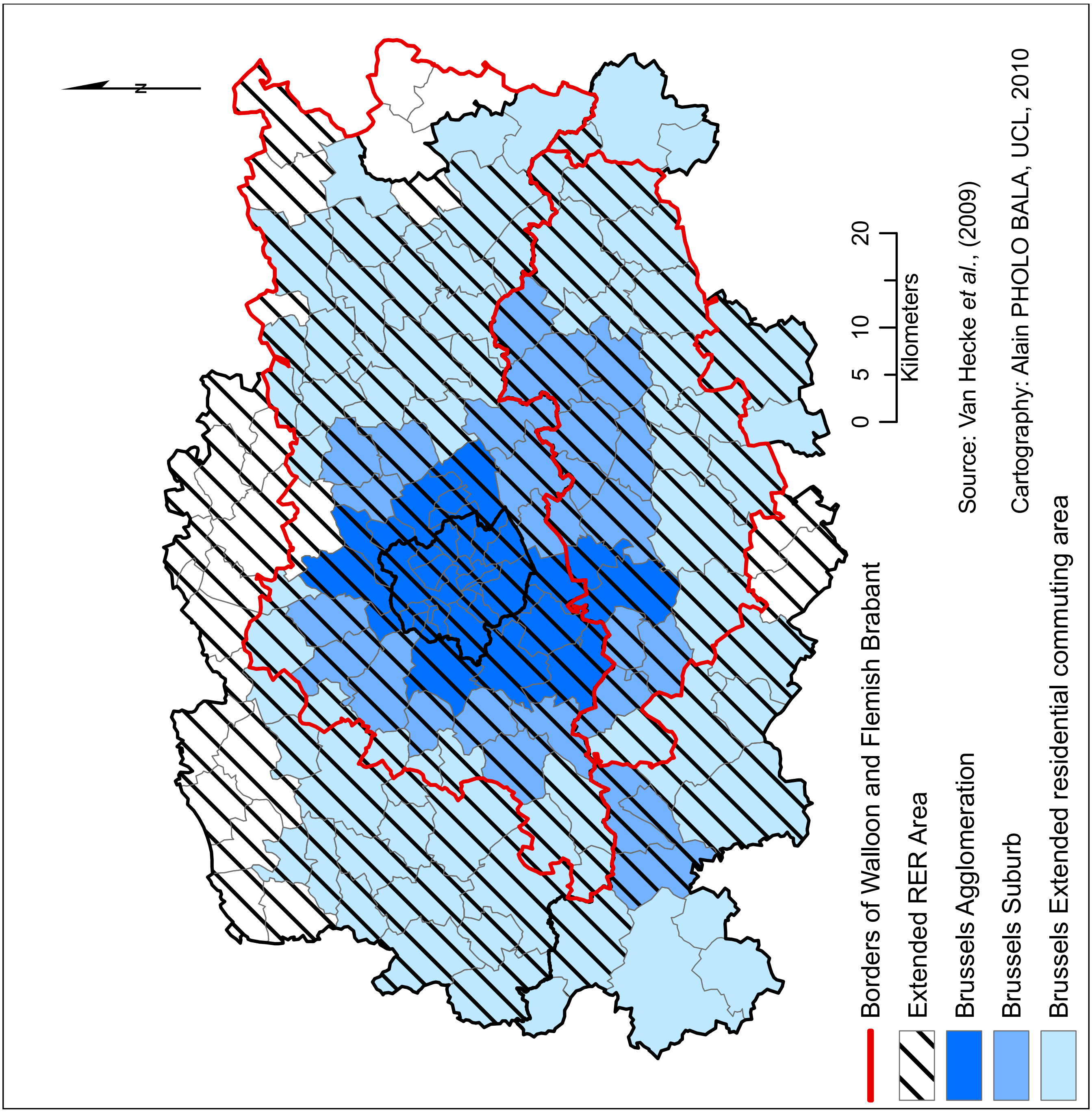


Figure 1: Different limits of Brussels agglomeration

4.2.2 Basic spatial units.

Environment quality attributes and neighborhood characteristics may be measured at different spatial scales. From the less to the more disaggregated level, we may distinguish the following spatial units: the municipality, the former township[3] and the statistical sector. Those spatial units are nested. Belgium is divided in 589 municipalities, 2,616 “anciennes communes” and 20,464 statistical sectors. Variables measured at the municipality level are followed by the label (COM), those measured at the former township level are followed by the label (AC), and those measured at the statistical sector level are followed by the label (SS).

4.3 Environment quality attributes.

4.3.1 Land Cover information.

The Belgian Corine (Coordination of Information on the Environment) Land Cover database provides land cover information classified according to a legend adapted for the whole European continent. It includes 44 themes organized in three hierarchical levels. This information is made available by the European Environment Agency (EEA) (see `http://www.eea.eu.int/products`) at a resolution of 250 m grid cells (minimum mapping unit = 25 ha), and is based on interpretations of remotely sensed photographs taken in the year 2000. The CORINE database was obtained in the form of a raster dataset that was used to produce the following synthetic variables at the commune level:

- Percentage of each municipality and “ancienne commune" covered by Forest (*Percent_Forest*). This is essentially obtained by the aggregation of three CORINE classes: broad–leaved forest, coniferous forest, and mixed forest. This proportion is computed as the percentage of the 250 m by 250 m grid cells entirely covered with forest in each municipality.

- Percentage of each municipality and “ancienne commune" covered by Agriculture (*Percent_Arable*). It is merely based on the aggregation of the arable land, permanent crop, pasture and heterogeneous agricultural areas classes in the CORINE database (class 2). This percentage represents the share of the 250 m by 250 m grid cells entirely covered with arable land in each commune.

[3]“Ancienne commune” in french.

4.3.2 Pollution indicator.

Several hedonic price studies attempt to find out whether air quality is associated with property value. Smith and Huang (1995) performed a good discussion of many of them in their formal summary of the hedonic studies in US from 1967 to 1988. In order to address the issue of whether housing market can value air quality, they used a comprehensive meta–analysis of hedonic property value model. Boyle and Kiel (2001) also realized a review of 12 hedonic studies. Some lessons may be drawn from those papers. Firstly, most of those studies suggest that air pollution affects, negatively, property value. Thus, they suggest that people are willing to pay for air quality improvement.

The Belgian Interregional Cell for the Environment (IRCEL–CELINE) provides information on air quality in all 3 Belgian Regions.[4] IRCEL–CELINE provides air concentration of PM_{10} through a raster file. This raster allowed us to build an indicator of PM_{10} concentration: the average concentration of PM_{10} in every Belgian municipality.

4.3.3 Slope indicator.

The average gradient of the relief, noted SLOPE, is obtained in each statistical sector, each former township and each municipality from a Digital Terrain Model. It is used as a proxy of the average landscape slope and it will be useful to test the assumption that hilly landscapes are more attractive to residents (Goffette-Nagot *et al.* (2010)).

4.4 Neighborhood attributes.

4.4.1 Median and average income.

Localities where most inhabitants have a high social and economic status are characterized by more expensive dwellings.

Data on median income by tax declaration (REVMED) for 2001 were obtained from Belgian National Statistical Institute and are computed at the level of each statistical sector and each municipality. Data on average income by tax declaration (REVMOY) are available for the same year at the level of each former township and each municipality.

[4] available online at `http://www.irceline.be/`.

4.4.2 Accessibility indicators.

Belgium is a densely–populated country with large commuting flows. Its small size and its high population density mean that several employment centres are often reachable from a given place (Goffette-Nagot *et al.* (2010)). Greater accessibility implies an increased quality of life for the individual (greater freedom to choose activities and more time to devote to them). Hence, we may expect an influence of accessibility to employment centers on residential land prices and dwelling rents. As pointed out by Goffette-Nagot *et al.* (2010), there exists no consensus about the definition and formulation of the concept of accessibility in the literature (for a recent review, see Geurs and Ritsema van Eck (2001)). A simple and intuitive measure of the mutual accessibility between two places is the straight–line distance between them. The distances between the centroids of the communes are easy to compute with a Geographical Information System (GIS). However, time units are more relevant in the computation of accessibility indices than than either Euclidian or Manhattan distances expressed in kilometers for several reasons. Firstly, travel time varies according to the type of roads used between two locations (e.g. motorways allow high speeds, which reduce the travelling time). Secondly, time units can include more time components (e.g. waiting time) than the simple journey time. Following the methodology of Vandenbulcke *et al.* (2009), indices of accessibility by car along the road network were computed from any Belgian former township to the five largest Belgian cities. We restrict our computation on this transport mode because in 2001 82% of all commuters journeys are made by car, while public transport only accounts for 14% (the other 4% represent travel made by bus companies). The basic data needed are: road network data, origins and destinations, i.e. the administrative frontiers of all Belgian former townships and the location of their centroids.

5 Results.

In this section, we present estimation results from different specifications. Let us firstly present results from the benchmark model which allows us to investigate on MAUP and delineation issues. Then, we discuss spatial autocorrelation issues with the “Spatial Interval Regression" Model.

5.1 Interval Regression Model (IRM).

Since it is much less computationally demanding than its spatial counterpart, it enables estimations with huge databases of tens thousands of observations. This allows to have enough variation on environment quality and neighborhood attributes to compare estimation results with different basic spatial units (BSU) and with different agglomeration delineations.

5.1.1 Impact of the choice of agglomeration delineation.

Table 3 displays results of the IRM for different macrozones and for a given set of dwelling structural characteristics. Most of the results for different delineations of Brussels agglomeration have the same sign but show differences in magnitude. This suggests that the choice of the limits of agglomeration has an impact on econometrics results.

For the dwelling structural characteristics, most of the results are as expected. The value of a dwelling increases with its number of rooms and its surface. Renovated dwellings and dwellings recently built have higher monthly rents. Lodgings with central heating (especially those with individual central heating) are dearer than those with other heating installations. Dwellings using coal or wood as the source of energy for heating are less expensive.[5] This is not surprising since heating with coal or wood is an indicator of the poor quality of a dwelling (Vanneste *et al.* (2007)). A contrario, lodgings using electricity as the source of energy are more expensive.[6] Dwellings with a garden, bathrooms, toilets, double glazing, and wall isolation are more valuable. Moreover, the more a dwelling have parking places the more it is expensive.

Other results are more puzzling. All other things being equal, appartments are cheaper than single family houses. Except in the Agglomeration where they are lower, rents of other types of dwellings are not significantly lower than those of single family houses. Studios and lofts are,

[5]However, results are not significant for the Agglomeration and the Brussels Capital Region macrozones. This may be explained by the low variance of the variable in the smaller macrozones and the small share of dwellings using coal or in Brussels CBD.

[6]For the Extended Residential Urban Complex, results are not significant.

ceteris paribus, more expensive than other dwellings. Furnished dwellings are less expensive. While this may seems paradoxical, this can be due to the fact that this category of dwellings targets mostly low income social categories such as students who can not afford to furnish and to renovate their lodgings.

Most of the results about environment quality variables and neighborhood attributes are in line with what is generally expected. Rental prices decline with the pollution indicator and with **LNACCE (AC)**, which is a reverse indicator of accessibility. This is not surprising since we expect dwellings located in more accessible and less polluted areas to be more demanded and therefore more valuable. The coefficient of **LNSLOPE (AC)** is positive. This confirms the assumption that hilly landscapes are more attractive to residents.[7] The coefficient obtained for the variable **LNPER_FOREST (AC)** is positive and significant. This indicates that dwellings located in neighborhoods covered by forests are sought-after.

The results obtained with the **LNSLOPE (AC)** and the **LNPER_FOREST (AC)** variables are clearcut. They emphasize that households value positively neighborhoods with environmental amenities. The sign of the **LNPER_ARABLE (AC)** coefficient is more difficult to interpret, it suggests that dwellings located in neighborhoods covered by agricultural areas are less valuable. This may be explained by the fact that such neighborhoods are deprived of infrastructures and amenities (schools, shopping centers, etc.) that are required by most households or that agriculture entails some negative neighborhood externalities. Finally, the coefficient of the variable **LNREVMED** is positive and significant, indicating that dwellings located in wealthy neighborhoods are more expensive.

The standard error estimate of the dependent variable (scale (σ) in table 3) decreases with the number of observations.[8] This suggests that our estimations are more precise when more observations are involved.

[7] Goffette-Nagot *et al.* (2010) tested this assumption by estimating a hedonic regression with data collected at the municipality level. However, as they obtained negative coefficients, they were not able to confirm that hypothesis. The use of household data helps greatly to improve the results regarding this variable.

[8] Except from the Union to the Extended Residential Urban Complex samples.

Table 3: Interval regression: estimation results for different study areas.

Variable	**Union**	**ERUC**	**Urban Region**	**Agglo.**	**Brussels Capital**
Intercept	**5.4538**	**5.7156**	**4.4735**	**4.0769**	**4.3542**
APP	**-0.0746**	**-0.0768**	**-0.0776**	**-0.0732**	**-0.0512**
OTHER	0.0244	0.0137	-0.0466	**-0.0914**	-0.0415
SING[1]					
LOFT	0.0169*	**0.0292**	**0.0642**	**0.0573**	**0.0566**
LNROOMS	**0.0958**	**0.1026**	**0.1253**	**0.1182**	**0.1193**
SURFA	**-0.4354**	**-0.4541**	**-0.5148**	**-0.5401**	**-0.5470**
SURFB	**-0.3889**	**-0.4050**	**-0.4644**	**-0.4903**	**-0.4972**
SURFC	**-0.3390**	**-0.3514**	**-0.4084**	**-0.4326**	**-0.4385**
SURFD	**-0.2520**	**-0.2596**	**-0.3030**	**-0.3157**	**-0.3130**
SURFE	**-0.1644**	**-0.1679**	**-0.1869**	**-0.1923**	**-0.1768**
SURFF[1]					
RECENTBUILT	**0.1726**	**0.1729**	**0.1992**	**0.2032**	**0.1897**
RENOVATION	**0.0641**	**0.0650**	**0.0821**	**0.0901**	**0.0980**
FURNISH	**-0.1917**	**-0.1895**	**-0.1828**	**-0.1674**	**-0.1603**
HEATA	**0.1908**	**0.1858**	**0.1695**	**0.1569**	**0.1428**
HEATB	**0.1789**	**0.1719**	**0.1505**	**0.1361**	**0.1294**
HEATC	**0.1864**	**0.1806**	**0.1560**	**0.1501**	**0.1497**
HEATD[1]					
FUELA	-0.0183	-0.0228	0.0591	0.0923*	0.0867
FUELB	**-0.1788**	**-0.1883**	**-0.1177**	-0.0761	-0.0420
FUELC	**-0.1040**	**-0.1147**	-0.1004*	-0.0240	0.0448
FUELD	0.0947	0.1094*	0.1555*	0.1449	0.1485*
FUELE	0.0702*	0.0600	**0.1168**	**0.1286**	**0.1159**
FUELF	-0.0075	-0.0157	0.0416	0.0642	0.0573
FUELG	-0.0470	-0.0506	0.0029	0.0424	0.0537
FUELH[1]					
DGLAZING	**0.0295**	**0.0267**	**0.0200**	**0.0186**	**0.0211**
WALLISO	**0.0406**	**0.0400**	**0.0347**	**0.0318**	**0.0265**
BATHROOM	**0.2525**	**0.2444**	**0.2172**	**0.2095**	**0.2001**
TOILET	**0.0401**	**0.0412**	**0.0404**	**0.0342**	**0.0375**
PARKA	**-0.1514**	**-0.1519**	**-0.1750**	**-0.1728**	**-0.1419**
PARKB	**-0.0468**	**-0.0435**	**-0.0430**	**-0.0317**	0.0085

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table3 – concluded from previous page

Variable	Union	ERUC	Urban Region	Agglo.	Brussels Capital
PARKC[1]					
GARDEN	0.0066*	0.0071*	**0.0148**	**0.0134**	**0.0186**
LNPM10 (COM)	**-0.1686**	**-0.2456**	**-0.1386**	**-0.1808**	**-0.2378**
LNACCE (AC)	**-0.5462**	**-0.5502**	**-0.4900**	**-0.4767**	**-0.5467**
LNSLOPE (AC)	**0.1299**	**0.1223**	**0.1508**	**0.1285**	**0.1607**
LNPER_FOREST (AC)	**0.0110**	**0.0101**	**0.0132**	**0.0104**	**0.0040**
LNPER_ARABLE (AC)	**-0.0223**	**-0.0255**	**-0.0263**	**-0.0273**	**-0.0360**
LNREVMED (SS)	**0.4035**	**0.4115**	**0.4686**	**0.5228**	**0.5503**
scale(σ)	0.3426	0.3430	0.3393	0.3338	0.3273
Nobs	70,839	62,695	44,319	37,805	30,315
Sample size	330,147	301,160	233,582	208,371	177,721
AIC	133,083.6	118,840.1	83,989.87	69,680.75	53,935.67
BIC	133,422.9	119,174.8	84,311.74	69,996.73	54,243.49

[1] Reference case.
Bold: significant at 0.05 level.
* significant at 0.10 level.

Table 4 provides estimation results for a specification that differs from the previous one only by the insertion of the composite variable **QUALITY**. As described in 4.1, this variable is a quality index mixing information from several dwelling physical attributes: existence of toilets, bathrooms, central heating, kitchen, double glazing, surface. Most of the results are in line with the previous findings. Moreover, they indicate that rents increase with quality. However, there are some discrepancies: dwellings using heat pump as source of energy are more valuable than those using electricity.

Table 4: Interval regression with the quality composite attribute: estimation results for different study areas.

Variable	**Union**	**ERUC**	**Urban Region**	**Agglo.**	**Brussels Capital**
Intercept	**5.4005**	**5.5848**	**4.1561**	**3.7583**	**4.2500**
APP	**-0.0504**	**-0.0553**	**-0.0733**	**-0.0771**	**-0.0602**
OTHER	0.0209	0.0077	**-0.0663**	**-0.1119**	-0.0894*
SING[1]					
LOFT	**0.0599**	**0.0749**	**0.1208**	**0.1164**	**0.1228**
LNROOMS	**0.1844**	**0.1985**	**0.2389**	**0.2371**	**0.2461**
RECENTBUILT	**0.1873**	**0.1862**	**0.2164**	**0.2225**	**0.2093**
RENOVATION	**0.0861**	**0.0868**	**0.1062**	**0.1157**	**0.1250**
FURNISH	**-0.1825**	**-0.1809**	**-0.1769**	**-0.1640**	**-0.1563**
QUALITY1	**-0.5176**	**-0.5122**	**-0.4936**	**-0.4908**	**-0.4753**
QUALITY2	**-0.2787**	**-0.2812**	**-0.2848**	**-0.2909**	**-0.2834**
QUALITY3	**-0.2300**	**-0.2393**	**-0.2586**	**-0.2698**	**-0.2669**
QUALITY4	**-0.1473**	**-0.1511**	**-0.1626**	**-0.1572**	**-0.1412**
QUALITY5[1]					
FUELA	0.0354	0.0181	**0.1497**	**0.1972**	**0.1908**
FUELB	**-0.2532**	**-0.2783**	**-0.1703**	-0.1096*	-0.0691
FUELC	**-0.1695**	**-0.1951**	**-0.1570**	-0.0773	-0.0215
FUELD	0.1190*	0.1197*	**0.2355**	**0.2449**	**0.2405**
FUELE	0.0329	0.0108	**0.1214**	**0.1517**	**0.1383**
FUELF	-0.0014	-0.0232	0.0793	**0.1133**	0.1013*
FUELG	**-0.0886**	**-0.1040**	0.0020	0.0484	0.0594
FUELH[1]					
WALLISO	**0.0696**	**0.0671**	**0.0600**	**0.0572**	**0.0521**
PARKA	**-0.1965**	**-0.1973**	**-0.2214**	**-0.2186**	**-0.1884**
PARKB	**-0.0632**	**-0.0592**	**-0.0561**	**-0.0416**	0.0066
PARKC[1]					
GARDEN	0.0075*	**0.0090**	**0.0183**	**0.0187**	**0.0257**
LNPM10 (COM)	**-0.1793**	**-0.2595**	**-0.1377**	**-0.1804**	**-0.2582**
LNACCE (AC)	**-0.6163**	**-0.6213**	**-0.5810**	**-0.5880**	**-0.6981**
LNSLOPE (AC)	**0.1478**	**0.1416**	**0.1766**	**0.1580**	**0.1934**
LNPER_FOREST (AC)	**0.0122**	**0.0109**	**0.0149**	**0.0116**	0.0032
LNPER_ARABLE (AC)	**-0.0267**	**-0.0304**	**-0.0316**	**-0.0330**	**-0.0383**
LNREVMED (SS)	**0.4835**	**0.5000**	**0.5687**	**0.6296**	**0.6638**

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table 4 – concluded from previous page

Variable	**Union**	**ERUC**	**Urban Region**	**Agglo.**	**Brussels Capital**
scale(σ)	0.3583	0.3588	0.3561	0.3508	0.3444
Nobs	72,950	64,496	45,455	38,781	31,070
Sample size	330,147	301,160	233,582	208,371	177,721
AIC	141,884.2	126,557.2	89,418.52	74,353.28	57,569.71
BIC	142,160.1	126,829.4	89,680.25	74,610.25	57,820.03

[1] Reference case.

Bold: significant at 0.05 level.

[*] significant at 0.10 level.

Tables 3 and 4 show the impact of the choice of the limits of the agglomeration on econometric results through estimations of separate samples. Theoretically, this is not the most efficient way to test the econometric impacts of different delineations. The econometrics literature recommends rather to perform an estimation with interaction terms based on the largest sample. This procedure has the advantage to allow testing whether each coefficient varies significantly, in a statistical viewpoint, across different delineations. However, as it adds to the previous specification several interaction terms that are likely to be correlated with the other regressors, this procedure has the shortcoming of increasing the multicollinearity between regressors. In table 10 in Appendix B, we present results of this interval regression with interaction terms for the union sample. Retaining only coefficients significant at the 10% level, we display in table 5 the coefficients implied by this specification. Table 5 clearly indicates that most of the regressors vary significantly across different definitions of the study area. Therefore, it gives a statistical backing to the existence of a sensitivity of statistical results to the delineation of the study area.

Table 5: Implied coefficients estimates in case of interaction terms

Variable	**BCR**	**Agglo.**	**Urban Region**	**RER**	**ERER**	**RUC**	**ERUC**
Intercept	8.372	6.7758	5.4761	10.3239	4.7307	4.2473	15.4603
APP	0.014	-0.069	-0.1531	-0.069	-0.069	-0.0006	-0.1198
OTHER			-0.1335				
SING[1]							
LOFT			0.12	-0.107			
LNROOMS	0.0838	0.0838	0.1516	0.0838	0.0838	0.0838	0.0838
SURFA	-0.3114	-0.2218	-0.4028	-0.2218	-0.2218	-0.2218	-0.2218
SURFB	-0.2813	-0.1928	-0.359	-0.1928	-0.1928	-0.1928	-0.1928
SURFC	-0.2288	-0.1503	-0.3058	-0.2076	-0.1503	-0.1503	-0.1503
SURFD	-0.0887	-0.0887	-0.2254	-0.1599	-0.0887	-0.0887	-0.0887
SURFE	-0.0568	-0.0568	-0.1499	-0.0568	-0.0568	-0.0568	-0.0568
RECENTBUILT	0.1686	0.1686	0.251	0.1686	0.1686	0.1163	0.1686
RENOVATION	0.1097	0.0549	0.0549	0.0549	0.0549	0.0549	
FURNISH	-0.2383	-0.2383	-0.28	-0.1681	-0.2383	-0.1244	-0.3237
HEATA	0.1314	0.2078	0.2422	0.2078	0.2078	0.2078	0.1879
HEATB	0.189	0.2233	0.2233	0.2233	0.2233	0.2233	0.1546
HEATC	0.2109	0.2109	0.2109	0.2109	0.2109	0.2109	0.2109
FUELA							
FUELB	-0.163	-0.163	-0.163-0.163	-0.163	-0.163	-0.163	-0.163
FUELC	-0.0852	-0.0852	-0.0852	-0.0852	-0.0852	-0.0852	-0.0852
FUELD							
FUELE	0.068	0.068	0.068	0.068	0.068	0.068	0.068
FUELF							
FUELG							

Table 5 – concluded from previous page

Variable	BCR	Agglo.	Urban Region	RER	ERER	RUC	ERUC
FUELH[1]							
DGLAZING	0.0481	0.0481	0.0198	0.0481	0.0481	0.0481	0.0481
WALLISO	0.004	0.0394	0.0394	0.0394	0.0394	0.0394	0.0394
BATHROOM	0.2223	0.3253	0.3253	0.2449	0.3253	0.3253	0.3253
TOILET				0.0776			
PARKA	-0.096	-0.1446	-0.2151	-0.1446	-0.1446	-0.1446	-0.1446
PARKB	0.0287	-0.0647	-0.0939	-0.0647	-0.0647	-0.0647	-0.0647
PARKC[1]							
GARDEN	-0.0277	-0.0494	-0.0494	-0.0494	-0.0033	-0.0494	-0.0494
LNPM10 (COM)	-0.2859		0.9084				-0.7052
LNACCE (AC)				-0.8014		0.7148	-0.7865
LNSLOPE (AC)			0.0971	0.0838			
LNPER_FOREST (AC)	-0.0167	-0.0296	0.0015	0.0019	-0.0167	-0.0167	-0.0167
LNPER_ARABLE (AC)			-0.0287			0.0351	-0.066
LNREVMED (SS)	0.0569	0.0945	-0.122	0.1517	0.4886	-0.122	-0.122

To sum up, choices of the limits of the study area are not benign regarding econometric results. As table 5 just showed, for some coefficients, estimates obtained with different macrozones may be significantly different. Therefore, the choice of the study area is a very sensitive issue regarding the precision of estimates. This further stresses the need to delineate the study in a way that is consistent with the problem under investigation. A failure to do so would entail biases that may mislead statistical inferences and the policy recommendations that they may drive. For instance, the absolue value of the elasticity of the accessibility indicator rises from 0.0223 for the Union macrozone to 0.036 for Brussels capital, a 60% increase. Such a discrepancy may have a dramatic impact in a microsimulation tool as Urbansim and may lead to highly mistaken conclusions in terms of land use and transport policies. In his literature review about regional convergence, Magrini (2004) warns against the measurement problems resulting from the mismatch between the spatial pattern of the process under study and the boundaries of the observational units. In the specific context of regional convergence analysis, the inadequate choice of the observational units might hide substantial dependence of income growth. However, Magrini's claim does not specifically concern the limits of the study area, but rather those of the basic spatial units. In the next subsection, we specifically address the question of the choice of basic spatial units.

5.1.2 Impacts of the choice of the basic spatial unit.

After having evidenced that statistical results are sensitive to the choice of the delineation of the study area, we now investigate on the impact of the aggregation scale on econometrics findings. To do so, we consider two indicators of neighborhood income: the logarithm of the median and the average income by tax declaration.

Table 6 compares estimation results when the median income by tax declaration is measured successively at the statistical sector and at the municipality level. It shows that the coefficient of the logarithm of the median income is higher when median income is measured at the municipality level. The same observation holds in table 7 where the average median income is successively captured at the statistical sector, the former township and at the municipality levels. Similar results are obtained for other neighborhood and environmental variables as shown in table 11 in Appendix C.

Such results are consistent with Gehlke and Biehl (1934) findings which outline that the correlation coefficient tends to increase as the size of spatial units increases. What is the rationale of such findings? A possible explanation of the tendency of those coefficients to increase with the size of the BSU may be the following: the higher the BSU the lower the variance of a variable. As the standard deviations of variables lie in the denominator of the correlation coefficient and the simple regression coefficient this may explain their increase when the size of a BSU increases. We may conjecture that a similar effect operates on variable coefficients in the interval regression model.

Table 6: Interval regression: impact of the choice of the basic spatial unit for the variable LNREVMED, Sample=Union.

	Coefficients	
Variable	**Specification1**	**Specification2**
Intercept	**5.4538**	**1.7213**
APP	**-0.0746**	**-0.0798**
OTHER	0.0244	0.0100
SING[1]		
LOFT	0.0169*	0.0153
LNROOMS	**0.0958**	**0.0957**
SURFA	**-0.4354**	**-0.4399**
SURFB	**-0.3889**	**-0.3918**
SURFC	**-0.3390**	**-0.3396**
SURFD	**-0.2520**	**-0.2520**
SURFE	**-0.1644**	**-0.1658**
SURFF[1]		
RECENTBUILT	**0.1726**	**0.1759**
RENOVATION	**0.0641**	**0.0564**
FURNISH	**-0.1917**	**-0.1947**
HEATA	**0.1908**	**0.1942**
HEATB	**0.1789**	**0.1864**
HEATC	**0.1864**	**0.1728**
HEATD[1]		
FUELA	-0.0183	-0.0007
FUELB	**-0.1788**	**-0.1612**
FUELC	**-0.1040**	**-0.0852**
FUELD	0.0947	0.0979
FUELE	0.0702*	**0.0828**
FUELF	-0.0075	0.0041
FUELG	-0.0470	-0.0264
FUELH[1]		
DGLAZING	**0.0295**	**0.0297**
WALLISO	**0.0406**	**0.0434**
BATHROOM	**0.2525**	**0.2583**
TOILET	**0.0401**	**0.0379**
PARKA	**-0.1514**	**-0.1642**
PARKB	**-0.0468**	**-0.0489**

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table6 – concluded from previous page

	Coefficients	
Variable	**Specification1**	**Specification2**
PARKC[1]		
GARDEN	0.0066*	**0.0161**
LNPM10 (COM)	**-0.1686**	0.0425
LNACCE (AC)	**-0.5462**	**-0.5171**
LNSLOPE (AC)	**0.1299**	**0.1273**
LNPER_FOREST (AC)	**0.0110**	**0.0102**
LNPER_ARABLE (AC)	**-0.0223**	**-0.0332**
LNREVMED (SS)	**0.4035**	
LNREVMED (COM)		**0.6931**
scale(σ)	0.3426	0.3440
Nobs	70,839	72,105
Sample size	330,147	
AIC	133,083.6	136,066.4
BIC	133,422.9	136,406.3

[1] Reference case.
Bold: significant at 0.05 level.
* significant at 0.10 level.

Changing the scale of the basic spatial unit for one variable also impacts the coefficients of the other variables. In table 6, the most important effects are observed in the intercept, which is almost divided by 3 in the second specification, in the **GARDEN** coefficient, which is significant only at 10% confidence level in the first specification and more than twice as large in the second specification, and in the **LNPM10 (COM)** coefficient which is not significant in the second specification.

In table 7, we also observe substantial changes in the intercept, which is almost divided by 5 in the third specification, the **GARDEN** coefficient, which is not significant in the first specification but increases by almost 50% from the second to the third specification, and the **LNPM10 (COM)** coefficient, which is not significant in the second specification and is positive — a surprising result — in the third specification. Therefore, the intercept, the garden and the pollution variables coefficients appear as very sensitive to change in the basic spatial unit.

Table 7: Interval regression: impact of the choice of the basic spatial unit for the variable LNREVMOY, Sample=Union.

	Coefficients		
Variable	**Specification1**	**Specification2**	**Specification3**
Intercept	**4.5262**	**3.1138**	**0.9209**
APP	**-0.0704**	**-0.0807**	**-0.0781**
OTHER	**0.0513**	0.0035	0.0111
SING[1]			
LOFT	0.0188*	0.0144	0.0137
LNROOMS	**0.0936**	**0.0940**	**0.0963**
SURFA	**-0.4275**	**-0.4379**	**-0.4295**
SURFB	**-0.3804**	**-0.3903**	**-0.3813**
SURFC	**-0.3310**	**-0.3389**	**-0.3300**
SURFD	**-0.2392**	**-0.2514**	**-0.2440**
SURFE	**-0.1491**	**-0.1628**	**-0.1592**
SURFF[1]			
RECENTBUILT	**0.1753**	**0.1765**	**0.1758**
RENOVATION	**0.0668**	**0.0579**	**0.0560**
FURNISH	**-0.1884**	**-0.1910**	**-0.1922**
HEATA	**0.1788**	**0.1910**	**0.1903**
HEATB	**0.1661**	**0.1833**	**0.1823**
HEATC	**0.1777**	**0.1739**	0.1672
HEATD[1]			
FUELA	-0.0307	-0.0165	-0.0037
FUELB	**-0.1789**	**-0.1750**	**-0.1615**
FUELC	**-0.1022**	**-0.1018**	**-0.0920**
FUELD	0.0406	0.0815	0.0924
FUELE	0.0505	0.0692*	**0.0800**
FUELF	-0.0204	-0.0080	0.0038
FUELG	-0.0479	-0.0423	-0.0296
FUELH[1]			
DGLAZING	**0.0272**	**0.0284**	**0.0287**
WALLISO	**0.0378**	**0.0432**	**0.0438**
BATHROOM	**0.2452**	**0.2559**	**0.2573**
TOILET	**0.0393**	**0.0377**	**0.0387**
PARKA	**-0.1374**	**-0.1565**	**-0.1610**
PARKB	**-0.0373**	**-0.0449**	**-0.0468**

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table7 – concluded from previous page

	Coefficients		
Variable	**Specification1**	**Specification2**	**Specification3**
PARKC[1]			
GARDEN	0.0002	**0.0086**	**0.0128**
LNPM10 (COM)	**-0.0922**	0.0004	**0.1789**
LNACCE (AC)	**-0.4469**	**-0.4311**	**-0.4066**
LNSLOPE (AC)	**0.1111**	**0.1029**	**0.0944**
LNPER_FOREST (AC)	**0.0085**	**0.0042**	**0.0056**
LNPER_ARABLE (AC)	**-0.0241**	**-0.0323**	**-0.0359**
LNREVMOY (SS)	**0.4067**		
LNREVMOY (AC)		**0.5071**	
LNREVMOY (COM)			**0.6477**
scale(σ)	0.3341	0.3418	0.3417
Nobs	63,426	70,828	72,105
Sample size		330,147	
AIC	116,497.5	132,953.3	135,340.9
BIC	116,832.7	133,292.5	135,680.8

[1] Reference case.
Bold: significant at 0.05 level.
[*] significant at 0.10 level.

To the best of our knowledge, up to now, no statistical test is available to check whether the difference between the same coefficients across the specifications is statistically significant. The criminology literature mentions tests of equality of regression coefficient across independent samples (Brame *et al.*, 1998). Such tests are not suitable in our case since our model is not a simple linear regression model and its coefficients are not characterized by standard normal distributions.

The results just described outline the necessity to use the finest spatial scale for the definition of environment quality and neighborhood attributes. Using larger basic spatial units for the definition of those variables, would result in inflated estimates. Once more, such biases may imply mislead conclusions in terms of the policy recommendations drawn upon econometric results. However, due to constraints in data availability, sometimes they are unavoidable as the information of some variables may only be obtained at specific spatial scales. This is precisely the case of the pollution indicator which, because of raster resolution, can only be computed at the municipality level. In such cases of spatial scale constraints, one should to be aware of the potential biases.

5.2 SAR Interval regression model.

Let us now investigate the substantive spatial dependence issue through our Spatial Autoregressive Interval Regression (SARIR) model. Unfortunately, we have not been able to run the SARIR algorithm on the full sample. Indeed, it is very demanding in terms of computational resources. So we ran our algorithm on 2 subsamples of respectively 2,969 observations and 62 statistical sectors (Sample 1) and 2,565 observations and 81 statistical sectors (Sample 2).

Dwellings of Sample 1 are located in the municipalities of Anderlecht, Berchem-Sainte-Agathe, and Molenbeek-Saint-Jean. The geographical locations of Sample 2 dwellings lie within the municipalities of Auderghem, Woluwe-Saint-Lambert, and Woluwe-Saint-Pierre.[9] Figures 2 and 3 depicts the average income in Brussels Capital Region (BCR) as well as the location of Sample 1 and Sample 2 dwellings in the BCR.

Unfortunately, since all the observations of the two samples refer to only 3 different municipalities, there is not enough variation in the environment and neighborhood attributes to assess reliably any potential impact of changes of the basic spatial unit. Table 8 displays the results of the basic interval regression model as well as those of the SARIR algorithm. The results of the two models differ substantially only for the spatial dependence parameter and the intercept. In the first sample estimations, the intercept is almost the half of the one obtained in the benchmark model. In the second sample estimation it is almost the eighth. The substantial reduction of the intercept in SARIR models is an indication that the omitted variable bias is significantly reduced in the spatial model.

While traditional hedonic estimation does not address omitted variable bias, the spatial autoregressive model mitigates that issue. Indeed, the spatial lag term $W\tilde{y}$ picks up unobserved influences that affect dwelling rent (Brasington and Hite (2005)). It relies on a linear combination of dwelling rents nearby in space. Unmeasured influences help determine the rent of neighboring dwellings and the rent of a given dwelling is related to the rents of neighboring dwellings as well. Hence, any dwelling rent is affected by the unmeasured influences of neighboring observations. Therefore, the $W\tilde{y}$ term incorporates the influence of omitted variables on the rent of a dwelling.

[9]Considering an axis dividing Brussels Capital Region from the South-East to the North-East, then the first sample lies in the part of Brussels above that axis and the second sample in the other side. As shown in Figure 2, most of the municipalities of Brussels above that axis, like Moleenbeek-Saint-Jean, Anderlecht, Saint-Josse have a lower average income (Ganshoren, Berchem-Sainte-Agathe, and Jette are exceptions characterized by higher incomes). Municipalities below that axis, especially Uccle, Watermael-Boitsfort, Auderghem and Woluwe-Saint-Pierre have a higher average income.

Figure 2: Average income in municipalities of Brussels Capital Region

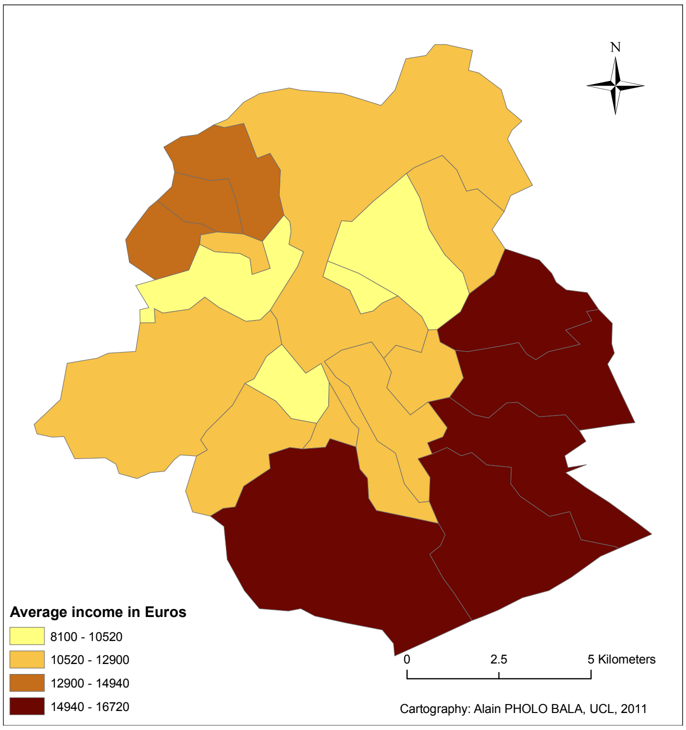
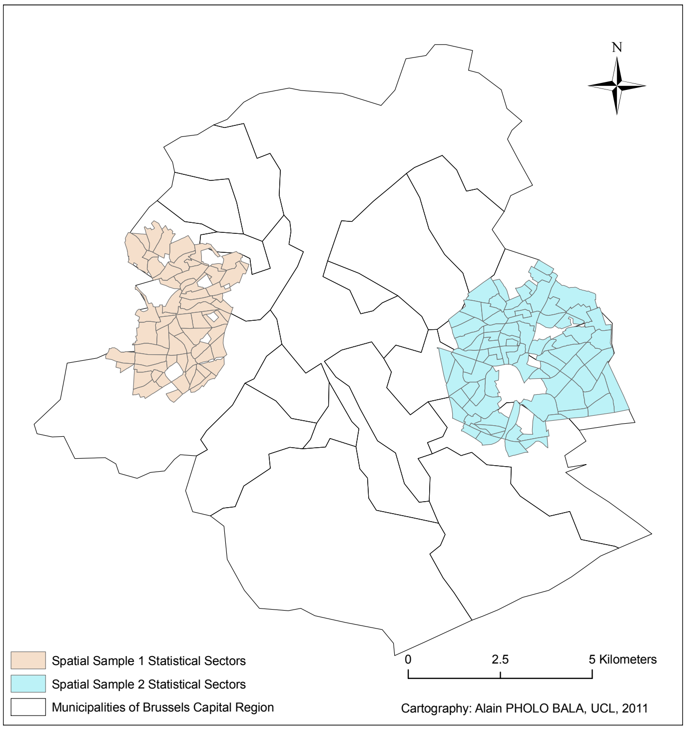


Figure 3: Location of Sample 1 and Sample 2 in Brussels Capital Region



Omitting the spatial lag term, as in the non spatial IRM, would definitely entail biased estimates. Such biases are more perceptible in the **LNREVMED (SS)** coefficient which is 18% higher in the non spatial IRM estimation based on Sample 1 and 7% higher in the non spatial IRM estimation based on Sample 2. By acknowledging the impact of dwelling rents nearby in space, the spatial model imply a lower impact of the median income. Another noteworthy observation may be made on **LNREVMED (SS)** coefficient: it is the only coefficient of an environment or a neighborhood attribute that is significant. A possible explanation is the following: since the other environment and neighborhood variables are measured at a more aggregated level ("ancienne commune" or municipality), they have not sufficient variation to allow enough precision in the measure of their estimates.

Another interesting result lies in the discrepancy between Sample 1 and Sample 2 SARIR model results. For instance, the spatial dependence parameter is higher for Sample 2 than for Sample 1, suggesting that dwellings rents of the neighborhood have a stronger impact on the dependant variable in Sample 2. A similar observation may be made about **LNREVMED (SS)** coefficient which is also higher in Sample 2 estimations. Both results indicate that the neighborhood has a stronger impact in the determination of Sample 2 dwelling rents.

The difference between Sample 1 and Sample 2 results yields some evidence of spatial heterogeneity in the estimation. Spatial heterogenity is related to the lack of stability over space of the relationship under study. It implies that functional forms and parameters vary with location (Anselin (1988)). While the analysis of this issue is beyond the scope of the paper, it may trigger interesting perspectives in terms of spatial issues investigation.

Table 8: SAR Interval regression: estimation results for different samples.

Variable	Sample1		Sample2	
	Inter.reg.	SAR	Inter.reg.	SAR
ρ		**0.1422**		**0.1735**
Intercept	23.7460	11.0174	-32.4745	-4.8084
SING	**0.0720**	**0.0703**	**0.1234**	**0.1191**
OTHER	**0.5674**	**0.5699**	**-0.5259**	**-0.5685**
APP[1]				
LOFT	**0.1837**	**0.1841**	**0.1171**	**0.1151**
LNROOMS	**0.2285**	**0.2291**	**0.2941**	**0.2937**
RECENTBUILT	**0.2435**	**0.2410**	**0.2438**	**0.2459**
RENOVATION	**0.0700**	**0.0701**	**0.0930**	**0.0976**
FURNISH	**-0.1300**	**-0.1324**	**-0.1185**	**-0.1200**
QUALITY2	**0.1463**	**0.1465**	**0.2507**	**0.2490**
QUALITY3	**0.2200**	**0.2201**	**0.2095**	**0.2058**
QUALITY4	**0.3062**	**0.3051**	**0.3281**	**0.3250**
QUALITY5	**0.3911**	**0.3908**	**0.5101**	**0.5112**
QUALITY1[1]				
FUELA	0.0149	0.0138	0.0107	0.0112
FUELB	**-0.2625**	**-0.2595**	**-0.8883**	**-0.9325**
FUELC	-1.5625	-1.6212	-0.2340	-0.2904
FUELH	-0.0390	-0.0396*	-0.0747*	**-0.0732**
FUELE[1]				
WALLISO	**0.0337**	**0.0334**	**0.0727**	**0.0709**
PARKB	**0.1256**	**0.1244**	**0.2611**	**0.2582**
PARKC	**0.1215**	**0.1204**	**0.2690**	**0.2673**
PARKA[1]				

[1] Reference case.

Bold: not significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table 8 – concluded from previous page

Variable	**Sample1**		**Sample2**	
	Inter.reg.	**SAR**	**Inter.reg.**	**SAR**
GARDEN	-0.0054	-0.0058	0.0107	0.0100
LNPM10 (COM)	-1.5882	-0.5385	5.8657	1.8797
LNACCE (AC)	-2.7523	-1.2245	2.2314	-0.3080
LNSLOPE (AC)	0.0111	0.0036	-0.0288	-0.0223
LNPER_FOREST (AC)			0.4000	0.1315
LNPER_ARABLE (AC)	0.0647	0.0318		
LNREVMED (SS)	**0.2800**	**0.2369**	**0.4660**	**0.4357**
scale(σ)	0.2801	0.0807	0.3642	0.1362
Nobs	2,969		2,565	

[1] Reference case.

Bold: not significant at 0.05 level.

[*] significant at 0.10 level.

6 Conclusion

As the central microsimulation tool used in the Sustaincity project, UrbanSim requires a massive amount of geocoded data, collected from several sources and often available at different spatial scales. Hence, choices have to be made about the relevant underlying basic spatial units (BSU), as well as the definition(s) of the studied area. Those choices are generally suspected to influence or even bias econometric results. Moreover, spatial autocorrelation is also likely to have significant impacts on statistical findings.

The main task of the UCL was to address those issues by carrying out sensitivity analyses and by developing adequate statistical tools. Therefore, on the basis of the hedonic regression model, the three aforementioned problems have been investigated in this Sustaincity deliverable.

First, the delineation of the metropolitan area highly impacts the statistical estimations. This was tested in the Basic Interval Regression model through interaction terms. We show that most of the coefficients highly and significantly vary with the definition of the study area. Hence, defining a city by functional or morphological criteria as each urban specialist or planner would do will lead to different results to that defined by a transportation regional planner (RER zone in Brussels).

A second spatial aspect addressed by the UCL team is the choice of the basic spatial units. The sensitivity of the coefficients to scale effect is empirically demonstrated on the example of Brussels. Our results are consistent with Gehlke and Biehl (1934) and the related literature. A possible explanation of such findings is that the larger the size of the BSUs, the lower the variance of the considered variable. As the standard deviations of variables lie in the denominator of the correlation coefficient and the simple regression coefficient, this may explain their increase when the size of a BSU increases. While there are no analytical expressions for variables coefficients in the interval regression model, we may conjecture that a similar effect operates on them.

Therefore, in order to minimize the biases, the delineation of the study area must be chosen in a way that is consistent with the phenomenon under investigation. Moreover, the finer the aggregation scale, the more precise the coefficients estimates. Indeed, bad choices in terms of the aggregation scale may lead to misspecification biases. In the ideal situation where all the statistical information is available at the individual level, biases inherent to the ecological fallacy would not exist.

A last issue is the impact of substantive spatial dependence. We accounted for it by considering one of the main components of the spatial econometrics toolbox: the Spatial AutoRegressive

Model (SAR). Since the information about our dependent variable is collected through a categorical variable, we have to resort on techniques designed to estimate spatially dependent discrete choice models. Therefore, we designed and estimated a "Spatial Autoregressive Interval Regression" model. As we obtained a statistically significant spatial dependence parameter from those estimations, our econometric results provide evidence of spatial dependence. The estimation of this Spatial Model is likely to mitigate the omitted variable bias which generally undermines traditional hedonic estimation.

Moreover, the difference in the results of the two samples used brings some evidence of spatial heterogeneity in the estimation. Indeed, spatial heterogenity is likely to occur in the estimation of econometric models on cross-sectional data set with dissimilar spatial units such as wealthy households in the South East of Brussels Capital Region and poorer households in the North West (Anselin (1988)). This gives ground for interesting future research.

As a concluding comment, let us simply insist on the fact that basic linear econometric estimations that do not consider spatial effects (delineation of the studied area, scale and spatial autocorrelation) may lead to erroneous decisions and conclusions. Therefore, one has to be careful in further interpretations.

Appendix A: Descriptive statistics of continuous variables.

Table 9: Descriptive statistics (Sample:“Union”).

Variable	Description	N	Min	Max	Mean	S.D.
dependent variable						
LLNRENT	ln of lower bound of rent interval	276,765	5.521	6.899	5.763	0.402
ULNRENT	ln of upper bound of rent interval	309,615	5.521	6.899	6.211	0.338
Regressors						
LNROOMS	ln of number of rooms	305,161	0	4.595	1.215	0.532
LNPM10 (COM)	ln of average concentration of PM_{10} by municipality	330,147	3.221	3.601	3.445	0.088
LNPER_FOREST (COM)	ln(Percent_Forest+1) by municipality	330,147	0	4.034	0.989	1.184
LNPER_FOREST (AC)	ln(Percent_Forest+1) by “Ancienne Commune"	330,118	0	4.035	0.847	1.204
LNPER_ARABLE (COM)	ln(Percent_Arable+1) by municipality	330,147	0	4.472	2.123	1.775
LNPER_ARABLE (AC)	ln(Percent_Arable+1) by “Ancienne Commune"	330,118	0	4.582	1.792	1.803
LNSLOPE (SS)	ln(SLOPE) (by statistical sector)	330,107	-0.680	2.945	1.012	0.523
LNSLOPE (AC)	ln(SLOPE) (by “Ancienne Commune")	330,118	-0.142	2.446	1.136	0.382
LNSLOPE (COM)	ln(SLOPE) (by municipality)	330,147	0.049	1.876	1.132	0.353
LNREVMED (SS)	ln(REVMED) (by statistical sector)	326,038	9.126	10.595	9.817	0.158
LNREVMED (COM)	ln(REVMED) (by municipality)	330,147	9.570	10.104	9.830	0.112
LNREVMOY (SS)	ln(REVMOY) (by statistical sector)	300,366	9.458	10.947	10.053	0.208
LNREVMOY (AC)	ln(REVMOY) (by “Ancienne Commune")	326,335	9.646	10.665	10.085	0.163
LNREVMOY (COM)	ln(REVMOY) (by municipality)	330,147	9.570	10.104	9.830	0.112
LNACCE (AC)	ln(ACCE+1) (by “Ancienne Commune")	330,096	5.594	6.164	5.750	0.113

Appendix B: Interval regression with interaction terms.

Table 10: Interval regression with interaction terms.

Variable	Estimate	Variable	Estimate
Intercept	**8.3720**	**DAGGLOP*LNSLOPE (AC)**	0.0243
DBRXCAP	-0.6966	**DAGGLOP*LNPER_FOREST (AC)**	**-0.0129**
DAGGLOP	**-1.5962**	**DAGGLOP*LNREVMED (SS)**	**0.2165**
DREGURB	**-2.8959**	**DREGURB*APP**	**-0.0841**
DCRU	**-4.1247**	**DREGURB*OTHER**	**-0.1335**
DCRUE	**7.0883**	**DREGURB*LOFT**	**0.1200**
DZRER	1.9519***	**DREGURB*LNROOMS**	**0.0678**
DZRERE	**-3.6413**	**DREGURB*SURFA**	**-0.1810**
APP	**-0.0690**	**DREGURB*SURFB**	**-0.1662**
OTHER	0.0666*	**DREGURB*SURFC**	**-0.1555**
SING[1]		**DREGURB*SURFD**	**-0.1367**
LOFT	0.0914 **	**DREGURB*SURFE**	**-0.0931**
LNROOMS	**0.0838**	**DREGURB*RECENTBUILT**	**0.0824**
SURFA	**-0.2218**	**DREGURB*FURNISH**	**-0.0417**
SURFB	**-0.1928**	**DREGURB*HEATA**	**0.0344**
SURFC	**-0.1503**	**DREGURB*HEATB**	0.0248**
SURFD	**-0.0887**	**DREGURB*HEATC**	-0.0416
SURFE	-0.0568***	**DREGURB*DGLAZING**	**-0.0283**
SURFF[1]		**DREGURB*PARKA**	**-0.0705**
RECENTBUILT	**0.1686**	**DREGURB*PARKB**	**-0.0292**
RENOVATION	**0.0549**	**DREGURB*LNPM10 (COM)**	**0.9084**
FURNISH	**-0.2383**	**DREGURB*LNSLOPE (AC)**	**0.0971**
HEATA	**0.2078**	**DREGURB*LNPER_FOREST (AC)**	**0.0182**
HEATB	**0.2233**	**DREGURB*LNPER_ARABLE (AC)**	**-0.0287**
HEATC	**0.2109**	**DCRU*APP**	**0.0684**
HEATD[1]		**DCRU*OTHER**	-0.0173
FUELA	-0.0096	**DCRU*LOFT**	0.0412
FUELB	**-0.1630**	**DCRU*RECENTBUILT**	**-0.0523**
FUELC	**-0.0852**	**DCRU*RENOVATION**	0.0191
FUELD	0.0975**	**DCRU*FURNISH**	**0.1139**

[1] Reference case.

Bold: significant at 0.05 level.

*** significant at 0.10 level.

** significant at 0.15 level.

* significant at 0.20 level.

Continued on next page

Table 10 – continued from previous page

Variable	Estimate	Variable	Estimate
FUELE	0.0680***	**DCRU*LNACCE (AC)**	**0.7148**
FUELF	-0.0124	**DCRU*LNSLOPE (AC)**	0.0219*
FUELG	-0.0367	**DCRU*LNPER_ARABLE (AC)**	**0.0351**
FUELH[1]		**DCRUE*APP**	**-0.0508**
DGLAZING	**0.0481**	**DCRUE*OTHER**	0.0511
WALLISO	**0.0394**	**DCRUE*LOFT**	-0.0835**
BATHROOM	**0.3253**	**DCRUE*RENOVATION**	-0.0318 *
TOILET	-0.0386	**DCRUE*FURNISH**	**-0.0854**
PARKA	**-0.1446**	**DCRUE*HEATA**	-0.0199***
PARKB	-0.0647***	**DCRUE*HEATB**	**-0.0687**
PARKC[1]		**DCRUE*HEATC**	-0.0417
GARDEN	**-0.0494**	**DCRUE*LNPM10 (COM)**	**-0.7052**
LNPM10 (COM)	0.0355	**DCRUE*LNACCE (AC)**	**-0.7865**
LNACCE (AC)	-0.3048**	**DCRUE*LNPER_ARABLE (AC)**	**-0.0660**
LNSLOPE	-0.0507**	**DZRER*LOFT**	**-0.1070**
LNPER_FOREST (AC)	**-0.0167**	**DZRER*LNROOMS**	-0.0284**
LNPER_ARABLE (AC)	0.0070	**DZRER*SURFA**	-0.0557*
LNREVMED (SS)	**-0.1220**	**DZRER*SURFB**	-0.0521**
DBRXCAP*APP	**0.0830**	**DZRER*SURFC**	-0.0573***
DBRXCAP*OTHER	-0.0073	**DZRER*SURFD**	**-0.0712**
DBRXCAP*SURFA	**-0.0896**	**DZRER*SURFE**	-0.0486**
DBRXCAP*SURFB	**-0.0885**	**DZRER*RECENTBUILT**	-0.0081
DBRXCAP*SURFC	**-0.0785**	**DZRER*FURNISH**	**0.0702**
DBRXCAP*SURFD	-0.0200	**DZRER*WALLISO**	0.0224 **
DBRXCAP*SURFE	0.0181	**DZRER*BATHROOM**	**-0.0804**
DBRXCAP*SURFF		**DZRER*TOILET**	**0.0776**
DBRXCAP*RENOVATION	**0.0548**	**DZRER*PARKA**	-0.0297
DBRXCAP*HEATA	**-0.0764**	**DZRER*PARKB**	-0.0064
DBRXCAP*HEATB	**-0.0343**	**DZRER*LNACCE (AC)**	**-0.8014**
DBRXCAP*HEATC	0.0365	**DZRER*LNSLOPE (AC)**	**0.0838**
DBRXCAP*WALLISO	**-0.0354**	**DZRER*LNPER_FOREST (AC)**	**0.0186**
DBRXCAP*BATHROOM	**-0.1030**	**DZRER*LNPER_ARABLE (AC)**	0.0167
DBRXCAP*PARKA	**0.0486**	**DZRER*LNREVMED (SS)**	**0.2737**
DBRXCAP*PARKB	**0.0934**	**DZRERE*BATHROOM**	0.0583
DBRXCAP*GARDEN	**0.0217**	**DZRERE*PARKA**	0.0542

[1] Reference case.

Bold: significant at 0.05 level.

*** significant at 0.10 level.

** significant at 0.15 level.

* significant at 0.20 level.

Continued on next page

Table 10 – concluded from previous page

Variable	Estimate	Variable	Estimate
DBRXCAP*LNPM10 (COM)	**-0.2859**	**DZRERE*PARKB**	0.0156
DBRXCAP*LNSLOPE (AC)	-0.0216	**DZRERE*GARDEN**	0.0461***
DBRXCAP*LNREVMED (SS)	**0.1789**	**DZRERE*LNACACCE**	**0.6106**
DAGGLOP*LNPM10 (COM)	-0.1618		
scale(σ)	0.3349		
Nobs	70,839		
Sample size	330,147		
AIC	130,664.0		
BIC	131,874.2		

[1] Reference case.
Bold: significant at 0.05 level.
*** significant at 0.10 level.
** significant at 0.15 level.
* significant at 0.20 level.

Appendix C: Effects of the choice of the basic spatial units on coefficients of IRM.

Table 11: Interval regression: impact of the choice of the basic spatial unit for the LNSLOPE, LNPER_FOREST and LNPER_ARABLE variables, Sample=Union.

Variable	BSU	(1)	BSU	(2)	BSU	(3)
LNSLOPE	SS	**0.0673**	AC	**0.1299**	COM	**0.1559**
LNPER_FOREST	AC	**0.0110**	COM	**0.0145**		
LNPER_ARABLE	AC	**-0.0223**	COM	**-0.0222**		

Each row represent, for a given variable, the effect on its coefficient of changes of the basic spatial unit at the level of which it is measured. The spatial scale of the other variables is kept constant.

7 References

Amrhein, C. (1995) Searching for the elusive aggregation effect: evidence from statistical simulations, *Environment and Planning A*, **27**, 105–119.

Amrhein, C. and R. Flowerdew (1992) The effect of data aggregation on a poisson regression model of canadian migration, *Environment and Planning A*, **24**, 1381–1391.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L. and A. Bera (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics, in A. Ullah and D. Giles (eds.) *Handbook of Applied Economic Statistics*, 237–289, Marcel Dekker, New York.

Bolduc, D., B. Fortin and S. Gordon (1997) Multinomial probit estimation of spatially interdependent choices: an empirical comparison of two new techniques, *International Regional Science Review*, **20**, 77–101.

Boyle, M. A. and K. A. Kiel (2001) A survey of house price hedonic studies of the impact of environmental externalities, *Journal of Real Estate Literature*, **9** (2) 117–144.

Brasington, D. M. and D. Hite (2005) Demand for environmental quality: a spatial hedonic analysis, *Regional Science and Urban Economics*, **35**, 57–82.

Briant, A., P. Combes and M. Lafourcade (2010) Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?, *Journal of Urban Economics*, **67**, 287–302.

Butler, R. (1982) The specification of hedonic indexes for urban housing, *Land Economics*, **58**, 96–108.

Case, A. (1992) Neighborhood influence and technological change, *Regional Science and Urban Economics*, **22**, 491–508.

Cassel, E. and R. Mendelsohn (1985) The choice of functional forms for hedonic price equations : comment, *Journal of Urban Economics*, **18**, 135–142.

Cheshire, P. (2010) Why brussels needs a city-region for the city.

Donnay, J.-P. and M. Lambinon (1997) Détermination des limites d'agglomération par télédetection: discussion méthodologique et application au cas de huy (belgique), paper presented at *Actes des journées scientifiques du réseau de télédétection de l'UREF : Télédétetection* *des espaces urbains et périurbains*, 239–246, Liège, Belgium, 1995.

Flemming, M. (2004) Techniques for estimating spatially dependent discrete choice models, in L. Anselin, R. Florax and S. Rey (eds.) *Advances in Spatial Econometrics. Methodology, Tools and Applications*, 145–166, Springer–Verlag, Berlin Heidelberg.

Fotheringham, A. and D. Wong (1991) The modifiable areal unit problem in multivariate statistical analysis, *Environment and Planning A*, **23**, 1025–1044.

Gawande, K. and H. Jenkins-Smith (2001) Nuclear waste transport and residential property values: Estimating the effects of perceived risks, *Journal of Environmental Economics and Management*, **42**, 207–233.

Gehlke, C. and K. Biehl (1934) Certain effects on grouping upon the size of the correlation coefficient in census tract material, *Journal of the American Statistical Association*, **185**, 169–170.

Geoghegan, J., L. Wainger and N. Bockstael (1997) Spatial landscape indices in a hedonic framework: an ecological economics analysis using gis, *Ecological Economics*, **23**, 251–264.

Geurs, K. and J. Ritsema van Eck (2001) Accessibility measures: review and applications, *Technical Report*, Urban Research Centre, Utrecht University. Evaluation of accessibility impacts of land-use transport scenarios, and related social and economic impacts.

Goffette-Nagot, F., I. Reginster and I. Thomas (2010) Spatial analysis of residential land prices in belgium: Accessibility, linguistic border, and environmental amenities, *Regional Studies*. First published on: 17 August 2010 (iFirst).

Goodman, A. (1978) Hedonic prices, price indices and housing markets, *Journal of Urban Economics*, **5**, 471–484.

Halvorsen, R. and H. Pollakowski (1981) Choice of functional form for hedonic price equations, *Journal of Urban Economics*, **10**, 37–49.

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

Koop, G. (2003) *Bayesian econometrics*, Wiley–Interscience, Chichester.

LeSage, J. and R. Pace (2009) *Introduction to spatial econometrics*, Chapman & Hall/CRC, Boca Raton.

Löchl, M. and K. Axhausen (2009) Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Working Paper*, **5XX**, Institut für Verkehrsplanung und Transportsysteme (IVT),ETH Zürich, Zürich. Arbeitsberichte Verkehrs–und Raumplanung.

Magrini, S. (2004) Regional (di)convergence, in J. V. Henderson and J. F. Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4, 2741–2796, Elsevier, Amsterdam.

Picard, N., C. Antoniou and A. de Palma (2010) Econometric models, *SustainCity Deliverable*, **2.4**, THEMA, Université de Cergy–Pontoise.

Pinkse, J. and M. Slade (1998) Contracting in space, an application of spatial statistics to discrete choice models, *Journal of Econometrics*, **85**, 125–154.

Rosen, S. (1974) Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy*, **82**, 34–35.

Smith, V. K. and J. Huang (1995) Can markets value air quality? a meta–analysis of hedonic property value models, *Journal of Political Economy*, **103**, 209–227.

Tannier, C., I. Thomas, G. Vuidel and P. Frankhauser (2010) A fractal approach to identifying urban boundaries, *Geographical Analysis*. In Press.

Van Hecke, E., J.-M. Halleux, J.-M. Decroly and B. Mérenne-Schoumaker (2009) Noyaux d'habitat et régions urbaines dans une belgique urbanisées, enquête socio–economique 2001, *Monographies*, Direction générale Statistique et Information économique.

Vandenbulcke, G., T. Steenberghen and I. Thomas (2009) Mapping accessibility in belgium: a tool for land–use and transport planning?, *Journal of Transport Geography*, **17**, 39–53.

Vandermotten, C., F. Vermoesen, I. Thomas, W. De Lannoy and S. De Corte (1999) Villes d'europe. cartographie comparative, *Bulletin du Crédit Communal*, (207–208).

Vanneste, D., I. Thomas and L. Goossens (2007) Le logement en belgique, *Monographie*, SPF Economie, P.M.E., Classes moyennes et Energie. Politique Scientifique Générale. Enquête Socio–économique 2001 Monographies 2. In coll. with P. De Decker, J. Laureys, I. Laureyssen, X. Querriau, L. Vanderstraeten, and W. Wevers.

Wooldridge, J. (2002) *Econometrics analysis of cross section and panel data*, The MIT Press, Cambridge and London.

Yusuf, A. (2004) Does air pollution affect property value? a hedonic price analysis in jakarta.