2013/36

Spatial issues on a hedonic estimation of rents in Brussels

Alain Pholo Bala, Dominique Peeters
and Isabelle Thomas

DISCUSSION PAPER

Center for Operations Research
and Econometrics

Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve
Belgium
http://www.uclouvain.be/core

CORE DISCUSSION PAPER
2013/36

Spatial issues on a hedonic estimation of rents in Brussels

Alain PHOLO BALA [1], Dominique PEETERS[2] and Isabelle THOMAS [3]

July 2013

Abstract

Using Belgian microdata, we assess the impact, on a hedonic regression, of the distortions arising from the choice of either a specific zoning system or the delineation of the study area. We also evaluate the biases that arise when spatial effects are not accounted for. Given that the dependent variable is interval-coded, controlling for spatial dependence in this context is challenging. We address this problem with two alternative strategies. Firstly, we use the Gibbs Sampling algorithm to estimate spatial econometric models which extends the interval regression model. A major drawback of this approach is that the implied estimation is proned to the endogeneity biases inherent to our hedonic regression model. To circumvent the endogeneity issues triggered by the first estimation strategy, we also use a two-stage estimation procedure with locational fixed effects. In all specifications, results are sensitive to the Modifiable Areal Unit Problem (MAUP) and to the choice of the delineation of the study area. Moreover, they confirm the existence of substantive spatial dependence. Conversely to the previous results with a negative elasticity for the percentage of the area covered by agriculture and a positive elasticity for the potential accessibility to jobs, the second approach implies opposite effects for those two variables. This indicates that dwellings close to agricultural areas and with a lower accessibility to the main employment centers are highly demanded and that endogeneity biases are not negligible.

Keywords: MAUP, interval regression, spatial dependence, spatial heterogeneity, Brussels.

JEL Classification: C21, C24, C25, C34, Q53, R21

[1] Department of Economics and Econometrics, University of Johannesburg, South Africa. E-mail: apholo@uj.ac.za
[2] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium.
E-mail: Dominique.peeters@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.
[3] Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium.
E-mail: isabelle.thomas@uclouvain.be. This author is also member of ECORE, the association between CORE and ECARES.

1 Introduction

The increasing concerns about sustainable development and the growth of urban areas have facilitated a renewed enthusiasm for the use of quantitative models in the field of transportation and spatial planning.

Some spatial issues may arise from the implementation of those quantitative models. One of them is that their implementation requires a massive amount of geographic data collected from various sources, and often at different spatial scales. Another issue is that the definition of agglomerations or, more broadly, the delineation of the study area may differ in the different case studies. All those problems are likely to influence and bias spatial econometric analyses. Moreover, spatial autocorrelation is also likely to have significant impacts on statistical findings.

In this paper, we check the magnitude of those "spatial" biases and we propose some suggestions to control or at least limit them. To do so we will base our econometric investigation on the first–stage hedonic regression model, which is well represented in the OPUS/UrbanSim platform as the Real Estate Price Model.

In a conceptual point of view, the problem of spatial autocorrelation and the issues of the choice of spatial scales and of the study area boils down to spatial dependence and spatial heterogeneity problems. 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 to be the existence of a functional relationship between what happens at one point in space and what happens elsewhere" (Anselin, 1988, p.11).

Two broad causes may lead to spatial dependence: the nuisance and the substantive spatial dependence (Magrini, 2004). The nuisance spatial dependence refers to the by-product of measurement errors for observations in contiguous spatial units. In several cases data are collected only at aggregate scale. As 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 problems of spatial aggregation or by arbitrary delineation of spatial units of observations. 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 (MAUP).

Several contributions have assessed the impact of the MAUP on multivariate statistics (Gehlke and Biehl, 1934; Fotheringham and Wong, 1991; Amrhein, 1995; Briant *et al.*, 2010). Gehlke and Biehl (1934) outline the tendency for the correlation coefficient to increase as the size of spatial units increases. In a recent contribution, Briant *et al.* (2010) analyze the impact of size distortions on the behavior of simple regression coefficients. The context of our study is somewhat different since, as the dependent variable and several covariates are individual dwelling attributes, aggregation biases apply only to a subset of regressors.

Nuisance spatial dependence may also arise because of the arbitrary delineation of basic spatial units (BSU). In a literature review on regional convergence, Magrini (2004) makes an interesting survey of the question. He 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. On the other hand, processes of decentralization or recentralization of residences relative to workplaces is likely to affect per capita income growth rates for administratively defined regions.

A related but less investigated issue is the one arising from the choice of the delineation of the study area. This issue points more to spatial heterogeneity, i.e. the lack of uniformity of the effects of space. Any structural instability of a given relationship across space would entail different econometric results for distinct study areas. More intuitively, different limits of agglomeration entail distinct geographic structures; and therefore unequal features in terms of degree of urbanization and accessibility. Our contribution focuses on Brussels. For this specific city several delineations may be considered: administrative delineations, morphological delineations (Donnay and Lambinon, 1997; Tannier *et al.*, 2011; 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 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 this paper, we analyze nuisance spatial dependance and spatial heterogeneity by investigating the impacts of choices of the aggregation scale and of the delineation of the study area.

The substantive spatial dependence is more fundamental and is 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). 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 also assess substantive spatial dependence by considering three components of the spatial econometrics toolbox: the Spatial AutoRegressive Model (SAR), the Spatial Durbin

Model (SDM) and the General Spatial Model (SAC). Several contributions investigate the spatial dependence issue in cross–sectional hedonic price analyses through the estimation of Spatial Models (Gawande and Jenkins–Smith, 2001; Kim *et al.*, 2003; Brasington and Hite, 2005; Löchl and Axhausen, 2010).

In most of these contributions, the dependent variable (house price or dwelling rent) is continuous. In this paper, we have to face 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 to techniques designed to estimate spatially dependent discrete choice models.

There are two ways to handle this issue. The first approach consists on using a Gibbs Sampling algorithm to design "Spatial Interval Regression" models. The second approach implies the use of a two-step procedure where we perform in the first step an interval regression on structural characteristics and fixed/locational effects. Then, in the second step we retrieve fixed/locational effects to obtain averages of log of rents within the basic spatial units and we regress them on a set of observed location characteristics. This last approach has the advantage of avoiding endogeneity bias caused by locational characteristics. Indeed, one may suspect reverse causality between dwelling rents and location characteristics like average income or accessibility.

This paper is organized as follows. The next section is devoted to a detailed presentation of the estimation strategy. The third section describes the study area and section 4 presents the data used for estimation. Section 5 presents the results of estimations and section 6 concludes the paper.

2 Estimation strategy

We estimate the hedonic model by means of interval regression and we analyze spatial effects. In this section we present the methodological aspects of our two estimation approaches.

2.1 Benchmark model: interval regression

In some databases the information on rent prices is collected through a categorical variable (cfr. 4.1). Therefore, they do not give the actual value of the rent y_i^*; they just provide 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 $j \in \{1, ..., J\}$ 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 an "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]

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). This computational procedure presents some important differences with the ordered probit model: in the latter the vector α is an ordered set of unknown cut points. Therefore, there is an identification issue in the ordered probit model and $\sigma^2 = Var(y^*|X)$ (with X the matrix of regressors) is normalized to one so that the model can estimate β, the vector of regressors, and α. In the interval regression model α is rather a set of known interval boundaries, thus, β and σ^2 may be jointly estimated.

As in Geoghegan *et al.* (1997), we opt for double-log estimation. This functional form has the clear advantage of simplifying the interpretation of the estimated coefficients of continuous variables. 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.

2.2 Spatial Interval Regression Models

2.2.1 Description

A first approach to control for spatial dependence is to extend the basic Interval Regression model by the following specification:

$$\begin{aligned} \tilde{y} &= \rho W_1 \tilde{y} + X\beta + u \\ u &= \lambda W_2 u + \epsilon \\ \epsilon &\sim \mathcal{N}\left(0, \sigma^2 I_N\right) \end{aligned} \tag{1}$$

where $\tilde{y} = \ln(y^*)$, N is the number of observations, X is a $N \times k$ matrix of regressors, ρ is the spatial dependence parameter. Whenever $W_1 \neq 0$ and $W_2 \neq 0$, we assume that $W_1 = W_2 = W$ which are $N \times N$ standardized spatial weight matrices.

W tells us whether any pair of observations are neighbors. For example, if dwelling i and dwelling j are neighbors then, $w_{ij} = 1$ and zero otherwise. Whether or not any pair of dwellings are neighbors is based on whether or not they are located in the same geographical entity or

[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 03, 2013).

in contiguous spatial units. We follow Kim *et al.* (2003) by considering two spatial units as contiguous when they share a common border.

We assume that there are S spatial entities. Any spatial entity l is populated by N_l individuals, with $\sum_{l=1}^{S} N_l = N$. Therefore, if $y^*_{s_l}$ denotes the $N_l \times 1$ vector of rents paid by the N_l households living in the lth spatial entity, the $N \times 1$ vector of rents paid by all the households of the sample is $y^* = \left({y^*_{s_1}}', \cdots, {y^*_{s_l}}', \cdots, {y^*_{s_S}}'\right)'$.

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.

Several techniques are used to estimate this spatially dependent discrete choice model in a purely cross-sectional setting. A comprehensive review of those techniques may be found in Flemming (2004). In this paper we opt for the Gibbs Sampling approach since it outperforms the most relevant alternative methods: the Recursive Importance Sampling (RIS) simulator and the Expectation Maximization (EM) algorithm. Indeed, while providing results that are similar to those of the RIS simulator, it is computationally and conceptually simpler (Bolduc *et al.*, 1997). Moreover, the Gibbs Sampler method overcomes the problem encountered in the estimation of standard errors by the EM algorithm because the standard errors of the estimates are derived directly from the posterior parameter distributions.

Specification (1) describes the SAC model. This model contains spatial dependence in both the dependent variable and the disturbances (LeSage and Pace, 2009). If $W_2 = 0$, then model (1) simplifies to the SAR model. The SAR model implicitly assumes that $W_1\tilde{y}$, the spatially weighted average of housing prices in a neighborhood, affects the price of each dwelling (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" (Kim *et al.*, 2003, p. 29).

Assuming that $W_2 = 0$ allows to develop an extension of the SAR model, the Spatial Durbin model (SDM) which controls for spatial dependence in the dependent variable and the explanatory variables. The formal expression of the SDM model is shown in (2):

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

where $\underline{X}$ denotes the matrix of regressors with the intercept excluded. The $W\underline{X}$ term allows the physical attributes of neighboring dwellings to impact the rent of each dwelling. It further captures how the price of houses in one BSU depends on the environment quality and the neighborhood characteristics of contiguous BSUs (Brasington and Hite, 2005).

The SAC model is the additive combination of the Spatial Error Model (SEM) and the SAR Model. Whenever $W_1 = 0$, the SAC model collapses to the SEM Model defined by the following specification:

$$
\tilde{y} = X\beta + u \quad u = \lambda W u + \epsilon \quad \epsilon \sim \mathcal{N}\left(0, \sigma^2 I_N\right). \tag{3}
$$

2.2.2 Estimation of SAR or SDM Interval Regression Models

SAR and SDM models can be written by the same expression

$$
\begin{aligned}
\tilde{y} &= \rho W \tilde{y} + Z\delta + \epsilon \\
\epsilon &\sim \mathcal{N}\left(0, \sigma^2 I_N\right)
\end{aligned} \tag{4}
$$

where $Z = X$ for the SAR model and $Z = [X \; W\underline{X}]$ for the SDM model.

This implies that the likelihood function for SAR and SDM models can be expressed in the same way

$$
L\left(\tilde{y}, W | \rho, \delta, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \left|I_N - \rho W\right| \exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\} \tag{5}
$$

where

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

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

$$
p\left(\delta, \sigma^2, \rho | \tilde{y}, Z, W\right) \propto \left|I_N - \rho W\right| \left(\sigma^2\right)^{-(N/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\} \tag{6}
$$

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

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

$$\begin{aligned}
\delta|\sigma^2_{(0)}, \rho_{(0)}, \tilde{y}_{(0)} &\sim \mathcal{N}\left(\tilde{\delta}, \sigma^2_{(0)}\left(Z'Z\right)^{-1}\right); \quad (7)\\
\tilde{\delta} &= \left(Z'Z\right)^{-1}\left(Z'A\tilde{y}\right); \quad (8)\\
A &= I_N - \rho W. \quad (9)
\end{aligned}$$

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

$$\sigma^2|\delta_{(1)}, \rho_{(0)}, \tilde{y}_{(0)} \sim \left(\sigma^2\right)^{-(N/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\}. \quad (10)$$

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

$$p\left(\rho|\delta, \sigma^2, \tilde{y}\right) \propto |A| \exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\}. \quad (11)$$

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

$$\begin{aligned}
\tilde{y}|\delta_{(1)}, \sigma^2_{(1)}, \rho_{(1)} &\sim TMVN\left(\mu, \Omega\right); \quad (12)\\
\mu &= \left(I_N - \rho W\right)^{-1} Z\delta; \quad (13)\\
\Omega &= \sigma^2\left[\left(I_N - \rho W\right)'\left(I_N - \rho W\right)\right]^{-1}; \quad (14)
\end{aligned}$$

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 δ and σ^2. Therefore, sampling for the parameter ρ must proceed using an alternative approach, such as numerical integration or Metropolis-Hastings.

2.2.3 Estimation of a General Spatial Interval Regression Model

Assuming that $W_1 = W_2 = W \neq 0$, we have to estimate the more general SAC model described in (1). With $Z = X$, the associated likelihood concentrated for the parameters δ and σ^2 take the following form (LeSage and Pace, 2009):

$$\begin{aligned}
p\left(\tilde{y}|\delta, \sigma^2, \rho, \lambda\right) &\propto |A|\,|B| \exp\left(-\frac{1}{2\sigma^2}\left(BAy - BZ\delta\right)'\left(BAy - BZ\delta\right)\right) \quad (15)\\
\text{with} \quad A &= I_N - \rho W,\\
\text{and} \quad B &= I_N - \lambda W
\end{aligned}$$

Following LeSage and Pace (2009), we draw the estimates of this distribution through the following 5 steps Gibbs sampler:

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

$$\begin{aligned} \delta|\sigma^2_{(0)}, \rho_{(0)}, \lambda_{(0)}, \tilde{y}_{(0)} &\sim \mathcal{N}\left(\tilde{\delta}, \sigma^2_{(0)} \left(Z'B'BZ\right)^{-1}\right); && (16) \\ \tilde{\delta} &= \left(Z'B'BZ\right)^{-1} \left(Z'B'A\tilde{y}\right). && (17) \end{aligned}$$

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

$$\begin{aligned} \sigma^2|\delta_{(1)}, \rho_{(0)}, \lambda_{(0)}, \tilde{y}_{(0)} &\sim \left(\sigma^2\right)^{-(N/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon'\epsilon\right)\right\}, && (18) \\ \text{with} \quad \epsilon &= B\left(A\tilde{y} - Z\delta\right) && (19) \end{aligned}$$

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

$$p\left(\rho|\delta, \sigma^2, \lambda, \tilde{y}\right) \propto |A| \left|B\left(\lambda_{(0)}\right)\right| \exp\left\{-\frac{1}{2\sigma^2}\left(\tilde{B}\left(A\tilde{y} - Z\delta\right)\right)'\left(\tilde{B}\left(A\tilde{y} - Z\delta\right)\right)\right\}. \quad (20)$$

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

$$p\left(\lambda|\delta, \sigma^2, \rho, \tilde{y}\right) \propto \left|A\left(\rho_{(0)}\right)\right| |B| \exp\left\{-\frac{1}{2\sigma^2}\left(B\left(\tilde{A}\tilde{y} - Z\delta\right)\right)'\left(B\left(\tilde{A}\tilde{y} - Z\delta\right)\right)\right\}. \quad (21)$$

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

$$\begin{aligned} \tilde{y}|\delta_{(1)}, \sigma^2_{(1)}, \rho_{(1)}, \lambda_{(1)} &\sim TMVN\left(\mu, \Omega\right); && (22) \\ \mu &= \left(I_N - \rho W\right)^{-1} Z\delta; && (23) \\ \Omega &= \sigma^2 \left[A'B'BA\right]^{-1}. && (24) \end{aligned}$$

When sampling for the parameter ρ (λ), we rely on the current value for λ (ρ) in $|B|$ $(|A|)$ which we denote $\left|B\left(\lambda_{(0)}\right)\right|$ $\left(\left|A\left(\rho_{(0)}\right)\right|\right)$, and $B\left(\lambda_{(0)}\right) = \tilde{B}$ $\left(A\left(\rho_{(0)}\right) = \tilde{A}\right)$. Therefore, we can still perform our univariate numerical integration scheme to find a normalizing constant and produce a CDF from which we can draw by inversion.

2.3 Two-step estimation with locational fixed effects

Spatial econometrics models described in section 2.2 are designed for purely cross-sectional settings. However, since our database has several observations (dwellings) for each basic spatial unit, it has the structure of a panel. Therefore, we may control for location heterogeneity by using an interval regression model with location/fixed effects.

Such an estimation strategy has an additional advantage as it allows to avoid the endogeneity bias caused by locational characteristics. One may suspect reverse causality between dwelling rents and observed location characteristics like average income or accessibility. While we may expect dwelling rents to be high as a resultant of high average income in the area, we may as well consider areas with high dwelling rents as a sign of attractiveness for high income households because they are expected to host better schools or because they host socio-economic peers.

Since reverse causality is more likely to concern location invariant attributes, we may address this endogeneity problem by retrieving locational fixed effects. Hence, we use an alternative approach to capture spatial dependence based on a two-stage procedure (Ahlfeldt, 2011). In the first stage we obtain estimated values of dwelling rents by carrying out an interval regression on a set of dwelling structural attributes as well as a set of locational-fixed effects that capture location heterogeneity

$$\begin{aligned} \tilde{y}_i &= z_i'\alpha + \mu_l + \varepsilon_i \ \text{ or equivalently} && (25) \\ \tilde{y} &= Z_s\alpha + Z_\mu \mu + \varepsilon && (26) \end{aligned}$$

with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_S),\ Z_s = \begin{bmatrix} z_1' \\ z_2' \\ \vdots \\ z_N' \end{bmatrix}$ and $\mu = (\mu_1, \mu_2, \dots, \mu_S)'$ where Z_s is a $N \times k_s$ matrix of dwellings structural characteristics[2], α is the corresponding vector of coefficients, μ is a vector of unobserved location fixed effects, and Z_μ is a $N \times S$ matrix of spatial entities dummies. After this first step, we retrieve the fixed effects and obtain an estimate of the average log of dwelling rent within former townships $LRENT$, adjusted for dwelling characteristics.

Then in a second stage, we regress $LRENT$ on location characteristics

$$LRENT_j = LOC_j'\gamma + \eta_j \tag{27}$$

where $LRENT$ is the average log of dwelling rents estimated from specification (26), LOC is a row-vector of location controls, γ is the corresponding column vector of coefficient estimates and η is a random error term.

This two-step estimation strategy ensures that there is no endogeneity bias due to the correlation between unobserved location characteristics and observed neighborhood and environment quality attributes.

[2] with k_s the number of regressors derived from the structural characteristics plus 1 for the constant.

We may account for spatial dependence in this second stage by extending (27) into a general spatial model

$$\begin{aligned} LRENT &= \delta M_1\, LRENT + LOC\gamma + \eta \\ \eta &= \phi M_2\, \eta + \xi \\ \xi &\sim \mathcal{N}\left(0, \sigma_\xi^2 I_S\right) \end{aligned} \tag{28}$$

where δ is a spatial dependence parameter, M_1 and M_2 are $S \times S$ standardized spatial weight matrices which tell whether two spatial entities are neighbors or not. As before, (28) simplifies to a SAR model if $M_2 = 0$. Moreover, in estimating (28) we will assume that $M_1 = M_2 = M$.

3 Delineation of the study area and basic spatial unit

3.1 Delineation of the study area

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, namely, administrative, morphological (Donnay and Lambinon, 1997; Tannier *et al.*, 2011; Van Hecke *et al.*, 2009) and functional (Cheshire, 2010; Van Hecke *et al.*, 2009; Vandermotten *et al.*, 1999; Blondel *et al.*, 2010[3], Thomas *et al.*, 2012). Table 2 and Figure 1 in Appendix B present macrozones that are consistent with eight delineations of Brussels. One of those delineations, the Brussels Capital Region, corresponds to a purely administrative definition of Brussels. Another delineation, Brussels (operational) agglomeration corresponds to a morphological definition of Brussels. It is one of the macrozones defined by the Van Hecke *et al.* (2009) nomenclature of Belgian urban regions.

The other delineations represent merely functional definitions of Brussels. They are macrozones that, because of the strong socio-economic ties of their peripheral rings with the Brussels urban center, may serve to define the Brussels urban functional region. From the Van Hecke *et al.* (2009) nomenclature, we may consider the Brussels urban region on one hand and the

[3]The methodology used by Blondel *et al.* (2010) to define a functional agglomeration is quite innovative. They use a mathematical method, based on origin-destination matrices, which allows networks to be divided into coherent groups in a natural and automatic manner (modularity criteria). This generates a mathematically optimal partition of space. Another innovative method is the one used by Rozenfeld *et al.* (2011). Rather than using informative but somewhat arbitrary legal or administrative definitions, they build on the City Clustering Algorithm (CCA) to construct cities based on geographical features of high-quality micro data.

Brussels residential urban complex on the other hand. We may also consider Brussels "extended residential commuting complex" which is the union of the Brussels and Leuven residential commuting complexes. Stratec, an independent consultancy company, has proposed other definitions: "Stratec RER Area" and "Stratec Extended RER Area" on the basis of commuting ties through the railway network transportation system.[4]

The 2001 Belgian census includes the information from 177,721 dwellings pertaining to the Brussels Capital Region 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, Brussels Capital Region concentrates more than half of the rented dwellings of the most extensive definitions of Brussels. In Figure 1, the "extended residential commuting complex" corresponds to the union of the "Extended residential commuting area" (light blue), the Suburb (medium blue), and the Agglomeration (dark blue). The hatched zone corresponds to "Stratec Extended RER Area" and the small and central area with a black border depicts the Brussels Capital Region.

3.2 Basic spatial units

Environment quality attributes and neighborhood characteristics can 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 and the statistical sector. Those spatial units are nested. The Union area is divided in 147 municipalities, 742 former townships and 5,127 statistical sectors. Variables measured at the municipal 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 Data description

In hedonic regression models, dwelling rent is characterized by 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 and accessibility. The third type of characteristic relates to environmental quality, such as air pollution and proportion of agricultural areas or forests.

[4]These spatial entities are essentially based on the Official RER Area defined by ministerial decree and composed of 126 municipalities (Moniteur Belge, 2004).

In principle, all the features pertinent for the characterization of market prices should be included in a hedonic regression. However, as Butler (1982) notices, this cannot be done in practice for two reasons. Firstly, the number of characteristics is unmanageably large and the 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, 2010). The variables used here are selected in that spirit.

4.1 Dwelling structural characteristics

The 2001 Belgian census includes several different housing attributes that can be taken into account into a hedonic regression: type of dwelling, number and type of rooms (separated kitchen(s), fitted kitchen(s) integrated in other rooms, separated lounge(s), bedroom(s), toilet(s), bathroom(s) etc.), total surface, building period, renovation, type of heating system, furnished or not, 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 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 below 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 than or equal to 991,57. As previously discussed, this way of coding the rent variable has an impact on the choice of the appropriate estimation strategy.

Table 3 in Appendix C provides the list of the dwelling attributes used in this paper.[5]

[5]The composite quality index presented in Table 3 is built on dwellings physical attributes by allocating 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 85 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 gazzling. Therefore, it is close to the index

4.2 Environment quality attributes

4.2.1 Land Cover information

As in Goffette–Nagot *et al.* (2011), we aggregate some of the data in the CORINE database to produce synthetic indicators of biophysical land cover at various aggregation levels. The CORINE (Coordination of Information on the Environment) land-cover database provides a detailed inventory of the biophysical land cover in Europe using forty-four classes. It was obtained in the form of a raster dataset that was used to produce the following synthetic variables:

- Percentage of each municipality and former township covered by Forest (PERFOR). This proportion is computed as the percentage of the 250m by 250m grid cells entirely covered with forest in each municipality and former township.

- Percentage of each municipality and former township covered by Agriculture (PERAR). This percentage represents the share of the 250m by 250m grid cells entirely covered with arable land in each municipality and former township.

4.2.2 Pollution indicator

Several hedonic price studies attempt to find out whether air quality is associated with property value. Most of those studies suggest that air pollution affects property value negatively (Boyle and Kiel, 2001; Kim *et al.*, 2003; Brasington and Hite, 2005; Anselin and Lozano–Gracia, 2008). The Belgian Interregional Cell for the Environment (IRCEL–CELINE) provides information on air quality in all 3 Belgian Regions through a raster file. This raster allowed us to compute the average concentration of PM_{10} in every Belgian municipality.[6]

4.2.3 Slope indicator

The average gradient of relief, noted SLOPE, is computed for 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.*, 2011).

---

proposed by Vanneste *et al.* (2007). 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.

[6] PM_{10} denotes particulate matter (suspended particles) with a diameter of $10\mu m$ or less. The use of the concentration of PM_{10} pollutants as a proxy of air quality may be found in other studies such as Anselin and Lozano–Gracia (2008) and Montero *et al.* (2008).

4.3 Neighborhood attributes

4.3.1 Median and average income

Localities where most inhabitants have a high social and economic status are characterized by more expensive dwellings. We use median or average income by tax declaration as indicators of social and economic status (respectively noted MEDINC and AVINC). Those data were obtained from Belgian National Statistical Institute for 2001. Median income is available at the level of each statistical sector and each municipality, while average income is available at all the spatial scales. Goffette–Nagot *et al.* (2011) also use median declared income in a spatial analysis of land price in Belgium.

4.3.2 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 point (Goffette–Nagot *et al.*, 2011). 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. Several indicators may be used to capture accessibility.

We may firstly use a potential accessibility measure computed by Vandenbulcke (2007), noted POT_JOBS. This measure of accessibility at zone i to all populations or jobs D_j in zone j (also considered as the attraction of destination j) is formally defined as $A_i = \sum_{j=1}^{n} D_j F(t_{ij})$ where t_{ij} is the traveling time between i and j along the Belgian road network,[7] and $F(t_{ij})$ the impedance function. Vandenbulcke (2007) computed the measures of potential accessibility to jobs and to population for all the 2616 Belgian former townships. We used their measure of the potential accessibility to jobs. The presence of D_j in the expression of this potential measure of accessibility allows it to somewhat capture the market potential of the basic spatial unit in terms of the number of jobs that may potentially access to it.

We use another measure of market potential, noted MARKPOT. Following Harris (1954), Head and Mayer (2004) propose the inverse-distance-weighted sum of incomes $M_i = \sum_{j=1}^{n} \frac{E_j}{d_{ij}}$ as a measure of market potential. In this expression d_{ij} is the radial distance between the centroids of

[7]Using exclusively the road network to compute the temporal distance between two locations is grounded on the fact that in 2001 82% of all commuters' journeys are made by car, while public transport only accounts for 14% (the other 4% represents travel by bus companies) (Vandenbulcke *et al.*, 2009).

the basic spatial units i and j, and E_j is the income in j. To measure the distance from a basic spatial unit to itself d_{ii}, we follow Head and Mayer (2004) by assuming that each basic spatial unit is a disk in which all production concentrates at the center and consumers are uniformly distributed throughout the rest of the area. Then, the average distance between a producer and a consumer is given by

$$d_{ii} = \int_0^R r \frac{2r}{R^2} dr, \tag{29}$$

where R denotes the radius of the disk, and $\frac{2r}{R^2}$ is the density of consumers at any given distance r to the center. R is obtained as the square root of the area A divided by π. From (29), we may derive the following expression of the internal distance $d_{ii} = \frac{2}{3}R = \frac{2}{3}\sqrt{A/\pi}$.

4.3.3 Population density

Urban theory and recent contributions in geographical economics outline the influence of population density on housing rents. Combes *et al.* (2011) find that land prices in France are higher in densely populated areas. Using 2001 data population from the Belgian Statistical institute, we computed the density of population at the level of each former township (noted DENS_POP).

5 Results

In this section, we present estimation results for different specifications. We firstly present results from the benchmark model which further allows us to investigate MAUP and delineation issues. We discuss spatial autocorrelation with the spatial interval regression model. Eventually, we discuss MAUP, delineation issues and spatial autocorrelation with the two-step estimation procedure.

5.1 Interval Regression Model

We use the Interval Regression Model (IRM) to serve as a benchmark for comparison with more appropriate models. Since it is less computationally demanding than the spatial interval regression model, the IRM enables estimations with huge databases. This allows for enough variation of environment quality and neighborhood attributes to compare estimation results with different basic spatial units and with different delineations.

5.1.1 Impact of the choice of the delineation of the study area

Table 8 displays results of the IRM for different delimitations.[8] Standard errors are clustered on statistical sectors in order to account for variance shifting in the error across space. Most of the results for different delineations of Brussels have the same sign but show differences in magnitude. This suggests that the choice of the limits of the agglomeration has a strong impact on econometric 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. Dwellings 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. This is not surprising since heating with coal or wood is an indicator of the poor quality of a dwelling (Vanneste *et al.*, 2007). Dwellings using electricity as the source of energy are more expensive. Dwellings with bathrooms, toilets, double glazing, and wall isolation are more valuable. Moreover, the more garage places a dwelling has, the more expensive it is. All other things being equal, apartments are cheaper than single-family houses.

Other results are more puzzling. Studios and lofts are, ceteris paribus, more expensive than other dwellings. Furnished dwellings are less expensive. While this may seem paradoxical, this can be due to the fact that this category of dwellings targets mostly low-income categories such as students who cannot afford to furnish and to renovate their dwellings.

Most of the results about environmental quality variables and neighborhood attributes are in line with expectations. Rental prices are low when the pollution indicator is high. This is not surprising since we expect greater demand for dwellings located in less polluted areas. The coefficient of **LNSLOPE (AC)** is positive. This confirms the assumption that hilly landscapes are more attractive to residents.[9] The coefficient obtained for the variable **LNPERFOR (AC)** is

[8]We chose the specification in Table 8, which uses the potential accessibility to jobs as a measure of accessibility, by comparing several alternative models in Table 4 in Appendix D in terms of goodness of fit criteria (AIC and SBC). In Table 4, our preferred specification is displayed in column (2). It outperforms the corresponding semilog specification shown in column (1). Moreover, it has a better fit than models in the column (3) and the column (4) where the proxies of accessibility are respectively the density of population and the Harris Market Potential. While the density of population is not a "natural" measure of accessibility, it is significantly correlated with the potential accessibility to jobs (the pairwise correlation between **LNPOT_JOBS** and **LNDENS_POP** is significant and equal to 0.7713). Therefore, due to multicollinearity, simultaneously inserting the potential accessibility to jobs and the density of population yields a non significant coefficient for **LNDENS_POP**.

[9]Goffette–Nagot *et al.* (2011) test this assumption by estimating a hedonic regression with data collected at the municipal level. However, as they obtained negative coefficients, they were not able to confirm that hypothesis.

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 **LNPERFOR (AC)** variables are clearcut. They emphasize that households value positively neighborhoods with environmental amenities.

The sign of the **LNPERAR (AC)** coefficient is more difficult to interpret; it suggests that dwellings located in neighborhoods largely covered by agriculture are less valuable. This may indicate 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. Dwelling rents increase with **LNPOT_JOBS** the indicator of potential accessibility to jobs, which suggests that there is a greater demand for more accessible residential areas.[10] Finally, the coefficient of the variable **LNMEDINC** is positive and significant, indicating that dwellings located in wealthy neighborhoods are more expensive.

Insert Table 8 about here.

Table 8 shows the impact of the choice of the limits of the agglomeration on econometric results through estimations of separate samples. The econometrics literature recommends rather to perform an estimation with interaction terms based on the largest sample. This procedure has the advantage of 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 potential shortcoming of increasing the multicollinearity between regressors. Table 5 in Appendix E shows that several interaction terms are significant which clearly imply that the coefficients of several regressors vary significantly across different definitions of the study area.[11] This further confirms the sensitivity of statistical results to the delineation of the study area.

The use of household data helps greatly to improve the results regarding this variable.

[10]Results for the **LNPERAR (AC)** and the **LNPOT_JOBS** will be challenged in the two-stage estimation approach where we control for endogeneity.

[11]Considering specifically the elasticity of potential accessibility and of its interaction terms, we find an elasticity of 0.3627 for the Union area and of 0.5230 for **DBRXCAP*LNPOT_JOBS** — the interaction term between potential accessibility to jobs and the dummy depicting the pertainance of a dwelling to Brussels Capital Region — and of -0.5026 for **DAGGLOP*LNPOT_JOBS** — the interaction term between potential accessibility to jobs and the dummy related to the agglomeration macrozone. Therefore, the implied elasticities for the Union area, Brussels Capital Region, and the Operational Agglomeration delineation are respectively 0.3627, 0.8857 and -0.1399. We can note that the corresponding results in Table 8 are respectively 0.3548, 0.3663 and 0.1917. While the ordering of the elasticities remains the same, the model with interaction terms overestimate the elasticity of potential accessibility for the Union area and Brussels Capital Region and underestimate it for the agglomeration macrozone. Therefore, the multicollinearity caused by the addition of interaction terms is likely to affect the precision of the estimates.

To sum up, choices regarding the limits of the study area are not benign regarding econometric results. The reason of those discrepancies may lie on a spatial heterogeneity argument. The different delineations imply distinct geographic structures: the larger the study area, the larger will be the implied proportion of rural hinterland; which is by nature less urbanized and less accessible from the main centers of activities and employment. 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 area 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 elasticity of the potential accessibility indicator rises from 0.1917 for the Operational Agglomeration macrozone to 0.3663 for the Brussels Capital Region, which implies a 91% increase. Such a discrepancy may lead to highly mistaken conclusions in terms of land use and transport policies. 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 delineation of the study area, we now investigate the impact of the aggregation scale on econometric findings. Briant *et al.* (2010) assess the impact of size and shape distortions on the behavior of simple regression coefficients. Concerning the size distortion, they find that if the aggregation distortion on the explanatory variables and the dependent variable 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. 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. Our dependent variable and the dwellings structural attributes are individual characteristics. Therefore, in our analysis aggregation biases concern only environmental and neighborhood variables.

Table 9 compares estimation results when the median income by tax declaration is measured successively at the statistical sector and at the municipal levels while Table 10 compares coefficient estimates when the average income by tax declaration is measured successively at the statistical sector, the former township and at the municipal levels. They show that the coefficients of the logarithm of the median and average income are higher when the scale of measure of those variables is higher. Similar results are obtained for other neighborhood and environmental variables as shown in Table 6 in Appendix F.

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 may be the following: the larger the BSU, the lower the variance of a variable. Since 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.

Insert Table 9 about here.

Changing the scale of the basic spatial unit for one variable also impacts the coefficients of the other variables. In Table 9, the most important effects are observed in the intercept, which is more than twice as large in the second specification, in the **GARDEN** coefficient, which is not significant in the first specification and more than twice as large in the second specification, and in the **LNPM10** coefficient which is not significant in the second specification.[12]

In Table 10, we also observe substantial changes in the intercept, which is more than doubled from the first to 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** 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 changes in the size of the basic spatial units.

Insert Table 10 about here.

The results just described outline the necessity of using the finest spatial scale for the definition of environmental 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 misleading conclusions in terms of the policy recommendations drawn from 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 with the pollution indicator which, because of raster resolution, can only be computed at the municipal level. In such cases of spatial scale constraints, one should be aware of the potential biases.

[12] By not significant we mean not significant at the 10% significance level.

Using inappropriate spatial scales may trigger endogeneity of the "error in variables" kind. In the case of a continuous dependent variable, Anselin and Lozano–Gracia (2008) use a Spatial 2SLS method to handle both spatial dependence and endogeneity generated by spatial interpolation of air quality values. While this approach may be relevant for handling biases due to spatial scales in cases where the regressand is continuous, it is tremendously challenging in the interval regression context. However, it may be considered in the second stage of the two-step procedure with locational fixed effects and deserves to be explored in future contributions.

5.2 Spatial Interval regression models

Let us now investigate the substantive spatial dependence issue through a Spatial Autoregressive Interval Regression (SARIR) model, a Spatial Durbin Interval Regression model (SDMIR) and a General Spatial Interval Regression (SACIR) model. We have not been able to run those spatial models algorithms on the full dataset. Indeed, they are very demanding in terms of computational resources. So we ran those algorithms on two subsamples of respectively 2,969 observations and 62 statistical sectors (Sample 1) and 2,565 observations and 81 statistical sectors (Sample 2).[13]

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.[14] Figures 2 and 3 depict the average income in Brussels Capital Region (BCR) as well as the location of Sample 1 and Sample 2 dwellings in the BCR.

Unfortunately, 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 11 displays the results of the basic interval regression model as well as those of the SARIR, the SDMIR and the SACIR algorithms. The SDMIR and the SACIR models seem less credible than the SARIR model because the disturbance spatial dependence parameter λ and the coefficients of

[13]All our Gibbs sampler algorithms are written in GAUSS 12. They imply 2,500 draws with an initial 500 draws "burn-in" sequence for the SAR and the SDM algorithms, and an initial 1000 draws "burn-in" sequence for the SAC algorithm. For a 2,969 observations sample, the SAR algorithm takes a total time of about 3 days, 11 hours, 39 minutes and 54 seconds on a Dell Precision workstation with a 64 GB RAM and with 2 processors of respectively 2.80 GHz and 2.79 GHz processor speed.

[14]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 and 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.

the regressors spatial lags are not significant. One has to be very cautious in comparing Interval Regression to SARIR, SDMIR and SACIR estimates. As for OLS, Interval Regression estimates have a straightforward interpretation as partial derivatives of the dependent variable with respect to an explanatory variable. Such an interpretation is possible because in the IR model the information set for an observation i contains only exogenous or predetermined variables associated with observation i. In spatial models which contain spatial lags of the dependent variable, the interpretation of parameters becomes richer and more complicated (LeSage and Pace, 2009). Basically, such models allow a change in the explanatory variable for a single dwelling to potentially affect the dependent variable in all other dwellings. Therefore, the marginal effects in those models have a formal expression way more complex than in OLS settings (the formal expression of those marginal effects can be found in LeSage and Pace (2009, p.35-36)). In Table 11 we do not compare marginal effects of the regressors across non-spatial and spatial models, but rather their direct effects (i.e. the "true" effect of the regressors, after that spatial dependence has been accounted for (Kim *et al.*, 2003).). The results of the IRM and the SARIR model differ substantially only for the spatial dependence parameter and the intercept. In the first sample estimations, the intercept is much lower in the spatial models than the one obtained in the benchmark model. For the second sample, there appears to be a significant discrepancy in the intercept in the SACIR model, but not in the SARIR and SDMIR models. The substantial reduction of intercept in the SAR and SAC models may be an indication that the omitted-variable bias is significantly reduced in the spatial models.

Pace and LeSage (2010) establish that, in the presence of spatial dependence, non spatial models like OLS amplify omitted variable bias. Moreover, they show that estimates from the SDM model shrink the bias relative to OLS. In Appendix G, we show formally how the non spatial models compound this omitted variable bias.

Insert Figures 2 and 3 about here.

Omitting the spatial lag term, as in the non-spatial IRM, would definitely entail biased estimates. Such biases are perceptible in the **LNMEDINC (SS)** coefficient which is 17% higher in the non-spatial IRM estimation based on Sample 1 and 8% higher in the IRM estimation based on Sample 2 than in SARIR corresponding estimations. By acknowledging the impact of dwelling rents nearby in space, the spatial model implies a lower impact of the median income. Another noteworthy observation may be made about the **LNMEDINC (SS)** coefficient: it is the only coefficient of an environment or a neighborhood attribute that is significant for all specifications. As explained earlier, since the other environment and neighborhood variables are measured at a more aggregated level, they do not have 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 and the **LNMEDINC (SS)** coefficient are higher for Sample 2 than for Sample 1. Those results indicate that the neighborhood has a stronger impact in the determination of Sample 2 dwelling rents than in Sample 1. The difference between Sample 1 and Sample 2 results yields some evidence of spatial heterogeneity in the estimation. Spatial heterogeneity 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).

Insert Table 11 about here.

5.3 Two-stage estimation with locational fixed effects

While the SARIR model may mitigate the omitted variable bias by inserting a spatial lag term, it does not control for the endogeneity generated by the simultaneous causality between the dwelling rents and the observed locational characteristics, like the average income. Indeed, while we may expect dwelling rents to be high as a resultant of high average income in the area, we may as well consider areas with high dwelling rents as a sign of attractiveness for high-income households because they are expected to host better schools or because they host socio-economic peers. Since reverse causality is more likely to concern location invariant attributes, the two-step estimation procedure address this endogeneity problem by retrieving locational fixed effects at the first step.

Insert Table 12 about here.

Table 7 in Appendix H provides results for the first stage. They are as expected and in line with the previous estimations. Before discussing the effects of a choice of a specific delineation, few remarks are in order. The first one is that conversely to earlier results where the percentage of the basic spatial unit covered by agriculture had a negative impact on dwelling rents while potential accessibility had a positive elasticity, Table 12 indicates reverse effects for those two variables. The private dwelling market of the Brussels Metropolitan area values dwellings close to agricultural areas and with a lower accessibility to the main employment centers. Clearly, inconveniences of urban life (congestion, pollution etc.) reflect in marginal prices.

Insert Table 13 about here.

Moreover, the SAC model is validated in this second-step estimation for the Urban Region and the ERUC samples since we obtain significant disturbance spatial dependence parameters. It provides more precise estimations (lower mean square error) and a better fit (Higher Pseudo $\bar{R}^2$). As before, choices of delineation impact statistical results. From the agglomeration to the Union samples, we obtain different estimates. This holds for OLS and spatial specifications. Therefore, controlling for substantive spatial dependence does not mitigate bias due to the delineation choices. This further highlights the necessity to delineate the study area consistently.

Regarding the impact of the choice of the basic spatial unit, the second-step estimation provides outcomes that are similar to the results obtained with the Interval Regression Model. Table 13 compares results when the median income by tax declaration is measured successively at the statistical sector and at the municipal levels. It confirms that the coefficient of the logarithm of the median income is higher when average income is measured at the municipal level.

6 Conclusion

Microsimulation tools require 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 limits 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 impact on statistical findings.

The goal of this paper was to address those issues by carrying out sensitivity analyses. Therefore, on the basis of the hedonic regression model, we have investigated the three aforementioned problems.

The delineation of the metropolitan area highly impacts the statistical estimations. We show that most of the coefficients vary significantly 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.

A second spatial aspect addressed by this paper 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.

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 coefficient 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.

We also accounted for the impact of substantive spatial dependence. As we obtained a statistically significant spatial dependence parameters from those estimations, our econometric results provide evidence of spatial dependence. The estimation of these spatial econometric models is likely to mitigate the omitted-variable bias which generally undermines traditional hedonic estimation.

However, the preceding approaches are "naive" in the sense that they do not address the endogeneity biases inherent to the hedonic regression model. To face this shortcoming, we apply a two-step estimation procedure with locational fixed effects. This estimation strategy shows that, while the explicit consideration of spatial effects in the econometric specification improves the fit of the model, it is still vulnerable to biases due to choices in terms of either the delineation of the study area or the basic spatial unit.

Conversely to the previous approaches for which we obtained a negative elasticity for the percentage of the basic spatial unit covered by agriculture and a positive elasticity for the potential accessibility, this method implies opposite effects for those two variables. This suggests that dwellings close to agricultural areas and with a lower accessibility to the main employment centers are sought-after and that endogeneity biases are not negligible.

Other results from this two-step estimation procedure further confirm previous findings. Therefore, even more attention should be paid to issues related to nuisance spatial dependence and to spatial heterogeneity. The future implementations of microsimulation models in the field of transportation and spatial planning should take that into consideration in order to avoid compounding all kinds of spatial biases.

Eventually, this contribution outlines two interesting research perspectives. The discrepancy between Sample 1 and Sample 2 SARIR estimation results provides strong evidence of spatial heterogeneity. Therefore, one could consider performing a careful investigation of that issue while addressing concomitantly endogeneity issues. Moreover, it would be interesting to use Spatial 2SLS for attempting to mitigate biases arising from choices of the spatial scale.

Acknowledgements. This research was conducted when the first author was appointed at CORE (UCLouvain) in the framework of the EU project named Sustaincity. We are grateful to all WP leaders and members of this project for their fruitful remarks. Special thanks are also addressed to Luc Bauwens and the anonymous referees for their valuable comments and suggestions on an earlier version of this paper. The usual disclaimer applies.

Appendix A: Descriptive statistics of continuous variables

Table 1: 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	ln of average concentration of PM_{10} by municipality	330,147	3.221	3.601	3.445	0.088
LNPERFOR (COM)	ln(PERFOR+1) by municipality	330,147	0	4.034	0.989	1.184
LNPERFOR (AC)	ln(PERFOR+1) by former township	330,118	0	4.035	0.847	1.204
LNPERAR (COM)	ln(PERAR+1) by municipality	330,147	0	4.472	2.123	1.775
LNPERAR (AC)	ln(PERAR+1) by former township	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 former township)	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
LNMEDINC (SS)	ln(MEDINC) (by statistical sector)	326,038	9.126	10.595	9.817	0.158
LNMEDINC (COM)	ln(MEDINC) (by municipality)	330,147	9.570	10.104	9.830	0.112
LNAVINC (SS)	ln(AVINC) (by statistical sector)	300,366	9.458	10.947	10.053	0.208
LNAVINC (AC)	ln(AVINC) (by former township)	326,335	9.646	10.665	10.085	0.163
LNAVINC (COM)	ln(AVINC) (by municipality)	330,147	9.756	10.574	10.102	0.149
LNDENS_POP	ln(DENS_POP) (by former township)	330,118	3.673	9.862	7.926	1.365
LNPOT_JOBS	ln(POT_JOBS) (by former township)	330,118	12.140	12.948	12.693	0.179
LMARKPOT	ln(MARKPOT) (by former township)	330,118	14.240	15.420	15.065	0.315

Appendix B: Different limits of Brussels

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

Spatial entity	Description	Nmun	Nobs
Brussels Capital Region (BCR)	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.

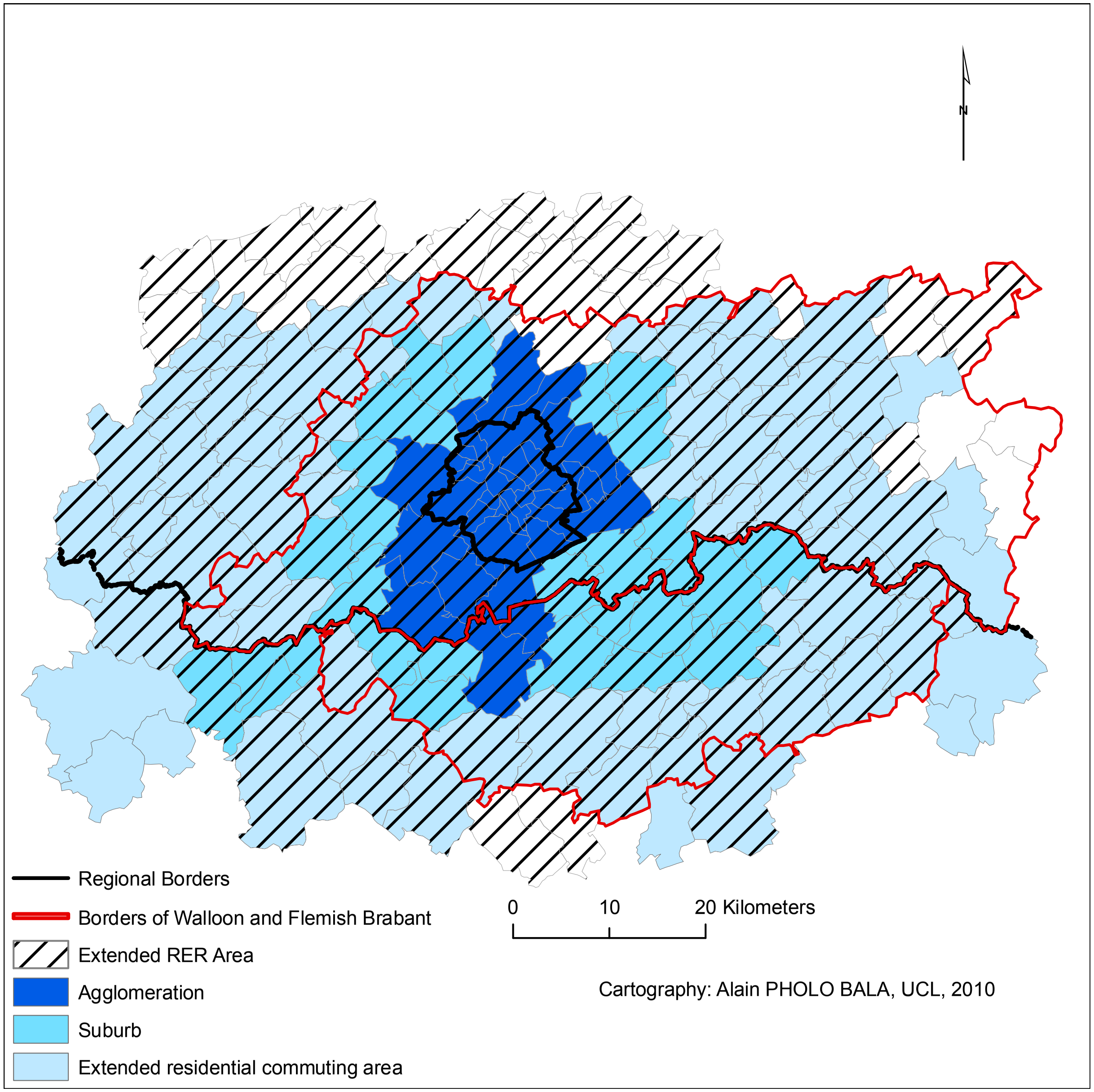

Figure 1: Different limits of Brussels agglomeration

Appendix C: List of dwelling structural characteristics

Table 3: List of variables linked to physical attributes of the dwellings.

Variable	Description
Type of dwelling dummies	
APP	dummy for apartments
OTHER	dummy for other kinds of dwelling
SING	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	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	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	Very good quality
Dummies related to the source of energy used for heating	
FUELA	Fuel
FUELB	Coal

Continued on next page

Table 3 – concluded from previous page

Variable	Description
FUELC	Wood
FUELD	Heat pump
FUELE	Electricity
FUELF	Natural gas
FUELG	Butane, propane
FUELH	Other source of energy
Garage dummies	
PARKA	No garage available
PARKB	Garage for one car
PARKC	Garage 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

Appendix D: Estimation results for various proxies of accessibility

Table 4: Interval regression: estimation results for various proxies of accessibility (Sample: Union).

Variable	**(1)**		**(2)**		**(3)**		**(4)**	
	Coeff.	**S.E**	**Coeff.**	**S.E**	**Coeff.**	**S.E**	**Coeff.**	**S.E**
Intercept	**4.6916**	(0.0812)	**-2.7275**	(0.5479)	**1.0568**	(0.3362)	**-1.2290**	(0.5392)
APP[1]								
OTHER	**0.0965**	(0.0413)	**0.1011**	(0.0414)	**0.0991**	(0.0415)	**0.0994**	(0.0414)
SING	**0.0702**	(0.0056)	**0.0738**	(0.0054)	**0.0674**	(0.0054)	**0.0736**	(0.0055)
LOFT	0.0188*	(0.0106)	0.0186*	(0.0106)	0.0183*	(0.0106)	0.0199*	(0.0104)
LNROOMS	**0.0964**	(0.0052)	**0.0964**	(0.0052)	**0.0952**	(0.0053)	**0.0965**	(0.0052)
SURFA[1]								
SURFB	**0.0474**	(0.0056)	**0.0463**	(0.0056)	**0.0452**	(0.0056)	**0.0450**	(0.0056)
SURFC	**0.0983**	(0.0060)	**0.0966**	(0.0060)	**0.0960**	(0.0059)	**0.0955**	(0.0060)
SURFD	**0.1854**	(0.0078)	**0.1833**	(0.0077)	**0.1837**	(0.0077)	**0.1823**	(0.0076)
SURFE	**0.2726**	(0.0102)	**0.2706**	(0.0101)	**0.2713**	(0.0102)	**0.2702**	(0.0101)
SURFF	**0.4376**	(0.0150)	**0.4357**	(0.0148)	**0.4397**	(0.0150)	**0.4362**	(0.0148)
RECENTBUILT	**0.1716**	(0.0061)	**0.1739**	(0.0064)	**0.1705**	(0.0064)	**0.1736**	(0.0066)
RENOVATION	**0.0651**	(0.0056)	**0.0648**	(0.0056)	**0.0644**	(0.0056)	**0.0647**	(0.0056)
FURNISH	**-0.1918**	(0.0095)	**-0.1896**	(0.0094)	**-0.1911**	(0.0095)	**-0.1879**	(0.0094)
HEATA[1]								
HEATB	**-0.0110**	(0.0056)	**-0.0121**	(0.0055)	**-0.0141**	(0.0054)	-0.0106*	(0.0055)
HEATC	-0.0016	(0.0186)	-0.0021	(0.0187)	-0.0057	(0.0188)	-0.0025	(0.0187)
HEATD	**-0.1925**	(0.0043)	**-0.1923**	(0.0042)	**-0.1956**	(0.0044)	**-0.1943**	(0.0042)
FUELA[1]								
FUELB	**-0.1675**	(0.0127)	**-0.1652**	(0.0127)	**-0.1655**	(0.0127)	**-0.1648**	(0.0127)
FUELC	**-0.0903**	(0.0238)	**-0.0880**	(0.0238)	**-0.0847**	(0.0234)	**-0.0875**	(0.0237)
FUELD	0.1001	(0.0639)	0.1070*	(0.0647)	0.1008	(0.0642)	0.1096*	(0.0645)
FUELE	**0.0820**	(0.0080)	**0.0847**	(0.0078)	**0.0858**	(0.0081)	**0.0879**	(0.0079)
FUELF	0.0028	(0.0050)	0.0076	(0.0048)	**0.0118**	(0.0050)	**0.0134**	(0.0049)
FUELG	-0.0275*	(0.0149)	-0.0281*	(0.0148)	-0.0270*	(0.0149)	-0.0291*	(0.0149)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table4 – concluded from previous page

Variable	**(1)**		**(2)**		**(3)**		**(4)**	
	Coeff.	**S.E**	**Coeff.**	**S.E**	**Coeff.**	**S.E**	**Coeff.**	**S.E**
FUELH	0.0048	(0.0386)	0.0126	(0.0390)	0.0052	(0.0388)	0.0148	(0.0391)
DGLAZING	**0.0309**	(0.0036)	**0.0296**	(0.0036)	**0.0309**	(0.0036)	**0.0289**	(0.0036)
WALLISO	**0.0388**	(0.0036)	**0.0408**	(0.0036)	**0.0404**	**0.0036**	**0.0407**	(0.0036)
BATHROOM	**0.2531**	(0.0078)	**0.2525**	(0.0080)	**0.2501**	(0.0080)	**0.2521**	(0.0080)
TOILET	**0.0395**	(0.0083)	**0.0401**	(0.0083)	**0.0401**	(0.0083)	**0.0413**	(0.0083)
PARKA[1]								
PARKB	**0.1041**	(0.0058)	**0.1039**	(0.0059)	**0.1051**	(0.0059)	**0.1023**	(0.0060)
PARKC	**0.1507**	(0.0080)	**0.1513**	(0.0080)	**0.1562**	(0.0082)	**0.1505**	(0.0080)
GARDEN	0.0078	(0.0052)	0.0072	(0.0052)	0.0090*	(0.0050)	0.0065263	(0.0051)
PM10	**-0.0068**	(0.0015)						
LNPM10			**-0.1554**	(0.0421)	**-0.1420**	(0.0481)	**-0.1752**	(0.0446)
POT_JOBS	**1.1**$\times10^{-6}$	(1.0×10^{-7})						
LNPOT_JOBS			**0.3548**	(0.0268)				
LNDENS_POP					**0.0362**	(0.0040)		
LNMARK_POT							**0.2084**	(0.0225)
SLOPE (AC)	**0.0443**	(0.0038)						
LNSLOPE (AC)			**0.1325**	(0.0076)	**0.1200**	(0.0075)	**0.1286**	(0.0074)
PERFOR (AC)	6.1×10^{-5}	(3.1×10^{-4})						
LNPERFOR (AC)			**0.0106**	(0.0024)	**0.0153**	(0.0026)	**0.0133**	(0.0025)
PERAR (AC)	**-0.0015**	(0.0002)						
LNPERAR (AC)			**-0.0198**	(0.0027)	**-0.0261**	(0.0039)	**-0.0173**	(0.0030)
MEDINC (SS)	**2.1**$\times10^{-5}$	(1.4×10^{-6})						
LNMEDINC (SS)			**0.4068**	(0.0285)	**0.4486**	(0.0257)	**0.3995**	(0.0276)
scale(σ)	0.3434		0.3428		0.3443		0.3435	
AIC	133,329.926		133,146.932		133,595.586		133,349.444	
BIC	133,669.148		133,486.154		133,934.808		133,688.666	
Nobs	70,839							

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Appendix E: Interval regression with interaction terms.

Table 5: Interval regression with interaction terms, Sample = Union.

Variable	Coeff.	S.E.	Variable	Coeff.	S.E.
Intercept	-0.3170	(0.7223)	**DAGGLOP*APP**	-0.0224	(0.0207)
DBRXCAP	**-6.7192**	(1.7876)	**DAGGLOP*OTHER**	-0.2719*	(0.1636)
DAGGLOP	**4.3241**	(1.9854)	**DAGGLOP*LOFT**	-0.0157	(0.0583)
DREGURB	-1.1976	(1.2842)	**DAGGLOP*LNROOMS**	-0.0373	(0.0244)
DCRU	**0.0318**	(0.0085)	**DAGGLOP*SURFA**	-0.0741	(0.0473)
APP[1]			**DAGGLOP*SURFB**	**-0.0837**	(0.0393)
OTHER	**0.1673**	(0.0546)	**DAGGLOP*SURFC**	**-0.0731**	(0.0364)
SING	**0.0709**	(0.0078)	**DAGGLOP*SURFD**	-0.0396	(0.0360)
LOFT	-0.0358	(0.0223)	**DAGGLOP*SURFE**	-0.0316	(0.0363)
LNROOMS	**0.0597**	(0.0070)	**DAGGLOP*RECENTBUILT**	0.0319	(0.0229)
SURFA[1]			**DAGGLOP*RENOVATION**	-0.0008	(0.0218)
SURFB	**0.0308**	(0.0083)	**DAGGLOP*FURNISH**	0.0711	(0.0498)
SURFC	**0.0734**	(0.0093)	**DAGGLOP*HEATA**	-0.0155	(0.0224)
SURFD	**0.1233**	(0.0098)	**DAGGLOP*HEATB**	-0.0304	(0.0299)
SURFE	**0.1732**	(0.0114)	**DAGGLOP*HEATC**	0.0049	(0.0810)
SURFF	**0.2778**	(0.0153)	**DAGGLOP*FUELA**	**0.3774**	(0.1566)
RECENTBUILT	**0.1327**	(0.0084)	**DAGGLOP*FUELB**	0.3280*	(0.1708)
RENOVATION	**0.0354**	(0.0090)	**DAGGLOP*FUELC**	**0.3899**	(0.1756)
FURNISH	**-0.1862**	(0.0162)	**DAGGLOP*FUELD**	0.0895	(0.1730)
HEATA[1]			**DAGGLOP*FUELE**	**0.3198**	(0.1577)
HEATB	-0.0078	(0.0094)	**DAGGLOP*FUELF**	**0.3516**	(0.1556)
HEATC	0.0050	(0.0379)	**DAGGLOP*FUELG**	0.3234*	(0.1683)
HEATD	**-0.2071**	(0.0064)	**DAGGLOP*DGAZZLING**	-0.0217	(0.0178)
FUELA[1]			**DAGGLOP*WALLISO**	-0.0012	(0.0183)
FUELB	**-0.1451**	(0.0149)	**DAGGLOP*BATHROOM**	0.0028	(0.0404)
FUELC	-0.0338	(0.0257)	**DAGGLOP*TOILET**	-0.0992*	(0.0515)
FUELD	0.1320	(0.0864)	**DAGGLOP*PARKA**	-0.0369	(0.0256)
FUELE	**0.1012**	(0.0108)	**DAGGLOP*PARKB**	-0.0026	(0.0284)
FUELF	**0.0258**	(0.0068)	**DAGGLOP*GARDEN**	-0.0370*	(0.0219)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table 5 – continued from previous page

Variable	Coeff.	S.E.	Variable	Coeff.	S.E.
FUELG	-0.0009	(0.0210)	**DAGGLOP*LNPM10**	0.0409	(0.2353)
FUELH	0.0518	(0.0476)	**DAGGLOP*LNPOT_JOBS**	**-0.5026**	(0.1204)
DGAZZLING	**0.0471**	(0.0064)	**DAGGLOP*LNSLOPE**	-0.0383	(0.0380)
WALLISO	**0.0576**	(0.0062)	**DAGGLOP*LNPERFOR (AC)**	**-0.0304**	(0.0104)
BATHROOM	**0.3197**	(0.0145)	**DAGGLOP*LNPERAR (AC)**	-0.0292	(0.0211)
TOILET	0.0342*	(0.0177)	**DAGGLOP*LNMEDINC (SS)**	**0.1959**	(0.0767)
PARKA[1]			**DREGURB*APP**	**-0.0486**	(0.0181)
PARKB	**0.0601**	(0.0067)	**DREGURB*OTHER**	-0.0005	(0.0863)
PARKC	**0.1167**	(0.0084)	**DREGURB*LOFT**	**0.1203**	(0.0422)
GARDEN	-0.0027	(0.0084)	**DREGURB*LNROOMS**	**0.0903**	(0.0174)
LNPM10	**-0.3042**	(0.0602)	**DREGURB*SURFA**	**-0.1322**	(0.0361)
LNPOT_JOBS	**0.3627**	(0.0382)	**DREGURB*SURFB**	**-0.1158**	(0.0287)
LNSLOPE	**0.0906**	(0.0110)	**DREGURB*SURFC**	**-0.1182**	(0.0262)
LNPERFOR (AC)	-0.0009	(0.0041)	**DREGURB*SURFD**	**-0.1202**	(0.0264)
LNPERAR (AC)	-0.0000	(0.0084)	**DREGURB*SURFE**	**-0.0795**	(0.0291)
LNMEDINC (SS)	**0.2034**	(0.0442)	**DREGURB*RECENTBUILT**	0.0435	(0.0159)
DBRXCAP*APP	**0.0911**	(0.0155)	**DREGURB*RENOVATION**	0.0191	(0.0167)
DBRXCAP*OTHER	0.1344	(0.1563)	**DREGURB*FURNISH**	**-0.0822**	(0.0368)
DBRXCAP*LOFT	-0.0122	(0.0445)	**DREGURB*HEATA**	0.0250	(0.0180)
DBRXCAP*LNROOMS	0.0079	(0.0196)	**DREGURB*HEATB**	0.0022	(0.0228)
DBRXCAP*SURFA	-0.0632	(0.0441)	**DREGURB*HEATC**	-0.0749	(0.0593)
DBRXCAP*SURFB	-0.0508	(0.0388)	**DREGURB*FUELA**	-0.1570*	(0.0846)
DBRXCAP*SURFC	-0.0430	(0.0386)	**DREGURB*FUELB**	**-0.2191**	(0.0876)
DBRXCAP*SURFD	0.0005	(0.0392)	**DREGURB*FUELC**	**-0.3054**	(0.0985)
DBRXCAP*SURFE	0.0381	(0.0379)	**DREGURB*FUELD**	0.0059	(0.1288)
DBRXCAP*RECENTBUILT	-0.0182	(0.0207)	**DREGURB*FUELE**	-0.1278	(0.0823)
DBRXCAP*RENOVATION	**0.0440**	(0.0190)	**DREGURB*FUELF**	**-0.1716**	(0.0836)
DBRXCAP*FURNISH	0.0369	(0.0376)	**DREGURB*FUELG**	**-0.2084**	(0.0811)
DBRXCAP*HEATA	**-0.0736**	(0.0156)	**DREGURB*DGAZZLING**	-0.0177	(0.0140)
DBRXCAP*HEATB	-0.0412*	(0.0215)	**DREGURB*WALLISO**	-0.0003	(0.0148)
DBRXCAP*HEATC	0.0078	(0.0759)	**DREGURB*BATHROOM**	-0.0465	(0.0407)
DBRXCAP*FUELA	-0.0834	(0.1560)	**DREGURB*TOILET**	0.0462	(0.0367)
DBRXCAP*FUELB	0.0446	(0.1623)	**DREGURB*PARKA**	**-0.0511**	(0.0244)
DBRXCAP*FUELC	0.0440	(0.1615)	**DREGURB*PARKB**	-0.0275	(0.0232)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table 5 – concluded from previous page

Variable	Coeff.	S.E.	Variable	Coeff.	S.E.
DBRXCAP*FUELD	-0.0316	(0.1872)	**DREGURB*GARDEN**	0.0225	(0.0175)
DBRXCAP*FUELE	-0.1281	(0.1563)	**DREGURB*LNPM10**	**0.4884**	(0.1902)
DBRXCAP*FUELF	-0.0991	(0.1540)	**DREGURB*LNPOT_JOBS**	-0.0174	(0.0751)
DBRXCAP*FUELG	-0.0124	(0.1693)	**DREGURB*LNSLOPE (AC)**	**0.0496**	(0.0222)
DBRXCAP*DGAZZLING	0.0139	(0.0135)	**DREGURB*LNPERFOR (AC)**	**0.0310**	(0.0081)
DBRXCAP*WALLISO	**-0.0296**	(0.0137)	**DREGURB*LNPERAR (AC)**	**-0.0246**	(0.0117)
DBRXCAP*BATHROOM	**0.0737**	(0.0260)	**DREGURB*LNMEDINC (SS)**	0.0087	(0.0637)
DBRXCAP*TOILET	0.0563	(0.0439)			
DBRXCAP*PARKA	**0.0615**	(0.0218)			
DBRXCAP*PARKB	**0.0951**	(0.0213)			
DBRXCAP*GARDEN	**0.0365**	(0.0155)			
DBRXCAP*LNPM10	**-0.4131**	(0.1638)			
DBRXCAP*LNPOT_JOBS	**0.5230**	(0.1185)			
DBRXCAP*LNSLOPE (AC)	0.0478	(0.0431)			
DBRXCAP*LNPERFOR (AC)	0.0033	(0.0087)			
DBRXCAP*LNPERAR (AC)	0.0243	(0.0190)			
DBRXCAP*LNMEDINC (SS)	**0.1422**	(0.0680)			
scale(σ)			0.3360		
Nobs			70,839		
Sample size			330,147		

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Appendix F: Effects of the choice of the size of the basic spatial units on coefficients of IRM

Table 6: Interval regression: impact of the choice of the size of the basic spatial unit for the LNSLOPE, LNPERFOR and LNPERAR variables, Sample=Union.

Variable	**BSU**	**(1)**		**BSU**	**(2)**		**BSU**	**(3)**	
		Coeff.	**S.E**		**Coeff.**	**S.E**		**Coeff.**	**S.E**
LNSLOPE	SS	**0.0674**	(0.0049)	AC	**0.1325**	(0.0076)	COM	**0.1573**	(0.0073)
LNPERFOR	AC	**0.0106**	(0.0024)	COM	**0.0151**	(0.0029)			
LNPERAR	AC	**-0.0198**	(0.0027)	COM	**-0.0200**	(0.0028)			

Each row represents, 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.

Appendix G: Spatial econometric models and omitted variable bias

Brasington and Hite (2005) assert that the use of spatial regression methods allows to mitigate the omitted variable bias. Here is a quick summary of their argumentation. Unmeasured influences help determine the rent of neighboring dwellings. As a linear combination of dwelling rents nearby in space, the spatial lag term $W\tilde{y}$ picks up unobserved influences that affect dwelling rent. Moreover, the rent of a given dwelling is related to the rents of neighboring dwellings. Hence, any dwelling rent is affected by the unmeasured effects of neighboring observations. Therefore, the $W\tilde{y}$ term should capture the influence of omitted variables on the rent of a dwelling.

To formally establish this conjecture, Pace and LeSage (2010) derive an expression for OLS omitted variable bias in a univariate model with spatial dependence in the dependent variable, in the disturbances and in the explanatory variables. They show that this type of spatial dependence in the presence of omitted variables exacerbates the usual bias that arises when applying OLS to this type of sample data. To derive the omitted variable bias implied by the use of OLS, Pace and LeSage (2010) use the following spatial econometric specification:

$$
\begin{aligned}
y &= x\beta + \alpha W y + \varepsilon & (30)\\
\varepsilon &= \rho W \varepsilon + \xi & (31)\\
\xi &= x\gamma + u & (32)\\
x &= \phi W x + \nu & (33)\\
u &\sim \mathcal{N}\left(0, \sigma_u^2 I_N\right) & (34)\\
\nu &\sim \mathcal{N}\left(0, \sigma_\nu^2 I_N\right) & (35)
\end{aligned}
$$

In (30) to (33), y is a $n \times 1$ vector of observations on the dependent variable, x represents an $n \times 1$ vector of observations on a non-constant explanatory variable, ε, ξ, u and ν represent various types of $n \times 1$ disturbance vectors. α, β, ρ, γ and ϕ represent scalar parameters, and W is an $n \times n$ non-negative symmetric spatial weight matrix with zeros on the diagonal.

In (30) to (35), Pace and LeSage (2010) extend the conventional SAC specification by implying a spatial dependence in the explanatory variable x, which is governed by a spatial autoregressive process with dependence parameter ϕ. Using $F(\alpha) = (I_n - \alpha W)^{-1}$, $G(\rho) = (I_n - \rho W)^{-1}$, and $H(\phi) = (I_n - \phi W)^{-1}$ we can solve for y, ε and x as follows

$$
\begin{aligned}
y &= F(\alpha)\, x\beta + F(\alpha)\, \varepsilon & (36)\\
\varepsilon &= G(\rho)\,(x\gamma + u) & (37)\\
x &= H(\phi)\, \nu & (38)
\end{aligned}
$$

Taking (36), (37) and (38) together lead to the following DGP

$$
y = F(\alpha)\, H(\phi)\, \nu\beta + F(\alpha)\, G(\rho)\, H(\phi)\, \nu\gamma + F(\alpha)\, G(\rho)\, u \qquad (39)
$$

OLS estimates $\hat{\beta}_0 = (x'x)^{-1} x'y$ represent BLUE estimates when the DGP is represented by the ordinary regression model: $y = x\beta + \varepsilon$. However, if the true DGP is (39), then an omitted variable bias may arise as the expression of $\hat{\beta}_0$ in terms of the true parameter value β is

$$
\begin{aligned}
\hat{\beta}_0 = {} & \frac{\nu' H(\phi)\, F(\alpha)\, H(\phi)\, \nu}{\nu' H(\phi)^2\, \nu}\beta + \frac{\nu' H(\phi)\, F(\alpha)\, G(\rho)\, H(\phi)\, \nu}{\nu' H(\phi)^2\, \nu}\gamma \\
& + \frac{\nu' H(\phi)\, F(\alpha)\, G(\rho)\, u}{\nu' H(\phi)^2\, \nu} \qquad (40)
\end{aligned}
$$

The probability limit of $\hat{\beta}_0$ is

$$\begin{aligned}
\text{plim}_{n\to\infty}\hat{\beta}_0 &= \frac{tr\left[H(\phi)^2 F(\alpha)\right]}{tr\left[H(\phi)^2\right]}\beta + \frac{tr\left[H(\phi)^2 F(\alpha) G(\rho)\right]}{tr\left[H(\phi)^2\right]}\gamma \\
&= T_\beta(\phi,\alpha)\beta + T_\gamma(\phi,\alpha,\rho)\gamma \qquad (41)
\end{aligned}$$

with

$$\begin{aligned}
T_\beta(\phi,\alpha) &= \frac{tr\left[H(\phi)^2 F(\alpha)\right]}{tr\left[H(\phi)^2\right]} \\
T_\gamma(\phi,\alpha,\rho) &= \frac{tr\left[H(\phi)^2 F(\alpha) G(\rho)\right]}{tr\left[H(\phi)^2\right]}
\end{aligned}$$

As the factors $T_\beta(\phi,\alpha)$ and $T_\gamma(\phi,\alpha,\rho)$ rise above 1, the bias of using OLS to produce estimates for a model with a dependent variable y generated by the spatial DGP from (9) can increase. This is further verified if β and γ have the same signs.

Pace and LeSage (2010) show that when spatial dependance in the dependent variable or in the disturbance exists ($\alpha > 0$ or $\rho > 0$), $T_\gamma(\phi,\alpha,\rho) > 1$ and that $T_\beta(\phi,\alpha) > 1$ for $\alpha > 0$. They also show that spatial dependence in the explanatory variable $\phi > 0$ further accentuate these factors.

Even in the absence of omitted bias ($\gamma = 0 \Rightarrow \text{plim}_{n\to\infty}\hat{\beta}_0 = T_\beta(\phi,\alpha)\beta$), spatial dependence in the dependent and the explanatory variable would induce a bias in OLS estimate ($\alpha > 0,\ \phi > 0 \Rightarrow T_\beta(\phi,\alpha) > 1$). This bias is further compounded in case of omitted variable bias ($\gamma \neq 0$). In that specific case, spatial dependence in y or in the disturbance ε amplifies the omitted variable bias as $\alpha > 0$ or $\rho > 0 \Rightarrow T_\gamma(\phi,\alpha,\rho) > 1$.

Appendix H: Two-stage estimation with fixed effects: first stage

Table 7: Interval regression with fixed effects: estimation results for different study areas.

Variable	Union		ERUC		Urban Region		Agglo.	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Intercept	**5.9009**	0.1092	**5.9163**	0.1102	**5.6521**	0.1735	**6.1178**	0.0600
APP	**-0.0850**	0.0045	**-0.0869**	0.0048	**-0.0805**	0.0057	**-0.0713**	0.0061
OTHER	0.0121	0.0217	0.0025	0.0241	-0.0407	0.0315	**-0.0811**	0.0375
SING[1]								
LOFT	0.0057	0.0106	0.0179*	0.0099	**0.0556**	0.0112	**0.0528**	0.0119
LNROOMS	**0.0965**	0.0044	**0.1035**	0.0048	**0.1261**	0.0060	**0.1212**	0.0066
SURFA	**-0.4137**	0.0090	**-0.4344**	0.0096	**-0.5041**	0.0118	**-0.5284**	0.0133
SURFB	**-0.3656**	0.0077	**-0.3837**	0.0082	**-0.4507**	0.0103	**-0.4753**	0.0118
SURFC	**-0.3167**	0.0072	**-0.3309**	0.0076	**-0.3925**	0.0095	**-0.4143**	0.0111
SURFD	**-0.2307**	0.0071	**-0.2399**	0.0076	**-0.2874**	0.0094	**-0.2995**	0.0110
SURFE	**-0.1482**	0.0077	**-0.1527**	0.0107	**-0.1766**	0.0102	**-0.1791**	0.0119
SURFF[1]								
RECENTBUIL	**0.1752**	0.0049	**0.1755**	0.0053	**0.1952**	0.0066	**0.1992**	0.0076
RENOVATION	**0.0582**	0.0048	**0.0599**	0.0051	**0.0704**	0.0061	**0.0738**	0.0066
FURNISH	**-0.1912**	0.0070	**-0.1904**	0.0073	**-0.1884**	0.0088	**-0.1757**	0.0094
HEATA	**0.1815**	0.0039	**0.1755**	0.0041	**0.1711**	0.0050	**0.1640**	0.0053
HEATB	**0.1677**	0.0052	**0.1603**	0.0055	**0.1534**	0.0062	**0.1470**	0.0065
HEATC	**0.1509**	0.0149	**0.1434**	0.0154	**0.1352**	0.0173	**0.1486**	0.0187
HEATD[1]								
FUELA	0.0148	0.0352	0.0230	0.0375	0.0666	0.0513	0.0904*	0.0538
FUELB	**-0.1394**	0.0366	**-0.1333**	0.0392	**-0.1116**	0.0541	-0.0820	0.0569
FUELC	-0.0724*	0.0398	-0.0723*	0.0429	-0.0975	0.0598	-0.0401	0.0671
FUELD	0.0904	0.0602	0.1091*	0.0634	0.1418*	0.0818	0.1394	0.0879
FUELE	**0.0906**	0.0355	**0.0956**	0.0379	**0.1221**	0.0517	**0.1333**	0.0544
FUELF	0.0104	0.0351	0.0133	0.0374	0.0419	0.0512	0.0618	0.0536
FUELG	-0.0059	0.0373	0.0033	0.0397	0.0171	0.0542	0.0384	0.0576

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table7 – concluded from previous page

Variable	Union		ERUC		Urban Region		Agglo.	
	Coeff.	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**
FUELH[1]								
DGLAZING	**0.0281**	0.0033	**0.0254**	0.0035	**0.0214**	0.0040	**0.0193**	0.0043
WALLISO	**0.0448**	0.0036	**0.0433**	0.0038	**0.0369**	0.0044	**0.0331**	0.0047
BATHROOM	**0.2525**	0.0068	**0.2451**	0.0071	**0.2216**	0.0081	**0.2155**	0.0083
TOILET	**0.0404**	0.0082	**0.0417**	0.0085	**0.0412**	0.0096	**0.0368**	0.0100
PARKA	**-0.1688**	0.0053	**-0.1712**	0.0057	**-0.1900**	0.0072	**-0.1876**	0.0082
PARKB	**-0.0469**	0.0051	**-0.0438**	0.0055	**-0.0402**	0.0071	**-0.0256**	0.0084
PARKC[1]								
GARDEN	**0.0129**	0.0039	**0.0135**	0.0042	**0.0230**	0.0050	**0.0222**	0.0053
scale(σ)	0.3345		0.3349		0.3360		0.3320	
Nobs	72,105		63,681		44,939		38,053	
Sample size	330,096		301,111		233,562		208,355	

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

References

[1] Ahlfeldt, G. (2011) If Alonso was right: modeling accessibility and explaining the residential land gradient, *Journal of Regional Science* 51, 318–338.

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

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

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

[5] Blondel, V., Krings, G., Thomas, I. (2010) Régions et frontières de téléphonie mobile en Belgique et dans l'aire métropolitaine bruxelloise, *Brussels Studies* **42**.

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

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

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

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

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

[11] Cheshire, P.C. (2010) *Why Brussels needs a City-Region for the City*. Mimeo.

[12] Combes, P.P., Duranton, G., Gobillon, L. (2011) The identification of agglomeration economies, *Journal of Economic Geography*, 11(2), 253–266.

[13] Donnay, J.-P., Lambinon, M. (1997) Détermination des limites d'agglomération par télédetection: discussion méthodologique et application au cas de Huy (Belgique), *Actes*

des journées scientifiques du réseau de télédétection de l'UREF: Télédétetection des espaces urbains et périurbains, Liège, 1995, Montréal: AUPELF, pp. 239–246;

[14] Flemming, M.M. (2004) Techniques for Estimating Spatially Dependent Discrete Choice Models. In: Anselin, L., R.J.G.M. Florax, and S.J. Rey (Eds.) *Advances in Spatial Econometrics. Methodology, Tools and Applications*. Berlin Heidelberg : Springer–Verlag , pp. 145–166.

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

[16] Gawande, K., Jenkins-Smith, H. (2001) Nuclear Waste Transport and Residential Property Values: Estimating the Effects of Perceived Risks, *Journal of Environmental Economics and Management* (42), 207–233.

[17] Gehlke, C., Biehl, K. (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.

[18] Geoghegan, J., Wainger, L.A., Bockstael, N.E. (1997) Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS, *Ecological Economics* (23), 251–264.

[19] Goffette–Nagot, F., Reginster, I., Thomas, I. (2011) Spatial Analysis of Residential Land Prices in Belgium: Accessibility, Linguistic Border, and Environmental Amenities, *Regional Studies*, 45, 1253–1268.

[20] Harris, C.D. (1954) The Market as a Factor in the Localization of Industry in the United States, *Annals of the Association of American Geographers* 44, 315–348.

[21] Head, K., Mayer, T. (2004) Market Potential and the Location of Japanese Investment in the European Union, *The Review of Economics and Statistics*, 86(4), 959-972.

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

[23] LeSage, J.P., Pace, R.K. (2009) *Introduction to spatial econometrics*. Boca Raton: Chapman & Hall/CRC.

[24] Löchl, M., Axhausen, K.W. (2010) Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Journal of Transport and Land Use*, 3, 39–63.

[25] Magrini, S. (2004) Regional (di)convergence, in: Henderson, J. V., Thisse, J. F. (ed.), *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam, Elsevier, pp. 2741–2796.

[26] Moniteur Belge (2004), Lundi 21 Juin 2004 Troisième Edition 174ème année.

[27] Montero, J.M., Larraz, B., Chasco, C. (2008) Building an Environmental Quality Index for a big city: a spatial interpolation approach with DP2, MPRA Paper #10736.

[28] Pace, R.K., LeSage, J.P. (2010) Omitted Variable Biases of OLS and Spatial Lag Models. In: A. Pàez, J. Gallo, R.N. Buliung and S. Dall'erba (Eds.) *Progress in Spatial Analysis Methods and Applications*. Berlin Heidelberg : Springer–Verlag , pp. 17–28.

[29] Rozenfeld, H.D., Rybski, D. , Gabaix, X., Makse, H.A. (2011) The Area and Population of Cities: New Insights from a Different Perspective on Cities, *American Economic Review* 101, 2205–2225.

[30] Tannier, C., Thomas, I. , Vuidel, G., Frankhauser, P. (2011) A Fractal Approach to Identifying Urban Boundaries, *Geographical Analysis* 43 (2), 211–227.

[31] Thomas, I., Cotteels, C., Jones, J., Peeters, D. (2012), Revisiting the extension of the Brussels urban agglomeration: new methods, new data new results?, *Belgeo* [Online] **1-2**. http://belgeo.revues.org/6074

[32] Van Hecke, E., Halleux, J.–M., Decroly, J.–M., Mérenne–Schoumaker, B. (2009), *Noyaux d'habitat et Régions urbaines dans une Belgique urbanisées, Enquête Socio–Economique 2001*, Monographies, Direction générale Statistique et Information économique.

[33] Vandenbulcke, G. (2007) *Accessibility indicators to places and transports.* Service Public Fédéral Mobilité et Transports. Mobilité et Sécurité Routière. Politique Scientifique Générale.

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

[35] Vandermotten, C., Vermoesen, F. , De Lannoy, W., De Corte, S. (1999), Villes d'Europe. Cartographie comparative, *Bulletin du Crédit Communal*, 207–208.

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

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

Table 8: Interval regression: estimation results for different study areas.

Variable	Union		ERUC		Urban Region		Agglo.		BCR	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E
Intercept	**-2.7275**	0.5479	**-2.4875**	0.6027	**-2.1854**	0.5961	**-1.9867**	0.7928	**-4.1692**	0.9286
APP[1]										
OTHER	**0.1011**	0.0414	**0.0937**	0.0435	0.0297	0.0503	-0.0202	0.0655	0.0099	0.0706
SING	**0.0738**	0.0054	**0.0758**	0.0058	**0.0768**	0.0072	**0.0722**	0.0082	**0.0510**	0.0090
LOFT	0.0186*	0.0106	**0.0310**	0.0106	**0.0645**	0.0111	**0.0576**	0.0117	**0.0570**	0.0120
LNROOMS	**0.0964**	0.0052	**0.1033**	0.0055	**0.1252**	0.0072	**0.1181**	0.0084	**0.1203**	0.0091
SURFA[1]										
SURFB	**0.0463**	0.0056	**0.0488**	0.0061	**0.0502**	0.0067	**0.0497**	0.0071	**0.0497**	0.0077
SURFC	**0.0966**	0.0060	**0.1028**	0.0065	**0.1062**	0.0071	**0.1073**	0.0075	**0.1084**	0.0080
SURFD	**0.1833**	0.0077	**0.1943**	0.0082	**0.2117**	0.0098	**0.2241**	0.0099	**0.2337**	0.0109
SURFE	**0.2706**	0.0101	**0.2856**	0.0107	**0.3280**	0.0130	**0.3480**	0.0133	**0.3694**	0.0162
SURFF	**0.4357**	0.0148	**0.4543**	0.0158	**0.5160**	0.0192	**0.5412**	0.0221	**0.5445**	0.0280
RECENTBUIL	**0.1739**	0.0064	**0.1743**	0.0070	**0.1997**	0.0081	**0.2030**	0.0094	**0.1894**	0.0131
RENOVATION	**0.0648**	0.0056	**0.0658**	0.0058	**0.0827**	0.0074	**0.0904**	0.0082	**0.0977**	0.0082
FURNISH	**-0.1896**	0.0094	**-0.1876**	0.0102	**-0.1817**	0.0118	**-0.1661**	0.0119	**-0.1594**	0.0128
HEATA[1]										
HEATB	**-0.0121**	0.0055	**-0.0139**	0.0057	**-0.0189**	0.0064	**-0.0203**	0.0068	-0.0131*	0.0073
HEATC	-0.0021	0.0187	-0.0029	0.0192	-0.0128	0.0210	-0.0067	0.0236	0.0071	0.0242
HEATD	**-0.1923**	0.0042	**-0.1874**	0.0045	**-0.1705**	0.0052	**-0.1582**	0.0053	**-0.1432**	0.0060
FUELA[1]										
FUELB	**-0.1652**	0.0127	**-0.1698**	0.0136	**-0.1782**	0.0226	**-0.1684**	0.0236	**-0.1275**	0.0259

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table8 – concluded from previous page

Variable	Union		ERUC		Urban Region		Agglo.		BCR	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E
FUELC	**-0.0880**	0.0238	**-0.0930**	0.0246	**-0.1597**	0.0386	**-0.1171**	0.0478	-0.0410	0.0693
FUELD	0.1070*	0.0647	**0.1257**	0.0623	0.0947	0.0889	0.0504	0.0902	0.0591	0.0970
FUELE	**0.0847**	0.0078	**0.0800**	0.0088	**0.0564**	0.0112	**0.0358**	0.0122	**0.0287**	0.0144
FUELF	0.0076	0.0048	0.0045	0.0049	**-0.0184**	0.0056	**-0.0282**	0.0062	**-0.0302**	0.0067
FUELG	-0.0281*	0.0148	-0.0270	0.0167	**-0.0565**	0.0219	**-0.0510**	0.0236	-0.0343	0.0262
FUELH	0.0126	0.0390	0.0153	0.0420	-0.0572	0.0517	-0.0905	0.0568	-0.0845	0.0583
DGLAZING	**0.0296**	0.0036	**0.0267**	0.0037	**0.0198**	0.0044	**0.0185**	0.0047	**0.0213**	0.0050
WALLISO	**0.0408**	0.0036	**0.0403**	0.0038	**0.0348**	0.0043	**0.0316**	0.0045	**0.0266**	0.0048
BATHROOM	**0.2525**	0.0080	**0.2444**	0.0083	**0.2171**	0.0089	**0.2094**	0.0085	**0.1999**	0.0082
TOILET	**0.0401**	0.0083	**0.0412**	0.0087	**0.0401**	0.0095	**0.0338**	0.0099	**0.0373**	0.0104
PARKA[1]										
PARKB	**0.1039**	0.0059	**0.1078**	0.0065	**0.1321**	0.0059	**0.1410**	0.0064	**0.1515**	0.0072
PARKC	**0.1513**	0.0080	**0.1520**	0.0094	**0.1763**	0.0112	**0.1738**	0.0121	**0.1428**	0.0140
GARDEN	0.0072	0.0052	0.0080	0.0055	**0.0149**	0.0053	**0.0135**	0.0061	**0.0192**	0.0063
LNPM10	**-0.1554**	0.0421	**-0.2209**	0.0510	-0.0483	0.0539	-0.0757	0.0603	**-0.1885**	0.0711
LNPOT_JOBS	**0.3548**	0.0268	**0.3483**	0.0292	**0.23433**	0.0379	**0.1917**	0.0541	**0.3663**	0.0658
LNSLOPE (AC)	**0.1325**	0.0076	**0.1193**	0.0104	**0.1362**	0.0136	**0.1156**	0.0174	**0.1497**	0.0268
LNPERFOR (AC)	**0.0106**	0.0024	**0.0103**	0.0025	**0.0143**	0.0031	**0.0106**	0.0032	0.0030	0.0049
LNPERAR (AC)	**-0.0198**	0.0027	**-0.0236**	0.0029	**-0.0277**	0.0033	**-0.0301**	0.0037	**-0.0295**	0.0050
LNMEDINC (SS)	**0.4068**	0.0285	**0.4152**	0.0294	**0.4694**	0.0259	**0.5199**	0.0248	**0.5499**	0.0278
scale(σ)	0.3428		0.3432		0.3396		0.3342		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	

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Table 9: Interval regression: impact of the choice of the size of the basic spatial unit for the variable LNMEDINC, Sample=Union.

Variable	**(1)**		**(2)**	
	Coeff.	**S.E**	**Coeff.**	**S.E**
Intercept	**-2.7275**	(0.5479)	**-6.1091**	(0.5307)
APP[1]				
OTHER	**0.1010**	(0.0414)	**0.0919**	(0.0396)
SING	**0.0738**	(0.0054)	**0.0790**	(0.0055)
LOFT	0.0186*	(0.0106)	0.0169	(0.0106)
LNROOMS	**0.0964**	(0.0052)	**0.0961**	(0.0052)
SURFA[1]				
SURFB	**0.0462**	(0.0056)	**0.0479**	(0.0057)
SURFC	**0.0966**	(0.0059)	**0.1005**	(0.0062)
SURFD	**0.1833**	(0.0077)	**0.1878**	(0.0080)
SURFE	**0.2706**	(0.0101)	**0.2738**	(0.0103)
SURFF	**0.4356**	(0.0148)	**0.4402**	(0.0151)
RECENTBUILT	**0.1738**	(0.0064)	**0.1772**	(0.0065)
RENOVATION	**0.0648**	(0.0055)	**0.0570**	(0.0054)
FURNISH	**-0.1896**	(0.0094)	**-0.1928**	(0.0094)
HEATA[1]				
HEATB	**-0.0121**	(0.0055)	-0.0080	(0.0058)
HEATC	-0.0021	(0.0187)	-0.0194	(0.0171)
HEATD	**-0.1923**	(0.0042)	**-0.1956**	(0.0041)
FUELA[1]				
FUELB	**-0.1651**	(0.0127)	**-0.1649**	(0.0126)
FUELC	**-0.0880**	(0.0237)	**-0.0863**	(0.0233)
FUELD	0.1070*	(0.0647)	0.0926	(0.0613)
FUELE	**0.0847**	(0.0078)	**0.0800**	(0.0079)
FUELF	0.0076	(0.0048)	0.0018	(0.0046)
FUELG	-0.0280*	(0.0148)	-0.0253*	(0.0144)
FUELH	0.0126	(0.0390)	-0.0051	(0.0366)
DGLAZING	**0.0296**	(0.0036)	**0.0298**	(0.0037)
WALLISO	**0.0408**	(0.0035)	**0.0436**	(0.0036)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table9 – concluded from previous page

Variable	(1)		(2)	
	Coeff.	S.E	Coeff.	S.E
BATHROOM	**0.2524**	(0.0080)	**0.2582**	(0.0077)
TOILET	**0.0401**	(0.0083)	**0.0380**	(0.0082)
PARKA[1]				
PARKB	**0.1039**	(0.0059)	**0.1146**	(0.0064)
PARKC	**0.1513**	(0.0080)	**0.1642**	(0.0084)
GARDEN	0.0072	(0.0051)	**0.0167**	(0.0052)
LNPM10	**-0.1554**	(0.0421)	0.0584	(0.0447)
LNPOT_JOBS	**0.3547**	(0.0268)	**0.3345**	(0.0237)
LNSLOPE (AC)	**0.1325**	(0.0076)	**0.1298**	(0.0076)
LNPERFOR (AC)	**0.0106**	(0.0024)	**0.0098**	(0.0025)
LNPERAR (AC)	**-0.0198**	(0.0027)	**-0.0311**	(0.0028)
LNMEDINC (SS)	**0.4068**	(0.0285)		
LNMEDINC (COM)			**0.7027**	(0.0376)
scale(σ)	0.3428		0.3442	
Nobs	70,839		72,105	
Sample size	330,147		330,147	

(1): Estimation with **LNMEDINC (SS)**
(log of median income measured at the Statistical Sector level)
(2): Estimation with **LNMEDINC (COM)**
(log of median income measured at the Municipality level)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Table 10: Interval regression: impact of the choice of the size of the basic spatial unit for the variable LNAVINC, Sample=Union.

Variable	**(1)**		**(2)**		**(3)**	
	Coeff.	**S.E**	**Coeff.**	**S.E**	**Coeff.**	**S.E**
Intercept	**-2.2814**	(0.5288)	**-3.4394**	(0.4774)	**-5.3756**	(0.4568)
APP[1]						
OTHER	**0.1222**	(0.0509)	**0.0857**	(0.0421)	**0.0910**	(0.0398)
SING	**0.0699**	(0.0055)	**0.0801**	(0.0055)	**0.0776**	(0.0055)
LOFT	0.0201*	(0.0112)	0.0156	(0.0104)	0.0149	(0.0105)
LNROOMS	**0.0940**	(0.0054)	**0.0944**	(0.0051)	**0.0967**	(0.0051)
SURFA[1]						
SURFB	**0.0469**	(0.0059)	**0.0474**	(0.0056)	**0.0481**	(0.0057)
SURFC	**0.0965**	(0.0060)	**0.0992**	(0.0060)	**0.0997**	(0.0062)
SURFD	**0.1881**	(0.0077)	**0.1865**	(0.0078)	**0.1855**	(0.0080)
SURFE	**0.2780**	(0.0103)	**0.2748**	(0.0097)	**0.2700**	(0.0101)
SURFF	**0.4272**	(0.0154)	**0.4380**	(0.0145)	**0.4296**	(0.0144)
RECENTBUIL	**0.1762**	(0.0069)	**0.1775**	(0.0064)	**0.1768**	(0.0064)
RENOVATION	**0.0673**	(0.0055)	**0.0583**	(0.0055)	**0.0565**	(0.0054)
FURNISH	**-0.1869**	(0.0103)	**-0.1895**	(0.0094)	**-0.1907**	(0.0093)
HEATA[1]						
HEATB	**-0.0129**	(0.0056)	-0.0079	(0.0056)	-0.0081	(0.0057)
HEATC	0.0001	(0.0189)	-0.0155	(0.0173)	-0.0215	(0.0170)
HEATD	**-0.1799**	(0.0044)	**-0.1921**	(0.0041)	**-0.1912**	(0.0039)
FUELA[1]						
FUELB	**-0.1521**	(0.0142)	**-0.1622**	(0.0128)	**-0.1612**	(0.0127)
FUELC	**-0.0735**	(0.0269)	**-0.0873**	(0.0256)	**-0.0898**	(0.0234)
FUELD	0.0661	(0.0697)	0.0932	(0.0620)	0.0912	(0.0612)
FUELE	**0.0780**	(0.0085)	**0.0828**	(0.0079)	**0.0808**	(0.0079)
FUELF	0.0077	(0.0051)	0.0062	(0.0047)	0.0050	(0.0045)
FUELG	-0.0171	(0.0157)	-0.0260*	(0.0145)	-0.0257*	(0.0142)
FUELH	0.0259	(0.0384)	0.0119	(0.0372)	-0.0009	(0.0366)
DGLAZING	**0.0272**	(0.0037)	**0.0283**	(0.0036)	**0.0288**	(0.0036)
WALLISO	**0.0380**	(0.0036)	**0.0434**	(0.0035)	**0.0439**	(0.0036)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table10 – concluded from previous page

Variable	(1)		(2)		(3)	
	Coeff.	**S.E**	**Coeff.**	**S.E**	**Coeff.**	**S.E**
BATHROOM	**0.2451**	(0.0086)	**0.2559**	(0.0078)	**0.2573**	(0.0076)
TOILET	**0.0391**	(0.0084)	**0.0376**	(0.0082)	**0.0387**	(0.0082)
PARKA[1]						
PARKB	**0.0999**	(0.0063)	**0.1111**	(0.0063)	**0.1136**	(0.0062)
PARKC	**0.1374**	(0.0081)	**0.1564**	(0.0081)	**0.1608**	(0.0080)
GARDEN	0.0006	(0.0054)	0.0090*	(0.0051)	**0.0133**	(0.0052)
LNPM10	-0.0765*	(0.0464)	0.0146	(0.0416)	**0.1930**	(0.0436)
LNPOT_JOBS	**0.2923**	(0.0277)	**0.2765**	(0.0273)	**0.2658**	(0.0220)
LNSLOPE (AC)	**0.1128**	(0.0075)	**0.1047**	(0.0071)	**0.0960**	(0.0071)
LNPERFOR (AC)	**0.0082**	(0.0023)	0.0039	(0.0026)	**0.0053**	(0.0025)
LNPERAR (AC)	**-0.0220**	(0.0027)	**-0.0307**	(0.0027)	**-0.0341**	(0.0028)
LNAVINC (SS)	**0.4088**	(0.0208)				
LNAVINC (AC)			**0.5121**	(0.0245)		
LNAVINC (COM)					**0.6564**	(0.0272)
scale(σ)	0.3342		0.3419		0.3418	
Nobs	63,426		70,828		72,105	
Sample size			330,147			

(1): Estimation with **LNAVINC (SS)**
(log of average income measured at the Statistical Sector level)
(2): Estimation with **LNAVINC (AC)**
(log of average income measured at the Former Township level)
(3): Estimation with **LNAVINC (COM)**
(log of average income measured at the Municipality level)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Table 11: SAR Interval regression: estimation results for different samples.

Variable	**Sample1**				**Sample2**			
	IR	**SAR**	**SDM**	**SAC**	**IR**	**SAR**	**SDM**	**SAC**
ρ		**0.1441**	**0.2766**	0.2000*		**0.1788**	0.1031*	0.1862*
		(0.0850)	(0.1356)	(0.1431)		(0.0910)	(0.0755)	(0.1331)
λ				-0.0443				-0.0408
				(0.2783)				(0.2700)
Intercept	16.0991	8.3069	9.9073	10.2833	-6.8674	-7.9728	-11.4137	-1.1954
	(13.2828)	(13.6757)	(16.0114)	(18.3413)	(7.2843)	(7.3068)	(12.6156)	(9.7381)
SING	**0.0725**	**0.0734**	**0.0744**	**0.0755**	**0.1236**	**0.1201**	**0.1217**	**0.1294**
	(0.0200)	(0.0206)	(0.0201)	(0.0211)	(0.0252)	(0.0260)	(0.0258)	(0.0275)
OTHER	**0.5719**	**0.5715**	**0.5776**	**0.6092**	**-0.5293**	**-0.5560**	**-0.5704**	**-0.5719**
	(0.1580)	(0.1558)	(0.1648)	(0.1701)	(0.2564)	(0.2645)	(0.2679)	(0.2778)
APP[1]								
LOFT	**0.1853**	**0.1837**	**0.1815**	**0.1773**	**0.1177**	**0.1165**	**0.1166**	**0.1190**
	(0.0400)	(0.0394)	(0.0416)	(0.0425)	(0.0529)	(0.0538)	(0.0536)	(0.0548)
LNROOMS	**0.2307**	**0.2297**	**0.2301**	**0.2283**	**0.2959**	**0.2975**	**0.2955**	**0.2995**
	(0.0206)	(0.0209)	(0.0212)	(0.0212)	(0.0265)	(0.0267)	(0.0270)	(0.0277)
RECENTBUILT	**0.2455**	**0.2439**	**0.2404**	**0.2429**	**0.2449**	**0.2461**	**0.2462**	**0.2497**
	(0.0224)	(0.0231)	(0.0228)	(0.0234)	(0.0315)	(0.0323)	(0.0320)	(0.0337)
RENOVATION	**0.0707**	**0.0722**	**0.0729**	**0.0668**	**0.0934**	**0.0971**	**0.0996**	**0.1019**
	(0.0253)	(0.0255)	(0.0249)	(0.0262)	(0.0304)	(0.0304)	(0.0313)	(0.0320)
FURNISH	**-0.1317**	**-0.1328**	**-0.1315**	**-0.1438**	**-0.1199**	**-0.1210**	**-0.1235**	**-0.1263**
	(0.0331)	(0.0341)	(0.0334)	(0.0360)	(0.0384)	(0.0395)	(0.0396)	(0.0409)
QUALITY2	**0.1478**	**0.1489**	**0.1491**	**0.1497**	**0.2526**	**0.2533**	**0.2525**	**0.2463**
	(0.0215)	(0.0224)	(0.0223)	(0.0232)	(0.0382)	(0.0391)	(0.0382)	(0.0389)
QUALITY3	**0.2226**	**0.2234**	**0.2231**	**0.2235**	**0.2116**	**0.2103**	**0.2092**	**0.1983**
	(0.0222)	(0.0232)	(0.0229)	(0.0239)	(0.0374)	(0.0382)	(0.0371)	(0.0385)
QUALITY4	**0.3096**	**0.3087**	**0.3094**	**0.3063**	**0.3307**	**0.3295**	**0.3268**	**0.3176**
	(0.0266)	(0.0273)	(0.0273)	(0.0283)	(0.0403)	(0.0411)	(0.0397)	(0.0414)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table 11 – continued from previous page

Variable	Sample1				Sample2			
	IR	SAR	SDM	SAC	IR	SAR	SDM	SAC
QUALITY5	**0.3942**	**0.3943**	**0.3931**	**0.3921**	**0.5131**	**0.5146**	**0.5102**	**0.5194**
	(0.0387)	(0.0406)	(0.0394)	(0.0407)	(0.0464)	(0.0480)	(0.0461)	(0.0484)
QUALITY1[1]								
FUELA	0.0150	0.0157	0.0159	0.0148	0.0108	0.0100	0.0119	0.0071
	(0.0333)	(0.0333)	(0.0331)	(0.0342)	(0.0420)	(0.0426)	(0.0439)	(0.0440)
FUELB	**-0.2653**	**-0.2656**	**-0.2692**	**-0.2746**	**-0.8951**	**-0.9456**	**-0.9576**	**-0.7536**
	(0.0775)	(0.0785)	(0.0768)	(0.0830)	(0.2524)	(0.2492)	(0.2587)	(0.2232)
FUELC	-1.5778	**-2.6890**	**-2.9592**	**-5.1116**	-0.2354	-0.2865	-0.3196	-0.5139*
	(881.8123)	(1.5379)	(1.5993)	(0.4398)	(0.2965)	(0.3058)	(0.3076)	(0.3830)
FUELH	-0.0394	-0.0399*	-0.0388	-0.0419*	-0.0752*	**-0.0738**	**-0.0706**	**-0.0752**
	(0.0308)	(0.0306)	(0.0308)	(0.0320)	(0.0408)	(0.0416)	(0.0428)	(0.0429)
FUELE[1]								
WALLISO	**0.0340**	**0.0331**	**0.0332**	**0.0342**	**0.0731**	**0.0703**	**0.0729**	**0.0711**
	(0.0146)	(0.0145)	(0.0152)	(0.0154)	(0.0185)	(0.0185)	(0.0185)	(0.0196)
PARKB	**0.1267**	**0.1252**	**0.1258**	**0.1194**	**0.2624**	**0.2610**	**0.2608**	**0.2589**
	(0.0168)	(0.0173)	(0.0169)	(0.0180)	(0.0188)	(0.0190)	(0.0192)	(0.0200)
PARKC	**0.1225**	**0.1223**	**0.1209**	**0.1147**	**0.2705**	**0.2710**	**0.2716**	**0.2791**
	(0.0246)	(0.0255)	(0.0240)	(0.0260)	(0.0321)	(0.0328)	(0.0318)	(0.0344)
PARKA[1]								
GARDEN	-0.0055	-0.0069	-0.0047	-0.0022	0.0107	0.0097	0.0094	0.0110
	(0.0166)	(0.0167)	(0.0168)	(0.0173)	(0.0208)	(0.0209)	(0.0208)	(0.0221)
LNPM10	**1.4044**	0.8260	2.1470	0.9098	4.7353*	2.0584	0.1302	2.6468
	(0.6670)	(0.7394)	(2.3173)	(1.1399)	(2.4996)	(2.8917)	(5.1493)	(3.4475)
LNPOT_JOBS (AC)	-1.4354	-0.7054	-2.1757	-0.9074	-0.7014	0.0607	-0.4415	-0.6514
	(1.1893)	(1.2322)	(3.8113)	(1.7109)	(1.1848)	(1.2560)	(2.5328)	(1.5762)
LNSLOPE (AC)	0.0113	0.0032	-0.0080	0.0515	-0.0291	-0.0228	-0.0221	-0.0579
	(0.0310)	(0.0307)	(0.0394)	(0.0483)	(0.0448)	(0.0449)	(0.0499)	(0.0519)
LNPERFOR (AC)					**0.3229**	0.1433	-0.0097	0.1808
					(0.1633)	(0.1904)	(0.3269)	(0.2274)

[1] Reference case.

Bold: significant at 0.05 level.

* significant at 0.10 level.

Continued on next page

Table 11 – concluded from previous page

Variable	Sample1				Sample2			
	IR	SAR	SDM	SAC	IR	SAR	SDM	SAC
LNPERAR (AC)	-0.0830	-0.0396	-0.1853	-0.0665				
	(0.0878)	(0.0889)	(0.2558)	(0.1195)				
LNMEDINC (SS)	**0.2831**	**0.2418**	**0.2938**	**0.2374**	**0.4685**	**0.4321**	**0.3684**	**0.4582**
	(0.0697)	(0.0747)	(0.0920)	(0.0999)	(0.0817)	(0.0848)	(0.1123)	(0.0914)
W*LNPM10			-1.7312				3.5457	
			(2.5462)				(7.2419)	
W*LNPOT_JOBS (AC)			1.5390				0.3131	
			(4.1943)				(3.6198)	
W*LNSLOPE (AC)			-0.0133				0.0540	
			(0.0897)				(0.1170)	
W*LNPERFOR (AC)							0.2475	
							(0.4690)	
W*LNPERAR (AC)			0.1631					
			(0.2862)					
W*LNMEDINC (SS)			-0.2376				0.2163	
			(0.2059)				(0.2843)	
scale(σ)	0.2827	0.2949	0.2852	0.2986	0.3664	0.3747	0.3682	0.3911
Nobs	2,969				2,565			

[1] Reference case.

Bold: significant at 0.05 level.

[*] significant at 0.10 level.

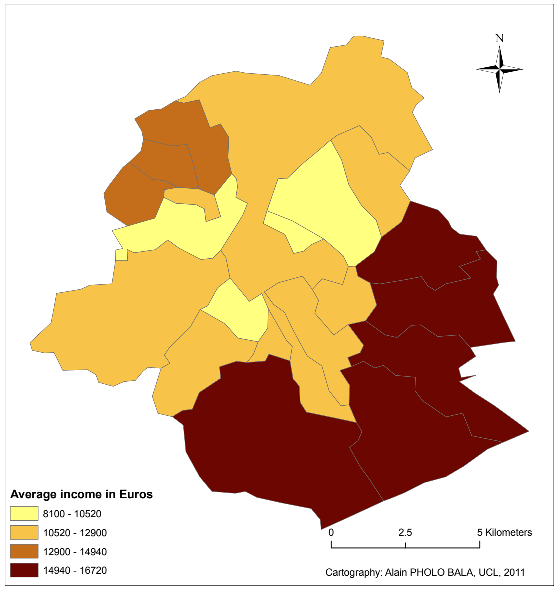

Figure 2: Average income in municipalities of Brussels Capital Region

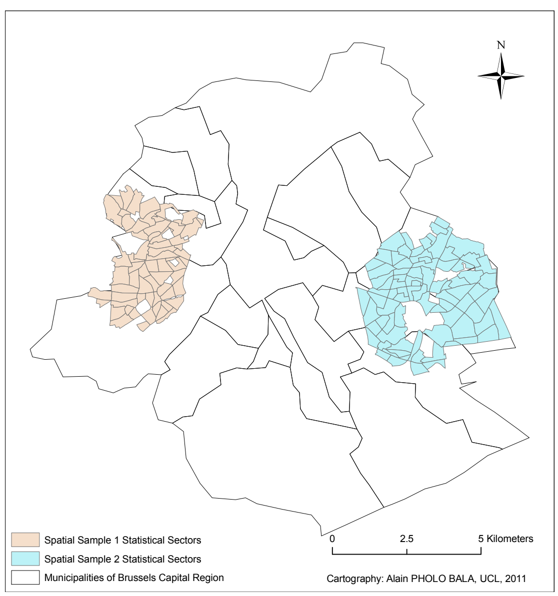

Figure 3: Location of Sample 1 and Sample 2 in Brussels Capital Region

Table 12: Two-stage estimation procedure with fixed effects: second step. Impact of the delineation of the study area.

Variable	**Agglomeration**			**Urban Region**			**ERUC**			**Union**		
	OLS	**SAR**	**SAC**	**OLS**	**SAR**	**SAC**	**OLS**	**SAR**	**SAC**	**OLS**	**SAR**	**SAC**
Intercept	4.5781*	0.5991	0.4843	**4.6113**	2.1905	0.1547	**4.1833**	**2.2157**	**1.4313**	**3.2357**	**1.4871**	0.8379*
	(2.3572)	(2.0517)	(2.2287)	(1.1623)	(1.3645)	(1.2352)	(0.6922)	(0.7154)	(0.6331)	(0.5856)	(0.5989)	(0.4891)
LNMEDINC (SS)	**0.7113**	**0.5163**	**0.5190**	**0.5011**	**0.4377**	**0.4286**	**0.4610**	**0.4152**	**0.3987**	**0.4599**	**0.4058**	**0.3897**
	(0.1156)	(0.0947)	(0.0983)	(0.0742)	(0.0699)	(0.0650)	(0.0474)	(0.0254)	(0.0128)	(0.0442)	(0.0119)	(0.0136)
LNPERAR (AC)	0.0076	0.0015	0.0008	**0.0250**	**0.0184**	0.0091	**0.0204**	**0.0144**	**0.0111**	**0.0245**	**0.0180**	**0.0148**
	(0.0103)	(0.0083)	(0.0084)	(0.0085)	(0.0083)	(0.0074)	(0.0069)	(0.0052)	(0.0052)	(0.0062)	(0.0050)	(0.0049)
LNPERFOR (AC)	0.0017	-0.0060	-0.0055	0.0088	0.0038	0.0027	0.0037	0.0019	0.0014	0.0051	0.0033	0.0029
	(0.0102)	(0.0080)	(0.0079)	(0.0067)	(0.0063)	(0.0050)	(0.0045)	(0.0042)	(0.0039)	(0.0040)	(0.0038)	(0.0035)
LNSLOPE (AC)	0.0260	0.0496*	0.0485*	0.0255	0.0276	0.0235	0.0223*	0.0199*	0.0186*	**0.0200**	0.0156	0.0138
	(0.0375)	(0.0297)	(0.0293)	(0.0219)	(0.0205)	(0.0157)	(0.0124)	(0.0119)	(0.0104)	(0.0101)	(0.0096)	(0.0085)
LNPOT_JOBS	**-0.3764**	-0.2289*	-0.2300*	**-0.2402**	**-0.1768**	**-0.1276**	**-0.0912**	**-0.0861**	**-0.0819**	-0.0522	**-0.0537**	**-0.0526**
	(0.1439)	(0.1172)	(0.1197)	(0.0640)	(0.0638)	(0.0240)	(0.0373)	(0.0240)	(0.0260)	(0.0336)	(0.0251)	(0.0254)
LNPM10	-0.1858	-0.0243	-0.0121	-0.2384*	-0.1140	-0.0525	**-0.4523**	**-0.2850**	**-0.2301**	**-0.3165**	**-0.1865**	**-0.1466**
	(0.1741)	(0.1437)	(0.1475)	(0.1400)	(0.1353)	(0.1067)	(0.0865)	(0.0835)	(0.0775)	(0.0703)	(0.0696)	(0.0617)
ρ		**0.5570**	**0.5670**		**0.3160**	**0.5430**		**0.2940**	**0.4110**		**0.3060**	**0.4150**
		(0.0971)	(0.1523)		(0.0886)	(0.1166)		(0.0417)	(0.0470)		(0.0412)	(0.0442)
λ			-0.0630			**-0.4480**			**-0.2070**			**-0.1699**
			(0.2832)			(0.2020)			(0.0341)			(0.0313)
scale(σ)	0.0866	0.0686	0.0686	0.0938	0.0877	0.0831	0.0985	0.0975	0.0927	0.0964	0.0922	0.0911
(Pseudo)R^2	0.7628	0.7871	0.8368	0.6533	0.6679	0.7192	0.3940	0.4046	0.4505	0.3621	0.3735	0.4243
(Pseudo)$\bar{R}^2$	0.7406	0.7671	0.8215	0.6404	0.6555	0.7087	0.3871	0.3978	0.4442	0.3558	0.3673	0.4186
Nobs		71			168			532			613	

Table 13: Two-stage estimation procedure with fixed effects, second step: impact of the choice of the BSU, Sample=ERUC

Variable	**(1)**			**(2)**		
	OLS	**SAR**	**SAC**	**OLS**	**SAR**	**SAC**
Intercept	**4.1833**	**2.2157**	**1.4163**	**2.2801**	1.2757*	1.3162*
	(0.6922)	(0.7154)	(0.8037)	(0.9028)	(0.7074)	(1.0159)
LNMEDINC (SS)	**0.4610**	**0.4152**	**0.3946**			
	(0.0474)	(0.0254)	(0.0468)			
LNMEDINC (COM)				**0.6136**	**0.5033**	**0.5020**
				(0.0760)	(0.0035)	(0.0867)
LNPERAR (AC)	**0.0204**	**0.0144**	**0.0109**	**0.0224**	**0.0191**	**0.0191**
	(0.0069)	(0.0052)	(0.0063)	(0.0071)	(0.0064)	(0.0071)
LNPERFOR (AC)	0.0037	0.0019	0.0013	0.0015	0.0006	0.0007
	(0.0045)	(0.0042)	(0.0040)	(0.0046)	(0.0044)	(0.0044)
LNSLOPE (AC)	0.0223*	0.0199*	**0.0182**	**0.0306**	**0.0275**	**0.0274**
	(0.0124)	(0.0119)	(0.0106)	(0.0127)	(0.0124)	(0.0125)
LNPOT_JOBS	**-0.0912**	**-0.0861**	**-0.0818**	**-0.0996**	**-0.0814**	**-0.0808**
	(0.0373)	(0.0240)	(0.0318)	(0.0396)	(0.0335)	(0.0396)
LNPM10	**-0.4523**	**-0.2850**	**-0.2269**	**-0.3092**	**-0.2104**	**-0.2122**
	(0.0865)	(0.0835)	(0.0840)	(0.0958)	(0.0899)	(0.0976)
ρ		**0.2940**	**0.4182**		**0.2510**	**0.2463**
		(0.0417)	(0.0933)		(0.0403)	(0.1275)
λ			-0.2209*			-0.0299
			(0.1390)			(0.1553)
scale(σ)	0.0985	0.0943	0.0087	0.1005	0.0980	0.0097
(Pseudo)R^2	0.3940	0.4048	0.4520	0.3639	0.3683	0.3907
(Pseudo)$\bar{R}^2$	0.3871	0.3980	0.4457	0.3566	0.3611	0.3838
Nobs			532			

(1): Estimation with **LNMEDINC (SS)**
(log of median income measured at the Statistical Sector level)
(2): Estimation with **LNMEDINC (COM)**
(log of median income measured at the Municipality level)

Recent titles

CORE Discussion Papers

2012/59 Tanguy ISAAC and Paolo PIACQUADIO. Equity and efficiency in an overlapping generation model.
2012/60 Luc BAUWENS, Giuseppe STORTI and Francesco VIOLANTE. Dynamic conditional correlation models for realized covariance matrices.
2012/61 Mikhail ESKAKOV and Alexey ISKAKOV. Equilibrium in secure strategies.
2012/62 Francis BLOCH and Axel GAUTIER. Strategic bypass deterrence.
2012/63 Olivier DURAND-LASSERVE, Axel PIERRU and Yves SMEERS. Sensitivity of policy simulation to benchmark scenarios in CGE models: illustration with carbon leakage.
2013/1 Pierre PESTIEAU and Maria RACIONERO. Harsh occupations, health status and social security.
2013/2 Thierry BRECHET and Henry TULKENS. Climate policies: a burden or a gain?
2013/3 Per J. AGRELL, Mehdi FARSI, Massimo FILIPPINI and Martin KOLLER. Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems.
2013/4 Adel HATAMI-MARBINI, Per J. AGRELL and Nazila AGHAYI. Imprecise data envelopment analysis for the two-stage process.
2013/5 Farhad HOSSEINZADEH LOTFI, Adel HATAMI-MARBINI, Per J. AGRELL, Kobra GHOLAMI and Zahra GHELEJ BEIGI. Centralized resource reduction and target setting under DEA control.
2013/6 Per J. AGRELL and Peter BOGETOFT. A three-stage supply chain investment model under asymmetric information.
2013/7 Per J. AGRELL and Pooria NIKNAZAR. Robustness, outliers and Mavericks in network regulation.
2013/8 Per J. AGRELL and Peter BOGETOFT. Benchmarking and regulation.
2013/9 Jacques H. DREZE. When Borch's Theorem does not apply: some key implications of market incompleteness, with policy relevance today.
2013/10 Jacques H. DREZE. Existence and multiplicity of temporary equilibria under nominal price rigidities.
2013/11 Jean HINDRIKS, Susana PERALTA and Shlomo WEBER. Local taxation of global corporation: a simple solution.
2013/12 Pierre DEHEZ and Sophie POUKENS. The Shapley value as a guide to FRAND licensing agreements.
2013/13 Jacques H. DREZE and Alain DURRE. Fiscal integration and growth stimulation in Europe.
2013/14 Luc BAUWENS and Edoardo OTRANTO. Modeling the dependence of conditional correlations on volatility.
2013/15 Jens L. HOUGAARD, Juan D. MORENO-TERNERO and Lars P. OSTERDAL. Assigning agents to a line.
2013/16 Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. First-order methods with inexact oracle: the strongly convex case.
2013/17 Olivier DEVOLDER, François GLINEUR and Yu. NESTEROV. Intermediate gradient methods for smooth convex problems with inexact oracle.
2013/18 Diane PIERRET. The systemic risk of energy markets.
2013/19 Pascal MOSSAY and Pierre M. PICARD. Spatial segregation and urban structure.
2013/20 Philippe DE DONDER and Marie-Louise LEROUX. Behavioral biases and long term care insurance: a political economy approach.
2013/21 Dominik DORSCH, Hubertus Th. JONGEN, Jan.-J. RÜCKMANN and Vladimir SHIKHMAN. On implicit functions in nonsmooth analysis.
2013/22 Christian M. HAFNER and Oliver LINTON. An almost closed form estimator for the EGARCH model.
2013/23 Johanna M. GOERTZ and François MANIQUET. Large elections with multiple alternatives: a Condorcet Jury Theorem and inefficient equilibria.

Recent titles

CORE Discussion Papers - continued

2013/24 Axel GAUTIER and Jean-Christophe POUDOU. Reforming the postal universal service.
2013/25 Fabian Y.R.P. BOCART and Christian M. HAFNER. Fair re-valuation of wine as an investment.
2013/26 Yu. NESTEROV. Universal gradient methods for convex optimization problems.
2013/27 Gérard CORNUEJOLS, Laurence WOLSEY and Sercan YILDIZ. Sufficiency of cut-generating functions.
2013/28 Manuel FORSTER, Michel GRABISCH and Agnieszka RUSINOWSKA. Anonymous social influence.
2013/29 Kent WANG, Shin-Huei WANG and Zheyao PAN. Can federal reserve policy deviation explain response patterns of financial markets over time?
2013/30 Nguyen Thang DAO and Julio DAVILA. Can geography lock a society in stagnation?
2013/31 Ana MAULEON, Jose SEMPERE-MONERRIS and Vincent VANNETELBOSCH. Contractually stable alliances.
2013/32 Jean-François CAULIER, Ana MAULEON and Vincent VANNETELBOSCH. Allocation rules for coalitional network games.
2013/33 Georg KIRCHSTEIGER, Marco MANTOVANI, Ana MAULEON and Vincent VANNETELBOSCH. Limited farsightedness in network formation.
2013/34 Ana MAULEON and Vincent VANNETELBOSCH. Relative concerns and delays in bargaining with private information.
2013/35 Kristof BOSMANS, Koen DECANCQ and Erwin OOGHE. What do normative indices of multidimensional inequality really measure?
2013/36 Alain PHOLO BALA, Dominique PEETERS and Isabelle THOMAS. Spatial issues on a hedonic estimation of rents in Brussels.

Books

G. DURANTON, Ph. MARTIN, Th. MAYER and F. MAYNERIS (2010), *The economics of clusters – Lessons from the French experience*. Oxford University Press.
J. HINDRIKS and I. VAN DE CLOOT (2011), *Notre pension en heritage*. Itinera Institute.
M. FLEURBAEY and F. MANIQUET (2011), *A theory of fairness and social welfare*. Cambridge University Press.
V. GINSBURGH and S. WEBER (2011), *How many languages make sense? The economics of linguistic diversity*. Princeton University Press.
I. THOMAS, D. VANNESTE and X. QUERRIAU (2011), *Atlas de Belgique – Tome 4 Habitat*. Academia Press.
W. GAERTNER and E. SCHOKKAERT (2012), *Empirical social choice*. Cambridge University Press.
L. BAUWENS, Ch. HAFNER and S. LAURENT (2012), *Handbook of volatility models and their applications*. Wiley.
J-C. PRAGER and J. THISSE (2012), *Economic geography and the unequal development of regions*. Routledge.
M. FLEURBAEY and F. MANIQUET (2012), *Equality of opportunity: the economics of responsibility*. World Scientific.
J. HINDRIKS (2012), *Gestion publique*. De Boeck.

CORE Lecture Series

R. AMIR (2002), Supermodularity and complementarity in economics.
R. WEISMANTEL (2006), Lectures on mixed nonlinear programming.
A. SHAPIRO (2010), Stochastic programming: modeling and theory.