MCMC BAYESIAN SPATIAL FILTERING FOR HEDONIC MODELS IN REAL ESTATE MARKETS

Gargallo, P. ; Miguel, J.A. and Salvador, M.J. *

emails: pigarga@unizar,es, jamiguel@unizar.es, salvador@unizar.es
Departamento of Estructura e Historia Económica y Economía Pública
Facultad de Economía y Empresa. University of Zaragoza (SPAIN)

Abstract

The traditional hedonic model postulates that housing prices depend on their characteristics and their location. However, this model assumes a constant relationship between the dependent and the independent variables. This assumption is unrealistic because empirical studies have shown that the regression coefficients depend on the housing location. For this reason, it is necessary to use models with spatially varying coefficients. The approaches proposed in the literature used to estimate this type of models do not incorporate the uncertainty associated with the estimation and selection of models and/or are computationally expensive. To improve these aspects, this paper proposes spatial filtering techniques to parsimoniously model the spatial dependencies of the hedonic coefficients and an adaptive MCMC algorithm of Bayesian variable selection to select the most appropriate filters. The method is illustrated through an application to the real estate market of Zaragoza, and a comparison with alternative procedures is conducted. Our results show that our valuation methodology has better goodness of fit and predictive performance properties than alternative methods. Although our proposal assumes normality and homoscedasticity of the model error distribution, the method is easy to implement and not very computationally demanding, which makes this approach attractive and useful from a practical viewpoint.

Keywords: *Bayesian Inference; Variable Selection; Adaptive MCMC; Hedonic Models; Real Estate Market; Spatial Filtering; Geographically Weighted Regression.*

* This work was supported by grants ECO2012-35029 and ECO2016-79392-P of the Ministerio de Economía y Competitividad (Spain)

1. Introduction

Real estate activities are linked to many sectors of the economy, including construction, finance, and insurance. Therefore, updated fair market housing values are extremely valuable to financial regulators and institutions, municipal assessors, housing index compilers, real estate developers, investors, and many others. Hedonic modelling is the most widely used method to estimate housing prices.

The traditional hedonic model postulates that housing prices mainly depend on their characteristics and locations. The model accounts for locational attributes (location

of the dwelling and proximity to central business districts), neighbourhood attributes (availability of public schools, income levels and population density) and random spatial effects. However, it does not account for spatial interaction effects among dwellings. Empirical evidence shows that prices of neighbouring houses tend to be similar because they share common local factors, such as physical characteristics (age, size, and exterior and interior features) and amenities (socioeconomic status, access to employment opportunities, shopping, public service facilities, and schools). Moreover, information spillovers exist in housing markets, which are manifested through the spatial autocorrelation in prices (Wheeler et al., 2014). Besides, the hedonic model assumes a constant relationship between the dependent variable and the independent variables, which is an unrealistic assumption in housing markets, where it has been observed that the regression coefficients depend on housing location (Goodman and Goodman and Thibodeau, 1998; Fotheringham et al., 2002; Páez et al., 2008; Wheeler et al., 2014).

Several reasons could explain the existence of relationship patterns that could be identified as market segments (Fotheringham et al., 2002; Wheeler et al., 2014). One reason relates to sampling variation because we would not expect the parameter estimates obtained from different samples to be the same. A second reason would be the spatial variations in the attitudes or preferences of people. For instance, the influence of the existence of a garage or a storage room on the price of a house is probably higher if the dwelling is located in the centre or the periphery of a city. A third reason could be the gross misspecification of the model due to omitted spatial explanatory variables or the assumption of an incorrect functional form. Hedonic theory provides little guidance on the choice of the functional form for the hedonic specification (Fleming, 1999).

To capture this spatial heterogeneity in housing markets, several modelling techniques have been proposed. Eckert (1990) suggested that, based on the assumption that subsets are characterized by a lower variance, models generated for housing submarkets should yield greater explanatory power (and predictive accuracy) than those computed at the overall market level. Goodman and Thibodeau (1998) introduced the concept of hierarchical linear modelling, whereby housing characteristics, neighbourhood characteristics, and submarkets interact to influence housing prices. Both of these approaches assume that the submarkets are previously known.

Brunsdon et al. (1996), Fotheringham et al. (1996, 2002) and Páez et al. (2002 a, b) did not assume a previous knowledge of submarkets and proposed to estimate the

regression coefficients for each dwelling using local geographically weighted regression (GWR) techniques. Using this method, an exploration of the variation of the parameters as well as a statistical analysis of the significance of this variation can be carried out. This methodology has received considerable attention in recent years, and some papers have applied GWR to housing markets (Brunsdon et al., 1999; Pavlov, 2000; Fotheringham et al., 2002, Yu, 2006 or Páez et al., 2008 among others). However, this method has been criticized because of the multicollinearity problems in the estimation of the parameters, which are due to the very similar characteristics of houses in the same area, which makes the estimation of the regression coefficients difficult (Wheeler and Tiefelsdorf, 2005; Griffith, 2008; Páez et al., 2011; Bárcena et al., 2014).

Several solutions have been proposed to address this problem. Wheeler (2007, 2009) and Bárcena et al. (2014) used penalized versions of GWR based on regularization methods (ridge and lasso regression) to build parsimonious models that weaken the multicollinearity problem and have good predictive and goodness of fit properties. Another alternative are the Bayesian spatially varying coefficients models (SVC) (Gelfand et al., 2003, 2004, Wheeler and Calder, 2007, Bárcena et al., 2014, Wheeler et al., 2014), which specify a single Bayesian hierarchical model that uses spatially varying coefficient processes to globally model the non-constant linear relationships between the variables.

SVC models have better performance than traditional linear regression models and GWR for both the estimation and prediction of hedonic prices (Wheeler and Calder, 2007; Wheeler et al., 2014). However, they are technically demanding, and their estimation procedures are very computationally intensive for larger samples (Páez and Wheeler, 2009).

Another solution was proposed by Griffith (2008), who formally established an indirect linkage between GWR and spatial filtering via interaction terms and noted how GWR could be viewed as a special case of indirect spatial filtering. The study used the spatial filters developed by Griffith (2000, 2003) to parsimoniously capture the spatial dependencies between the regression coefficients, which significantly alleviated the multicollinearity problem. This methodology is easier to implement than alternative methods and furnishes a way to include non-spatially varying coefficients in the model specification. However, the method is limited by the ability to compute the eigenvectors of a transformed contiguity matrix, and it is computationally intensive due to the large number of possible filters, which increases with the sample size. To solve this last

problem, Griffith (2008) proposed the use of forward-selection variable procedures to select the most relevant filters. Nevertheless, these methods only select one model and do not consider the uncertainty associated with the model selection process or that several models with similar fitness to the data may be possible.

In this paper, we focus our attention on this method, and we propose the use of the adaptive Monte Carlo Markov Chain (MCMC) algorithm from Lamnisos et al. (2013), which is very useful in linear regression problems with a high number of independent variables. This Bayesian selection method allows us to carry out an exhaustive exploration of the model space and takes into account the uncertainty associated with the model estimation and selection processes, which can be very important when the number of filters is high. In addition, and given that the analysis is conjugate, the method is easy to implement and not very computationally demanding. The proposed method is illustrated with an application to the real estate market in the Spanish city of Zaragoza, and a comparison with alternative procedures is also carried out.

The paper is structured as follows. Section 2 reviews the spatial filters model of Griffith (2008), and Section 3 describes the methodology. In Section 4, a case study on the housing prices in the Zaragoza real estate market is presented, and a comparison with alternative methods is carried out. In addition, we present a sensitivity study on several of the model hyperparameters. Section 5 presents the conclusion and identifies future lines of research. A mathematical appendix containing the description of the comparison criteria used in the paper is also included.

2. Spatial filtering in linear regression models with space-varying coefficients

Let $\{P_i;\ i = 1,\ldots,N\}$ be the sale prices of a set of N dwellings and let $\{\mathbf{x}_i = (x_{i1}, \ldots, x_{ip});\ i = 1,\ldots,N\}$ be the values of their hedonic characteristics $(X_1, \ldots, X_p)$'.

Let

$$y_i = \beta_0(u_i,v_i)+\sum_{k=1}^{p} x_{ik}\beta_k(u_i,v_i)+\varepsilon_i\ ;\ i = 1,\ldots, N \qquad (1)$$

be a hedonic linear regression model with spatially varying coefficients where $y_i = \log(P_i)$ is the logarithm of the sale price of dwelling i; (u_i, v_i) denotes the UTM coordinates of the location of dwelling i; $\{\varepsilon_i;\ i = 1,\ldots,N\}$ are independent and identically distributed, verifying that $E[\varepsilon_i] = 0$ and $Var(\varepsilon_i) = \sigma^2$.

This model allows the regression coefficients $\{\boldsymbol{\beta}(u_i,v_i) = (\beta_1(u_i,v_i), \ldots, \beta_P(u_i,v_i))'; i = 1,\ldots,N\}$ to vary across space to capture the local effects of the influence of the dwelling covariates $(X_1, \ldots, X_p)$ on the price of the dwelling.

It seems sensible to assume that observations close to an observation i should have a greater influence on the estimation of the regression coefficients $\boldsymbol{\beta}(u_i,v_i)$. Based on ideas of local regression, Fotheringham et al. (2002) proposed estimating these coefficients using the weighted least squares estimator:

$$\hat{\boldsymbol{\beta}}(u_i,v_i) = \left(\mathbf{X}'\mathbf{W}(u_i,v_i)\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{W}(u_i,v_i)\mathbf{Y}\right) \quad (2)$$

where $\mathbf{X'} = (\mathbf{x}_1,\ldots,\mathbf{x}_N)$ and $\mathbf{W}(u_i,v_i)$ is an (NxN) matrix of the weights that inversely depend on the distance d_{ij} from (u_i,v_i) to other locations (u_j,v_j); j = 1,…,N.

However, as Wheeler and Tiefelsdorf (2005) and Bárcena et al. (2014) show, estimations using expression (2) usually have serious multicollinearity problems because geographically close transactions often have similar hedonic characteristics. Among the suggested solutions to this problem, we can highlight that of Griffith (2008) who proposed the use of spatial filters to capture the most relevant spatial dependencies among the coefficients $\{\beta_k(u_i,v_i); k = 0,1,\ldots p; i = 1,\ldots,N\}$.

Griffith (2008) uses linear expressions of the form:

$$\beta_k(u_i,v_i) = \alpha_k + \sum_{j=1}^{N}\alpha_{kE_j}E_{j,i} \text{ for } k = 0,\ldots,p \quad (3)$$

where $\{\mathbf{E}_j = (E_{j,1}, \ldots, E_{j,N})'; j = 1,\ldots,N\}$ are the eigenvectors of the matrix $\left(\mathbf{I}-\frac{1}{N}\mathbf{1}\mathbf{1}'\right)\mathbf{C}\left(\mathbf{I}-\frac{1}{N}\mathbf{1}\mathbf{1}'\right)$ sorted in a non-decreasing way and $\mathbf{C} = (c_{ij})$ is an (NxN) matrix of connectivity between dwelling locations $\{(u_i,v_i); i = 1,\ldots,N\}$. **C** is usually a 0-1 binary array so that $c_{ij} = 1$ if locations (u_i,v_i) and (u_j,v_j) are geographically connected and 0 otherwise. Griffith (2000, 2003) shows that the first eigenvector, $\mathbf{E}_1$, provides a set of numerical values with the highest value of the spatial correlation Moran coefficient attainable by any set of real numbers observed at the locations $\{(u_j,v_j); j = 1,\ldots,N\}$ whose spatial connections are described by **C**. The second eigenvector, $\mathbf{E}_2$, has the next highest value of the Moran coefficient that is uncorrelated with $\mathbf{E}_1$. The third eigenvector, $\mathbf{E}_3$, has the next highest value of the Moran coefficient that is uncorrelated with $\mathbf{E}_1$ and $\mathbf{E}_2$, and so on. As Griffith (2000) argues, these eigenvectors provide different map patterns describing possible spatial dependencies between locations $\{(u_j,v_j); j = 1,\ldots,N\}$ whose connectivities are described by **C**.

Using (3), model (1) can be written as

$$y_i = \alpha_0 + \sum_{j=1}^{N} \alpha_{0E_j} E_{j,i} + \sum_{k=1}^{p} \alpha_k x_{ik} + \sum_{k=1}^{p} \sum_{j=1}^{N} \alpha_{kE_j} x_{ik} E_{j,i} + \varepsilon_i \; ; \; i = 1,\ldots,N \qquad (4)$$

If N or p is large, expression (4) contains an excessively large number of parameters, most of which might not be significant. Griffith (2008) used forward selection variable procedures to determine which coefficients should be incorporated into the model. However, his approach neglects the uncertainty associated with the estimation and selection of models. To avoid this uncertainty, in the following section, we propose the use of Bayesian procedures that take this uncertainty into account and carry out a deeper exploration of the model space.

3. Statistical Methodology

In this section, we set up a Bayesian variable selection approach to the problem and describe an adaptive MCMC algorithm to solve it based on the method proposed by Lamnisos et al. (2013). Finally, we detail several methods to carry out the model selection using the information provided by our proposed algorithm.

3.1. Bayesian set-up

Let us consider the model $M_{\boldsymbol{\gamma}}$ given by the expression:

$$\mathbf{Y} = \alpha_0 \mathbf{1} + \mathbf{Z}_{\boldsymbol{\gamma}} \boldsymbol{\alpha}_{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon} \quad \text{with } \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)' \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N) \qquad (5)$$

where

$\mathbf{Y} = (y_1, \ldots, y_N)'$;

$\mathbf{1}$ is the vector of N ones;

$\boldsymbol{\gamma} = \left(\gamma_{0E_1},\ldots,\gamma_{0E_N},\gamma_1,\gamma_{1E_1},\ldots,\gamma_{1E_N},\ldots,\gamma_p,\gamma_{pE_1},\ldots,\gamma_{pE_N}\right)' \in \boldsymbol{\Gamma} = \{0,1\}^{p(N+1)+N}$;

$\boldsymbol{\gamma}$ is a vector of indicators whose components take a value of 1 or 0 depending on whether the corresponding variable is included in the model or not

$\mathbf{Z}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ denote the submatrices of matrix $\mathbf{Z}$ and vector $\boldsymbol{\alpha}$ containing the columns and components corresponding to the indicators of $\boldsymbol{\gamma}$ taking the value 1, respectively.

$\mathbf{Z} = \left(\mathbf{E}_1,\ldots,\mathbf{E}_N,\mathbf{X}_1,\mathbf{X}_1 \circ \mathbf{E}_1,\ldots,\mathbf{X}_1 \circ \mathbf{E}_N,\ldots,\mathbf{X}_p,\ldots,\mathbf{X}_p \circ \mathbf{E}_N\right)$ where $\mathbf{X}_j = (x_{1j}, \ldots, x_{Nj})'$ is the vector of observed values of the variable X_j for j = 1,…, p and ◦ denotes the Hadamard product.

$\boldsymbol{\alpha} = \left(\alpha_{0E_1},\ldots,\alpha_{0E_N},\alpha_1,\alpha_{1E_1},\ldots,\alpha_{1E_N},\ldots,\alpha_p,\ldots,\alpha_{pE_N}\right)'$

3.1.2. Model Estimation

The estimation of the parameters of model $M_{\boldsymbol{\gamma}}$ is performed from their posterior distribution calculated using Bayes theorem. As done by Lamnisos et al. (2013), for the intercept, we take the commonly used non-informative improper prior for location parameters:

$$\pi(\alpha_0) \propto 1 \tag{6}$$

while for the regression coefficients, $\boldsymbol{\alpha}_{\gamma}$, we use the multivariate normal prior:

$$\boldsymbol{\alpha}_{\gamma}|\sigma^2, \mathbf{Z}_{\gamma}, M_{\gamma} \sim N_{p_{\gamma}}\left(0, \sigma^2 \mathbf{V}_{\gamma}\right) \tag{7}$$

where $\mathbf{V}_{\gamma} = c\mathbf{I}_{p_{\gamma}}$ (c>0) and p_{γ} is the dimension of $\boldsymbol{\alpha}_{\gamma}$. For the error variance σ^2, we use the usual non-informative improper prior:

$$\pi\left(\sigma^2\right) \propto \frac{1}{\sigma^2} \tag{8}$$

Finally, we take $\pi(\gamma_i = 1) = \frac{1}{2}$, independently, for i = 1,…, p(N+1)+N.

Notice that

$$\boldsymbol{\beta}(u,v) = \begin{pmatrix} \beta_0(u,v) \\ \beta_1(u,v) \\ \dots \\ \beta_p(u,v) \end{pmatrix} = \begin{pmatrix} \alpha_0 + \sum_{j=1}^{N} \mathbf{E}_j \gamma_{0,\mathbf{E}_j} \alpha_{0,\mathbf{E}_j} \\ \alpha_1 + \sum_{j=1}^{N} \mathbf{E}_j \gamma_{1,\mathbf{E}_j} \alpha_{1,\mathbf{E}_j} \\ \dots \\ \alpha_p + \sum_{j=1}^{N} \mathbf{E}_j \gamma_{p,\mathbf{E}_j} \alpha_{p,\mathbf{E}_j} \end{pmatrix} = \begin{pmatrix} \mathbf{h}'_{\mathbf{0},\gamma} \\ \mathbf{h}'_{1,\gamma} \\ \dots \\ \mathbf{h}'_{p,\gamma} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \boldsymbol{\alpha}_{\gamma} \end{pmatrix} = \mathbf{H}_{\boldsymbol{\gamma}} \begin{pmatrix} \alpha_0 \\ \boldsymbol{\alpha}_{\gamma} \end{pmatrix} \tag{9}$$

for the appropriate vectors $\{\mathbf{h}_{j,\gamma};\ j = 0,\dots,p\}$. Using this result, standard calculations prove that the posterior distribution of the regression coefficients $\boldsymbol{\beta}(u,v)|\ M_{\boldsymbol{\gamma}}, \mathbf{Y}, \mathbf{X}$ will be $t_p\left(\mathbf{H}_{\gamma}\hat{\boldsymbol{\mu}}_{\gamma}, \mathbf{H}_{\gamma}\hat{\boldsymbol{\Sigma}}_{\gamma}\mathbf{H}'_{\gamma}, N\text{-}1\right)$, where $t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},df)$ denotes the p-dimensional Student t-distribution with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and df degrees of freedom.

$$\hat{\boldsymbol{\mu}}_{\gamma} = \left(\breve{\mathbf{Z}}'_{\gamma}\breve{\mathbf{Z}}_{\gamma} + \breve{\mathbf{V}}_{\gamma}\right)^{-1}\left(\breve{\mathbf{Z}}'_{\gamma}\mathbf{Y}\right),\ \hat{\boldsymbol{\Sigma}}_{\gamma} = \hat{\sigma}^2_{\gamma}\left(\breve{\mathbf{Z}}'_{\gamma}\breve{\mathbf{Z}}_{\gamma} + \breve{\mathbf{V}}_{\gamma}\right)^{-1},$$

$$\hat{\sigma}^2_{\gamma} = \frac{1}{N-1}\left(\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\breve{\mathbf{Z}}_{\gamma}\left(\breve{\mathbf{Z}}'_{\gamma}\breve{\mathbf{Z}}_{\gamma} + \breve{\mathbf{V}}_{\gamma}\right)^{-1}\breve{\mathbf{Z}}'_{\gamma}\mathbf{Y}\right)$$

where $\breve{\mathbf{Z}}_{\gamma} = \left(\mathbf{1}, \mathbf{Z}_{\gamma}\right)$ and $\breve{\mathbf{V}}_{\gamma} = \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{V}_{\gamma}^{-1} \end{pmatrix}$.

3.2. Model Selection

Given that $\boldsymbol{\gamma}$ is unknown, we use a Bayesian model selection process based on the calculation of the posterior probabilities of each $M_{\boldsymbol{\gamma}}$ model. Applying Bayes theorem, these probabilities are given by the expression:

$$\pi(M_{\boldsymbol{\gamma}}|\mathbf{Y},\mathbf{X}) = \frac{f\left(\mathbf{Y}|\mathbf{Z}_{\gamma},M_{\gamma}\right)\pi\left(M_{\gamma}\right)}{\sum_{\gamma\in\Gamma} f\left(\mathbf{Y}|\mathbf{Z}_{\gamma},M_{\gamma}\right)\pi\left(M_{\gamma}\right)} \tag{10}$$

where $f\left(\mathbf{Y}|\mathbf{Z}_{\gamma},M_{\gamma}\right) = \int f\left(\mathbf{Y}|\mathbf{Z}_{\gamma},\alpha_0,\boldsymbol{\alpha}_{\gamma},\sigma^2,M_{\gamma}\right)\pi\left(\alpha_0,\boldsymbol{\alpha}_{\gamma},\sigma^2|M_{\gamma}\right)d\alpha_0 d\boldsymbol{\alpha}_{\gamma} d\sigma^2$ is the marginal density of $M_{\boldsymbol{\gamma}}$ with $\pi\left(\alpha_0,\boldsymbol{\alpha}_{\gamma},\sigma^2|M_{\gamma}\right)$ being the prior distribution of its parameters and $\pi(M_{\boldsymbol{\gamma}}) = \prod_{i=1}^{N}\pi(\gamma_i) = \frac{1}{2^{p(N+1)+N}}$ is its prior probability. In our case:

$$f\left(\mathbf{Y}\middle|\mathbf{Z}_{\gamma},M_{\gamma}\right) \propto \left|\tilde{\mathbf{Z}}_{\gamma}'\tilde{\mathbf{Z}}_{\gamma}+\mathbf{V}_{\gamma}^{-1}\right|^{-1/2}\left|\mathbf{V}_{\gamma}\right|^{-1/2}\left(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}-\tilde{\mathbf{Y}}'\tilde{\mathbf{Z}}_{\gamma}\left(\tilde{\mathbf{Z}}_{\gamma}'\tilde{\mathbf{Z}}_{\gamma}+\mathbf{V}_{\gamma}^{-1}\right)^{-1}\tilde{\mathbf{Z}}_{\gamma}'\mathbf{Y}\right)^{-(N-1)/2} \tag{11}$$

where $\tilde{\mathbf{Y}} = \tilde{\mathbf{Y}} - \overline{Y}\mathbf{1}_N$, $\overline{Y}$ is the mean of the response Y and $\tilde{\mathbf{Z}}_{\gamma} = \mathbf{Z}_{\gamma} - \overline{\mathbf{Z}}_{\gamma}\mathbf{1}_{p_{\gamma}}'$.

The analytical expression for the marginal density, $f\left(\mathbf{Y}\middle|\mathbf{Z}_{\gamma},M_{\gamma}\right)$, facilitates the application of adaptive MCMC methods for variable selection proposed by Lamnisos et al. (2009, 2013), which are very efficient when the number of independent variables is large. These algorithms are based on the Random Walk Metropolis sampler with three possible movements: "Addition", "Deletion" and "Swapping" of regressors, which are uniformly chosen at random. If "Addition" is selected, then $K^{(t)} + 1$ regressors are chosen to be added to those included in M_{γ}; if "Deletion" is selected, then $K^{(t)} + 1$ regressors are chosen to be removed from the model; and if "Swapping" is selected, then $K^{(t)} + 1$ included regressors are swapped with $K^{(t)} + 1$ excluded regressors without changing the model size (provided $p_{\gamma} \geq K^{(t)} + 1$; if not, the "Addition" step is chosen for $p_{\gamma} < K^{(t)} + 1$, and either the "Addition" or "Swapping" step is chosen for $p_{\gamma} = K^{(t)} + 1$). In the last two cases, the model proposal and reverse model proposal are slightly adjusted to consider these restrictions.

Using this basic scheme, Lamnisos et al. (2013) proposed different algorithms for variable selection in linear and generalized linear regression models, which have large acceptance rates in the Hastings-Metropolis step and good mixing properties. In this paper, we use their Adaptive Metropolis-Hastings Linear Regression (ADMH-LR) algorithm, which proceeds as follows:

Algorithm (ADMH-LR)

1. Select the parameters K (maximum number of variables that can be changed in a movement), ξ_0 (initial probability of the number of variables proposed to be changed) and $\bar{\tau}$ (target acceptance rate of the movement).
2. Draw $\gamma^{(0)}$ from its prior distribution and set $\xi^{(0)} = \xi_0$. Fix $iter_{max}$, which is the maximum number of iterations. Set the number of iterations t = 1.
3. Draw $K^{(t)}$ from a binomial distribution Bi(K-1,$\xi^{(t)}$).
4. Select a movement (Addition, Swapping or Deletion) at random:
   a. If "Addition" is selected, then $\min\{K^{(t)} + 1, p(N+1)+N-p_{\gamma^{(t-1)}}\}$ regressors are chosen to be added to those included in $M_{\gamma^{(t-1)}}$ to form $M_{\gamma'}$.
   b. If "Deletion" is selected, then $\min\{K^{(t)}+1, p_{\gamma^{(t-1)}}\}$ regressors are chosen to be removed from the model $M_{\gamma^{(t-1)}}$ to form $M_{\gamma'}$.
   c. If "Swapping" is selected, then $K^{(t)} + 1$ included regressors are swapped with $K^{(t)} + 1$ excluded regressors without changing the model size (provided $p_\gamma \geq K^{(t)} + 1$; if not, the "Addition" or "Deletion" step is chosen for $p_\gamma < K^{(t)} + 1$). In the last two cases, the model proposal and reverse model proposal are slightly adjusted to consider these restrictions. With the new regressors, form $M_{\gamma'}$.
5. Set $\gamma^{(t)} = \gamma$' with probability

$$\alpha\left(M_{\gamma^{(t-1)}}, M_{\gamma'}\right) = \min\left\{1, \frac{f\left(\mathbf{Y}\middle|\mathbf{Z}_{\gamma'}, M_{\gamma'}\right)\pi\left(M_{\gamma'}\right)q_{\xi^{(t-1)}}\left(M_{\gamma^{(t-1)}}\middle|M_{\gamma'}\right)}{f\left(\mathbf{Y}\middle|\mathbf{Z}_{\gamma^{(t-1)}}, M_{\gamma^{(t-1)}}\right)\pi\left(M_{\gamma^{(t-1)}}\right)q_{\xi^{(t-1)}}\left(M_{\gamma'}\middle|M_{\gamma^{(t-1)}}\right)}\right\}$$

where

$$q_\xi\left(M_{\gamma'}\middle|M_\gamma\right) = \begin{cases} \frac{1}{3\left|\boldsymbol{\gamma}^{+}\right|}\binom{K-1}{j-1}\xi^{j-1}(1-\xi)^{K-j} & \text{if} \quad \sum_{i=1}^{p}\left|\gamma_i' - \gamma_i\right| = j, \text{ addition} \\ \frac{1}{3\left|\boldsymbol{\gamma}^{0}\right|}\binom{K-1}{j-1}\xi^{j-1}(1-\xi)^{K-j} & \text{if} \quad \sum_{i=1}^{p}\left|\gamma_i' - \gamma_i\right| = 2j, \text{ swapping} \\ \frac{1}{3\left|\boldsymbol{\gamma}^{-}\right|}\binom{K-1}{j-1}\xi^{j-1}(1-\xi)^{K-j} & \text{if} \quad \sum_{i=1}^{p}\left|\gamma_i' - \gamma_i\right| = j, \text{ deletion} \\ 0 & \text{otherwise} \end{cases}$$

is the probability of the movement, $|\gamma^+|$ = the number of neighbouring models of M_γ with dimension $p_\gamma + j$, $|\gamma^0|$ = the number of neighbouring models of M_γ with dimension p_γ and $|\gamma^-|$ = the number of neighbouring models of M_γ with dimension $p_\gamma - j$. Otherwise, set $\gamma^{(t)} = \gamma^{(t-1)}$.

6. Compute

$$\xi^{(t)} = \rho\left(\xi^{(t-1)} + s^{(t)}\left(\alpha\left(M_{\gamma^{(t-1)}}, M_{\gamma'}\right) - \bar{\tau}\right)\right)$$

where $\rho(\xi) = \begin{cases} 0 & \text{if} \quad \xi < 0 \\ \xi & \text{if} \quad \xi \in [0,1] \\ 1 & \text{if} \quad \xi > 1 \end{cases}$ and $s^{(t)} = \frac{\xi_0}{t}$.

7. Set t = t+1. If $t \leq iter_{max}$, return to step 3. Otherwise, stop.

As a result of the algorithm, we obtain a sample of models $\hat{\Gamma} = \left\{M_{\gamma^{(t)}}; t = B+1,..., iter_{max}\right\}$ from the posterior distribution $\pi(M|\mathbf{Y},\mathbf{X})$, where B is a burning period to achieve convergence to this distribution. From this sample, we can make inferences about the regression coefficients $\beta(u,v)$ and the valuations of the price P_{new} of a house with some given hedonic characteristics $\mathbf{x}_{new} = (x_{new,1}, \ldots, x_{new,p})'$ and UTM coordinates (u_{new},v_{new}). In this paper, we compare the results of three methods commonly used in the literature. The first model is based on the use of the most likely model $M_{\hat{\gamma}_{max}}$ so that $\pi(M_{\hat{\gamma}_{max}}|\mathbf{Y},\mathbf{Z}) = \max_{\gamma \in \hat{\Gamma}} \{\pi(M_\gamma|\mathbf{Y},\mathbf{Z})\}$. The second model is the median model $M_{\hat{\gamma}_{median}}$ proposed in Barbieri and Berger (2004), which verifies that $\hat{\gamma}_{median,i} = 1$ if $\pi(\gamma_i = 1|\mathbf{Y},\mathbf{Z}) \geq 0.5$ and 0 otherwise, where $\pi(\gamma_i = 1|\mathbf{Y},\mathbf{Z}) = \sum_{\gamma \in \hat{\Gamma}: \gamma_i = 1} \pi\left(M_\gamma \mid \mathbf{Y},\mathbf{Z}\right)$. Finally, the third procedure consists of using a mixture of models $\hat{\Gamma}$ so that the inference for any quantity of interest Δ is based on the following mixture:

$$\hat{\pi}(\Delta \mid \mathbf{Y},\mathbf{Z}) = \sum_{\gamma \in \hat{\Gamma}} f\left(\Delta \mid M_\gamma, \mathbf{Y}, \mathbf{Z}_\gamma\right) \hat{\pi}\left(M_\gamma \mid \mathbf{Y}, \mathbf{Z}_\gamma\right) \tag{12}$$

with $\hat{\pi}\left(M_\gamma \mid \mathbf{Y}, \mathbf{Z}_\gamma\right) = \frac{f\left(\mathbf{Y} \mid M_\gamma, \mathbf{Z}_\gamma\right)\pi\left(M_\gamma\right)}{\sum_{\gamma \in \hat{\Gamma}} f\left(\mathbf{Y} \mid M_\gamma, \mathbf{Z}_\gamma\right)\pi\left(M_\gamma\right)}$.

In particular, if $\Delta = y_{new} = \log(P_{new})$, we calculate the values of the explanatory variables $\mathbf{z}_{new}$ and the inferences about the price P_{new} will be made using a Monte Carlo method from the posterior predictive distributions:

$$f\left(y_{new} \mid M_{\hat{\gamma}_{max}}, \mathbf{z}_{\hat{\gamma}_{max},new}, \mathbf{Y}, \mathbf{Z}\right) \text{ if we use } M_{\hat{\gamma}_{max}} \quad (13)$$

$$f\left(y_{new} \mid M_{\hat{\gamma}_{median}}, \mathbf{z}_{\hat{\gamma}_{median},new}, \mathbf{Y}, \mathbf{Z}\right) \text{ if we use } M_{\hat{\gamma}_{median}} \quad (14)$$

or

$$\sum_{\gamma \in \hat{\Gamma}} f\left(y_{new} \mid M_{\gamma}, \mathbf{z}_{\gamma,new}, \mathbf{Y}, \mathbf{Z}\right) \hat{\pi}\left(M_{\gamma} \mid \mathbf{Y}, \mathbf{Z}_{\gamma}\right) \quad (15)$$

if we use the mixture (12). In cases where the number of models in $\hat{\Gamma}$ is large and, therefore, the calculation of (15) is computationally demanding, we have found it very convenient to use the Occam window $\widetilde{\Gamma} = \left\{ \gamma \in \hat{\Gamma} : \frac{\hat{\pi}\left(M_{\hat{\gamma}_{max}} \mid \mathbf{Y}, \mathbf{Z}_{\gamma}\right)}{\hat{\pi}\left(M_{\gamma} \mid \mathbf{Y}, \mathbf{Z}_{\gamma}\right)} \leq C_{max} \right\}$ from Madigan and Raftery (1995). When $C_{max} = 100$, the results from the Occam window are indistinguishable from those obtained with (15).

4. Empirical application: analysis of the Zaragoza real estate market.

In this section, the methodology described in Section 3 is applied to the Zaragoza real estate market. To provide a more exhaustive study, we compare our methodology with a set of alternative procedures widely used in the literature that correspond to several strategies used to estimate model (4).

4.1. Data

The analysed data correspond to 1,268 resale housing transactions in Zaragoza in 2013. The *Colegio Nacional de Registradores de la Propiedad* provided us with the data on the registered price, the living area, the age of the building and the geographical position measured by UTM coordinates for each dwelling.

Table 1 shows the characteristics of the dwellings. The maximum price corresponds to an apartment located in the centre of Zaragoza, the maximum living area corresponds to a house in an area of luxury chalets, and the maximum age of the building corresponds to an apartment in the old part of the city.

(Insert Table 1 about here)

To clarify the interpretation of each regression coefficient, the variables *living area* and *age of the building* are centred. Thus, the intercept $\beta_0(u,v)$ can be interpreted as the logarithm of the price of a dwelling with mean characteristics located at the geographic coordinates (u, v).

The connectivity matrix **C** was given by $c_{ij} = 1$ if $d_{ij} \leq d_{max} = 300$ metres and 0 otherwise and it is based on the procedure used by the estate agents of the housing

market of Zaragoza to assess housing prices. The eigenvectors $\{\mathbf{E}_i; i = 1,\ldots, 1268\}$ were calculated, and those that verified $100\frac{\lambda_i}{\lambda_{max}} \geq c_{min} = 25\%$ were selected, where λ_i is the eigenvalue associated with $\mathbf{E}_i$ described in Section 2, and λ_{max} is the maximum of the eigenvalues of $\mathbf{C}$. This value of c_{min} was selected after carrying out an out-of-sample validation procedure for a series of values of c_{min} (from 0 to 100 with increases of 1%). To do this, we took a random estimation sample of 1,000 transactions and a validation sample of 268 transactions, and we repeated this procedure 100 times using a stepwise procedure.

Finally, 54 eigenvectors were selected, and as p = 2 in our example, we had a selection problem of 3x54-1 = 161 independent variables, which results in 2^{161} = 2.923 x 10^{48} models, a figure that is clearly unworkable. For this reason, it is necessary to design efficient searching methods to locate the models that best fit the data.

4.2. Estimation and model selection procedures

We used 7 estimation and model selection procedures, namely:

a) *Constant:* A constant model where we assume that the regression coefficients $\beta(u,v) = \beta$ are constant and without any spatial dependence. Therefore, we take $Z_\gamma = (\mathbf{X}_1,\ldots, \mathbf{X}_p)$.

b) *Stepwise:* A stepwise selection procedure implemented using the routine *stepwisefit* programmed in *MATLAB R2015 b* with an input p-value of 0.05 and output p-value of 0.10.

c) *Lasso:* A Bayesian lasso estimation procedure implemented using the routine *blasso* programmed in R and included in the package *monomvn*. This provides inferences for Bayesian lasso by Gibbs sampling from the Bayesian posterior distribution augmented with a Reversible Jump for model selection (Park and Casella, 2008).

d) *Ridge:* A Bayesian ridge estimation procedure implemented using the routine *bridge* programmed in R and included in the package *monomvn*. This is a special case of *blasso* with the argument *case* ="*ridge*".

e) *Most Probable:* Algorithm ADMH-LR, selecting the most likely model $M_{\hat{\gamma}_{max}}$.

f) *Median:* Algorithm ADMH-LR, selecting the median model of Barbieri and Berger (2004).

g) *Mixture:* Algorithm ADMH-LR, using the Bayesian mixture (12) to make inferences.

In the *Stepwise* procedure, the estimations of the regression coefficients were obtained using the maximum likelihood method and, for confidence intervals of 95% and 99%, we assumed asymptotic normality for the estimator. In the Bayesian *Lasso*, the Bayesian *Ridge* and the ADMH-LR procedures, we took c = 100, and we calculated the posterior median as the point estimate together with the Bayesian credibility intervals of 95% and 99%, which were obtained from the corresponding quantiles. In addition, in the ADMH-LR algorithm, we used the values proposed in Lamnisos et al. (2013) with N = 4, $\xi_0 = 0.5$, $\bar{\tau} = 0.3$ and we took $iter_{max}$ = 100,000 iterations with a burning period B = 10,000. The computations were completed using the programs *MATLAB R2015 b* and *R 3.3.2.*

Constant was chosen to test if there were significant spatial dependencies in the regression coefficients. *Stepwise* is similar to that proposed by Griffith (2008), and it is used as a reference. *Lasso* and *Ridge* are widely used in the literature as alternatives to the ordinary least squares (OLS) method, and they improve the average prediction error and implicitly provide a variable selection procedure (Hans, 2009). Finally, *Most Probable*, *Median* and *Mixture* are the Bayesian procedures proposed in this paper. Only the *Mixture* procedure takes into account the uncertainty associated with the estimation and the model selection processes because it builds a mixture of all the relevant models found weighted by their posterior probability. These probabilities measure their fit to the data.

4.3. Results

4.3.1. Estimations of the regression coefficients

Table 2 shows the correlations between the estimations of the parameters $\{\beta_k(u_i,v_i);i = 1,\ldots,N\}$ for k = 0,1,2 and Figure 1 shows the scatter plot matrix of the estimations obtained by each procedure[1]. Except for the *Constant* procedure, which estimates $\hat{\beta}_0 = 11.4348$, $\hat{\beta}_1 = 1.0285$ and $\hat{\beta}_2 = -0.2548$, all procedures provide very similar estimations with correlations larger than 0.89, 0.80 and 0.65 between the intercept, living area and building age regression coefficients, respectively. This result can also be seen in Figure 1, where a non-parametric regression line was added to each

[1] Stepwise estimations were obtained by MLE. For the rest of the procedures, the estimations were equal to the posterior median.

scatter plot that it is approximately linear for most of them. The larger differences correspond to those obtained with the *Ridge* and *Lasso* procedures (especially in the living area and building age coefficients). A slight shrinkage effect towards 0, which provides a trend to obtain smaller absolute values of these coefficients, can be seen.

(Insert Table 2 and Figure 1 about here)

Figures 3-5 show the estimations of parameters β_0 (Figure 3), β_1 (Figure 4) and β_2 (Figure 5) for the logarithm of the price of a dwelling with average characteristics and the price elasticities relative to the living area and the age of the building, respectively. Additionally, Figure 2 displays a map of the city in which the names of the main districts are shown to make it easier to read the results. Given that the three ADMH-LR Bayesian methods provided essentially the same results, Figures 3-5 only show the estimations corresponding to the *Mixture* procedure and they are compared with those provided by the *Stepwise*, *Lasso* and *Ridge* procedures.

(Insert Figure 2 about here)

The prices of a dwelling with average characteristics are distributed among the different urban areas of the city as expected (see Figure 3). The highest prices are in the centre and residential areas (*Actur* and *Universidad*), and the lowest prices are in the old part of the city and the traditionally working-class neighbourhoods (*Las Fuentes*, *San José*, *Las Delicias* and the oldest houses of *Torrero* and *El Rabal*).

(Insert Figure 3 about here)

The signs of the price elasticities with respect to the living area and the age of the building were as expected: positive for the living area, indicating that the bigger the living area, the higher the price (see Figure 4), and negative for the age of the building, reflecting an age penalization on the price of a dwelling (see Figure 5).

The strongest effects of the living area correspond to houses located in the *San José* district and residential areas near the centre of the city, all with heterogeneous living areas. The weakest effects correspond to working-class neighbourhoods, such as *Las Delicias* or the old part of *Torrero*, where houses are fairly homogeneous in living areas, and to the most expensive dwellings in downtown areas where the location is more important than living area. The major differences among the methods compared are found in some peripheral parts of the city (*Actur, Casablanca, Montecanal* and *Valdespartera*). The Bayesian estimations (*Lasso*, *Ridge* and *Mixture* methods) are more reasonable because these dwellings are very homogeneous in living area, so its

influence must be lower. In these areas, the shrinkage effect towards 0 of the *Lasso* and *Ridge* methods is more notable, providing lower absolute values than the other two methods.

(Insert Figure 4 about here)

The strongest effect of the age of the building can be seen in the *Las Delicias* district where there is more heterogeneity of dwellings with respect to age. By contrast, the weakest effects are found in the areas that are more homogeneous in age such as *El Rabal* and *Casco Histórico* (old houses in both cases) or the new part of *Torrero* (new housing). The major differences among the methods compared are seen in *Las Fuentes, Actur* and some peripheral parts (*Montecanal, Valdefierro, Valdespartera, Casablanca*). Again, the Bayesian estimations seem more realistic in *Las Fuentes* due to the coexistence of ancient poor-quality dwellings and higher quality houses, which leads to notable differences in price and a strong influence of age. In contrast, all other areas correspond to expansion zones of the city where the ages of the dwellings are very similar and therefore the influence of age is lower. Finally, in these areas, the shrinkage effect towards 0 of the *Lasso* and *Ridge* procedures can also be seen.

(Insert Figure 5 about here)

4.3.2. Goodness of fit and intra-sample model comparison

Table 3 shows the intra-sample value of the predictive criteria described in the Appendix for the seven estimation and model selection procedures. Most of the Bayesian methods yield similar results, and they are better than those from the *Stepwise* procedure (and the *Most Probable* method, which selects the same model) in all criteria. The worst performance in all criteria corresponds to the *Constant* procedure, which provides statistical evidence against the constant regression coefficient hypothesis. The most significant differences correspond to the LPRED criterion (and, with lower intensity, to the LOSS-GR criteria), which evaluates the goodness of fit of models in terms of the posterior predictive density. The *Mixture* procedure has the best performance, highlighting the advantages of considering the uncertainty associated not only with the estimation of the parameters of the model but also with the model selection process.

Empirical coverage levels of the predictive intervals are well fitted at 99% but show a significant tendency for over-coverage at 95% in all procedures, which is most likely due to the use of a normal distribution for the error term. It would be interesting

to use more general distributions (Student t and GED), which will be left for future research.

(Insert Table 3 about here)

Figure 6 compares the point forecast of the logarithm of the prices with their true values and shows the 95% and 99% limits of the credibility intervals obtained by the *Mixture* procedure. The behaviours of the predictions obtained by all procedures are very similar and reasonable, and the confidence bounds adequately reflect the uncertainty associated with these predictions.

(Insert Figure 6 about here)

4.3.3 Out-of-sample analysis

Finally, an out-of-sample validation process was performed, where 1,000 transactions were taken at random to estimate the model and the rest were used for the validation, resulting in a total of 268 out-of-sample observations. The process was repeated 100 times. Table 4 presents the values of the comparison criteria while Figure 7 shows the boxplot of their values. In general, all methods adequately capture the evolution of prices with coverage levels of the predictive intervals similar to their confidence/credibility levels. The worst performance again corresponds to the *Constant* method, so the hypothesis of constant regression coefficients is rejected. The best performance corresponds to the *Lasso* procedure with the exceptions of the PMAD and LPRED criteria where the *Mixture* procedure performs better. However, the differences are small with the only exception being the LPRED criterion, in which the ADMH-LR procedures have a significantly better performance than the rest. This difference highlights the advantages of these algorithms for obtaining a better goodness of fit to data in terms of the posterior predictive density.

(Insert Table 4 and Figure 7 about here)

In summary, the ADMH-LR procedures tend to select models with better goodness of fit properties in terms of predictive densities both intra-sample and out-of-sample. These procedures have a similar performance to the other procedures with respect to the rest of the criteria. In particular, the *Mixture* procedure, which takes into account not only the uncertainty associated with the estimation of the parameters but also that associated with the model selection process, tends to have the best predictive performance. Our recommendation is to use the *Mixture* procedure to estimate linear regression models with spatially varying coefficients.

4.4. Sensitivity studies

In this section, we carry out a sensitivity study with respect to some of the model hyper-parameters. We study the sensitivity of the results to the diffuseness of parameter c of the prior distribution and to the c_{min} and d_{max} thresholds, which determine the number of eigenvectors $\{\mathbf{E}_i;\ i = 1,\ldots, 1268\}$ and the connectivity matrix **C** used to describe the spatial dependencies of the regression coefficients, respectively. In all cases, we use our preferred *Mixture* procedure to carry out the estimation and model selection processes.

4.4.1. Sensitivity analysis with respect to the prior

We considered four different priors (7) of increasing orders of magnitude for non-informativeness corresponding to c = 10, 100, 1,000 and 10,000. The rest of the hyperparameters were fixed at their previous values, i.e., c_{min} = 25% and d_{max} = 300. Table S.1[2] shows the distribution of the number of variables selected by the ADMH-LR algorithm. The greater the non-informative character of the prior, the fewer variables selected.

This is a consequence of the well-known Lindley paradox (Lindley, 1957) which proved that a very non-informative prior could favour the selection of excessively parsimonious models. In our case, and given that the amount of data is large, it follows from (11) that if c is large, then

$$f\left(\mathbf{Y}\middle|\mathbf{Z}_\gamma, M_\gamma\right) \approx \left|\tilde{\mathbf{Z}}_\gamma'\tilde{\mathbf{Z}}_\gamma\right|^{-1/2} c^{-\frac{p_\gamma}{2}} \left(\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} - \tilde{\mathbf{Y}}'\tilde{\mathbf{Z}}_\gamma\left(\tilde{\mathbf{Z}}_\gamma'\tilde{\mathbf{Z}}_\gamma\right)^{-1}\tilde{\mathbf{Z}}_\gamma'\mathbf{Y}\right)^{-(N-1)/2}$$

where the term $c^{-\frac{p_\gamma}{2}}$ penalizes complex models (large p_γ) by decreasing their posterior probabilities and making their selection more difficult. For this reason, it is convenient to study the sensitivity of the results obtained by this parameter.

Table S.2 and Figure S.1 show the correlations and the scatter plot matrix between the point estimations of the regression coefficients, respectively. In general, all correlations are high with values above 0.75 (see Table S.2) and Figure S.1 also displays a high level of concordance between all estimations. The closer the values of c, the higher the correlations. Furthermore, the higher the values of c, the lower the shrinkage effect towards 0 of the prior (7). This result provides estimations of the regression coefficients, especially those of the living area and building age variables,

[2] All tables and figures referring to this sensitivity analysis have been placed in the Supplementary Material section. Therefore, their numbering begins with the letter S, for example, ... Table S.1, Table S.2, ... Figure S.1, Figure S.2, ...

which are slightly smaller in absolute value. The maps of the coefficients estimated using each prior did not reveal the existence of patterns significantly different from those shown in Figures 3 to 5 and are omitted for the sake of brevity but are available on request.

Finally, Table S.3 and Figure S.2 show the results of the intra-sample and out-of-sample validation of the models selected for each prior[3]. The results of the empirical coverage levels of the predictive intervals are very similar in all cases and are similar to their confidence/credibility levels with the only exception being the 95% intra-sample predictive intervals, in which a significant tendency for over-coverage is observed. The values of the intra and out-of-sample comparison criteria of the selected models are very similar. The selected models with c = 10 have the best performances in most of the criteria. Only in the LPRED criterion are there significant differences among the models selected with c = 10, 1,000 and 10,000, with the first having a better performance. This fact highlights the importance of carrying out sensitivity studies with respect to the prior distribution to avoid the selection of models with poor goodness of fit due to the effects of the Lindley paradox.

4.4.2. Sensitivity analysis with respect to c_{min}

This parameter controls the number of eigenvectors $\{\mathbf{E}_i; i = 1,\ldots, 1268\}$ that are used to model the spatial dependencies of the regression coefficients. The higher the value of c_{min}, the bigger the number of eigenvectors and, therefore, the higher the flexibility of the model to capture more complex spatial dependencies and the greater the possibility of capturing spurious dependencies. We considered three different values of c_{min} (20%, 25% and 30%) while the rest of the hyperparameters were fixed at their previous values, i.e., c = 100 and d_{max} = 300.

Table S.4 shows the distribution of the number of variables selected by the ADMH-LR algorithm. It can be observed that this number is very similar for c_{min} = 20% and 25% and slightly smaller for c = 30%.

Table S.5 and Figure S.3 show the correlations and the scatter plot matrix of the estimations of the regression coefficients, respectively. Table S.6 and Figure S.4 show the results of the intra-sample and out-of-sample validation of the models selected for each value of c_{min}, respectively. In general, all correlations are high with values above 0.60 and a high level of concordance between all solutions (see Figure S.3). The closer

[3] As in Section 4.3.3, we selected 1,000 observations to estimate the model and 268 to validate it, and we replicated this process 100 times in all sensitivity studies.

the values of c_{min}, the higher the correlations and the concordance of the estimations. The main differences are between the estimations obtained with c_{min} = 20% and 30%, especially in the living area and building age coefficients (see Figure S.3), where the smallest values of the correlation are obtained (see Table S.5). This difference is due to the elimination of some eigenvectors $\{\mathbf{E}_i;\ i = 1,\ldots, 1268\}$ when c_{min} = 30%, which reduces the flexibility of the model to describe the spatial dependency modelling process and lowers the intra-sample goodness of fit and the predictive performance of the selected models. However, a map of the coefficients estimated by each model did not reveal the existence of patterns significantly different from those shown in Figures 3 to 5. Furthermore, the out-of-sample performances of the three estimated mixture models are very similar (see Table S.6 and Figure S.4). Therefore, we conclude that our results are reasonably robust to the value of c_{min}.

4.4.3 Sensitivity analysis with respect to $\mathbf{d}_{max}$

The hyperparameter d_{max} determines the values of the connectivity matrix **C** and the neighbouring transactions of a given transaction in such a way that the higher its value, the larger the number of neighbouring transactions. We considered three different values of d_{max} (250, 300 and 350 metres) while the rest of the hyperparameters were fixed at their previous values, i.e., c = 100 and c_{min} = 25%.

Table S.7 shows the distribution of the number of variables selected by the ADMH-LR algorithm. Table S.8 and Figure S.5 compare the estimations of the regression coefficients. Finally, Table S.9 and Figure S.6 show the results of the intra-sample and out-of-sample validation of the selected models.

The more complex models selected correspond to d_{max} = 300, which tend to have the best intra-sample and out-of-sample performances for most criteria. The regression coefficients estimated by the mixture models selected when d_{max} = 250 and 350 are significantly correlated with those estimated when d_{max} = 300; all of the correlations are larger than 0.5. Finally, a map of the coefficients estimated by each method did not reveal the existence of patterns significantly different from those shown in Figures 3 to 5. Therefore, we can conclude that the value of d_{max} provided by the estate agents of Zaragoza is well supported empirically.

5. Conclusions

This paper has proposed a Bayesian procedure for selecting spatial filters to conduct valuation procedures for dwellings using regression models with spatially

varying coefficients. The procedure is based on the method proposed by Griffith (2008) and applies Bayesian techniques of variable selection and, more specifically, MCMC adaptive methods to determine the most relevant spatial dependencies in the data. Using the selected models, we have proposed three parameter estimation procedures: the most likely model, the median model and the Bayesian mixture model. The proposed methods are illustrated by applying them to the real estate market of Zaragoza, and a comparison with alternative procedures has been carried out.

The methodology is very flexible and parsimoniously describes the spatially varying relationships of the regression coefficients. In addition, the method is easy to implement and not very computationally demanding, as it does not require excessively high computational times. In simulated examples similar to the empirical example considered in this study, the average CPU time of the estimation algorithm was approximately 320 seconds with a standard deviation of approximately 70 seconds using an Ultrabook Toshiba Kira computer with an Intel Core i7 processor. Moreover, the results show that the valuation methodology proposed in the paper and the Bayesian Mixture procedure improve the results obtained by alternative procedures and more adequately reflect the uncertainty associated with the estimation and selection of models.

One limitation of our work is the assumption of the normality and homoscedasticity of the error distribution. Both of these hypotheses are questionable and might be responsible for the over-coverage at the 95% level observed in all procedures. A right-skewed error distribution or a heteroscedastic error might be more appropriate. It would be interesting to extend our procedures to address these situations in future studies.

Other prior distributions of the regression coefficients could be used in the line of the g-priors of Zellner (1986) and Liang et al. (2008) to avoid problems of local multicollinearity. Moreover, given that the number of independent variables might be much higher than the number of data points, high dimensional statistical techniques of variable selection (Giraud, 2014) could be applied. Similarly, Bayesian comparison procedures of non-nested models could be used to select variables from several forms of the connectivity matrix **C**.

An interesting alternative approach to those considered here is the structured additive regression (STAR) model of Fahrmeir et al. (2004, 2010). They proposed the use of Bayesian spatiotemporal extensions of generalized additive and varying

coefficients models for analysing space-time regression data, which include the GWR models. They used a two-dimensional version of penalized splines to capture the spatial dependencies of the regression coefficients and used MCMC methods (among other possibilities) to estimate the coefficients. Their methodology is very flexible and general and can be extended to generalized additive (Brezger and Land, 2006) and hierarchical regression models (Lang et al., 2014) and to variable selection problems (Fahrmeir et al., 2010). It would be interesting to compare the STAR methodology with our procedures in future research. All of these aspects are on our agenda for future research, the results of which will be presented in later papers.

References

Barbieri, M. and Berger, J.O. (2004): "Optimal Predictive Model Selection". *The Annals of Statistics*, 32 (3), 870-897.

Bárcena, M.J.; Ménendez, P.; Palacios, M.B. and Tusell, F.T. (2014): "Alleviating the effect of collinearity in geographically weighted regression". *Journal of Geographical Systems,* 16, 4, 441-466.

Brezger, A. and Lang S. (2006): "Generalized structured additive regression based on Bayesian P-splines". *Computational Statistics & Data Analysis*, 50, 967-991.

Brunsdon, C.; Fotheringham, A.S. and Charlton, M. (1996): "Geographically weighted regression: a method for exploring spatial non-stationary". *Geographical Analysis,* 28, 281-298.

Brunsdon, C.; Fotheringham, A.S. and Charlton, M. (1999): "Some notes on parametric tests for geographically weighted regression". *Journal of Regional Sciences,* 39, 3, 497-524.

Eckert, J.K. (1990) *Property appraisal and assessment administration.* Chicago, II: International Association of Assessing Officers.

Fahrmeir, L.; Kneib, T. and Lang, S. (2004): "Penalized Structured Additive Regression for Space-Time Data: A Bayesian Perspective". *Statistica Sinica,* 14, 731-761.

Fahrmeir, L; Kneib, T. and Konrath, S. (2010): "Bayesian regularization in structured additive regression: a unifying perspective on shrinkage, smoothing and predictor selection". *Statistics and Computing,* 20, 203-219.

Fleming, M.M. (1999) Growth controls and fragmented suburban development: the effect on land values. *Geographical Information Sciences,* 5, 154-162.

Fotheringham, S.; Charlton, M. and Brundsdon, C. (1996): "The geography of parameter space: an investigation of spatial non-stationary". *International Journal of Geographical Information Science,* 10, 605-627.

Fotheringham, S.; Charlton, M. and Brundsdon, C. (2002): *Geographically weighted regression: the analysis of spatially varying relationships. Wiley. New York*

Gelfand, A.E.; Ecker, M.D.; Knigh and J.R., Sirmans, C.F. (2004): "The dynamics of location in home price". *Journal of Real Estate Finance and Econometrics*, 29 (2) , 149–167.

Gelfand, A.E.; Kim, H.J.; Sirmans, C. and Banerjee, S. (2003): "Spatial modeling with spatially varying coefficient processes". *Journal of the American Statistical Association* 98 (462), p. 387-396.

Giraund, C. (2014): *Introduction to High-Dimensional Statistics*. Monographs on Statistics and Applied Probability 139. CRC Press.

Gneiting, T. and Raftery, A.E. (2007): "Strictly Proper Scoring, Rules, Prediction, and Estimation". *Journal of the American Statistical Association,* 102 (477), 359-378.

Goodman, A.C. and Thibodeau, T.G. (1998): "Housing market segmentation". *Journal of Housing Economics,* 7, 121-143.

Griffith, D.A. (2000): "A linear regression solution to the spatial autocorrelation problem". *Journal of Geographical Systems,* 2 (2), 141-156.

Griffith, D.A. (2003): *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization. Springer, Berlin.*

Griffith, D.A. (2008): "Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR)". *Environment and Planning Series A*, 40 (11), 2751-2769.

Hans, C. (2009): "Bayesian lasso regression". *Biometrika,* 96 (4), 835-845.

Lamnisos, D.; Griffin, J.E. and Steel, M.F.J. (2009): "Transdimensional Sampling Algorithms for Bayesian Variable Selection in Classification Problems with many more variables than observations". *Journal of Computational and Graphical Statistics,* 18, 592-612.

Lamnisos, D.; Griffin, J.E. and Steel, M.F.J. (2013): "Adaptive Monte Carlo for Bayesian Variable Selection in Regression Models". *Journal of Computational and Graphical Statistics*, 22 (3), p. 729-748.

Lang, S.; Umlauf, N.; Wechselberger, P.; Harttgen, K. and Kneib, T. (2014) Multilevel structured additive regression. *Statistics and Computing,* 24, 223-238.

Liang, F.; Paulo, R.; Molina, G.; Clyde, M. and Berger, J. (2008): "Mixtures of g-Priors for Bayesian Variable Selection". *Journal of the American Statistical Association,* 103 (481), 410-423.

Lindley, D. W. (1957): "A statistical paradox". *Biometrika,* 44 (1-2), 187-192.

Madigan, A. and Raftery, A.E. (1995): "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window". *Journal of the American Statistical Association,* 89 (428), 1535-1546.

Páez, A.; Uchida, T. and Miyamoto, K. (2002 a): "A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity". *Environment and Planning Series A*, 34 (4), 733-754.

Páez, A.; Uchida, T. and Miyamoto, K. (2002 b): "A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests". *Environment and Planning Series A*, 34 (4), 883-904.

Páez, A; Farber, S. and Wheeler, D.C. (2011): "A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships". *Environment and Planning Series A*, 43 (12), 2992–3010.

Páez, A; Long, F. and Farber, S. (2008): "Moving Window Approaches for Hedonic Price Estimation: An Empirical Comparison of Modeling Techniques". *Urban Studies,* 45 (8), 1565-1581.

Páez, A and Wheeler, D. C. (2009): "Geographically weighted regression", in *International Encyclopedia of Human Geography.* Eds Kitchin, R and Thrift, N. (Elsevier, Amsterdam).

Park, T., Casella, G. (2008): "The Bayesian Lasso". *Journal of the American Statistical Association*, 103(482), 681-686.

Pavlov, A. (2000): “Space varying regression coefficients: a semi-parametric approach applied to real estate markets”. *Real Estate Economics*, 28 249-283.

Wheeler, D.C. (2007): “Diagnostic tools and a remedial method for collinearity in linear regression models with spatially varying coefficients”. *Environment and Planning Series A*, 39 (10), 2464-2481.

Wheeler, D.C. (2009): “Simultaneous coefficient penalization and model selection in geographically weighted regression”. *Environment and Planning Series A*, 41 (3), 722-742.

Wheeler, D.C. and Calder, C. (2007): “An assessment of coefficient accuracy in linear regression models with spatially varying coefficients”. *Journal of Geographical Systems,* 9 (2), 145-166.

Wheeler, D.C.; Páez, A.; Spinney, J. and Waller, L. (2014): “A Bayesian approach to hedonic price analysis”. *Papers in Regional Science,* 93 (3), 663-683.

Wheeler, D.C. and Tiefelsdorf, M. (2005): “Multicollinearity and correlation among local regression coefficients in geographically weighted regression”. *Journal of Geographical Systems,* 7 (2), 161-187.

Yu, D.L. (2006): “Spatially varying development mechanisms in the Greater Beijing area: a geographically weighted regression investigation”. *The Annals of Regional Science,* 40, 1, 173-190.

Zellner, A. (1986): “On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions”, in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, Amsterdam: North-Holland,* 233-243.

APPENDIX

Predictive criteria

Let $\mathbf{y}_{val} = (y_{val,1},\ldots,y_{val,nval})'$ be the sample of validation and $\mathbf{y}_{est} = (y_{est,1},\ldots,y_{est,nest})'$ be the sample of estimation where *nval* and *nest* are the corresponding sample sizes. Let $\mathbf{Z}_{val} = (\mathbf{z}_{val,1},\ldots,\mathbf{z}_{val,nval})$ and $\mathbf{Z}_{est} = (\mathbf{z}_{est,1},\ldots,\mathbf{z}_{est,nest})$ be their corresponding covariates.

In the in-sampling validation we take $\mathbf{y}_{val} = \mathbf{y}_{est} = \mathbf{y}$ and $\mathbf{Z}_{val} = \mathbf{Z}_{est} = \mathbf{Z}$**.** In the out-sampling validation we take $\mathbf{y}_{est} = (y_1,\ldots,y_{nest})'$ and $\mathbf{y}_{val} = (y_{nest+1}, \ldots, y_n)'$ and $\mathbf{Z}_{est} = (\mathbf{z}_1,\ldots,\mathbf{z}_{nest})$, $\mathbf{Z}_{val} = (\mathbf{z}_{nest+1},\ldots,\mathbf{z}_n)$ where the components of $\mathbf{y}$ have been previously randomly sorted.

We have used the following criteria:

1. Root of the Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{nval}}\sum_{i=1}^{\mathrm{nval}}\left(y_{\mathrm{nval,i}} - \mathrm{E}\left[y_{\mathrm{val,i}} \mid \mathbf{z}_{\mathrm{val,i}}, \left(\mathbf{y}_{\mathrm{est}}, \mathbf{Z}_{\mathrm{est}}\right), \hat{\mathbf{M}}\right]\right)^2}$$

where $\hat{\mathbf{M}}$ is the model selected by the evaluated procedure; $\mathrm{E}\left[y_{\mathrm{val,i}} \mid \mathbf{z}_{\mathrm{val,i}}, \left(\mathbf{y}_{\mathrm{est}}, \mathbf{Z}_{\mathrm{est}}\right), \hat{\mathbf{M}}\right] = \mathbf{z}'_{\hat{\gamma},\mathrm{val,i}}\hat{\boldsymbol{\mu}}_{\hat{\gamma}}$, where $\hat{\gamma} = 1_{N+p+Np}$ in the case of the *Lasso* and *Ridge* procedures, $\hat{\gamma} = \left(\hat{\gamma}_i; i = 1,..., N+p+Np\right)'$ with $\hat{\gamma}_i = 1$ if the i-th variable of $\mathbf{Z}$ is selected and 0 otherwise in the case of the *Stepwise* and *Bayesian Most Probable and Median* procedures; $\hat{\mathrm{M}}$ is the mixture given by $\sum_{\gamma\in\hat{\Gamma}}\pi\left(\mathrm{M}_\gamma \middle| \mathbf{Y}, \mathbf{Z}_\gamma\right)\mathrm{M}_\gamma$ in the case of *the* Bayesian *Mixture* procedure and $\sum_{\gamma\in\hat{\Gamma}}\mathrm{w}\left(\mathrm{M}_\gamma \middle| \mathbf{Y}, \mathbf{Z}_\gamma\right)\mathrm{M}_\gamma$ with $w\left(\mathrm{M}_\gamma | \mathbf{Z}_\gamma, \mathbf{Y}\right) = \dfrac{\mathrm{f}\left(\mathbf{Y} | \mathbf{Z}_\gamma, \mathbf{Y}, \mathrm{M}_\gamma\right)}{\sum_{\gamma\in\hat{\Gamma}}\mathrm{f}\left(\mathbf{Y} | \mathbf{Z}_\gamma, \mathbf{Y}, \mathrm{M}_\gamma\right)}$; $\hat{\boldsymbol{\mu}}_{\hat{\gamma}}$ is an estimation of $\left(\alpha_0, \boldsymbol{\alpha}'_{\hat{\gamma}}\right)'$ obtained by maximum likelihood from model $\mathbf{M}_{\hat{\gamma}}$ in the case of the *Stepwise* procedure, and equal to the posterior mean of $\left(\alpha_0, \boldsymbol{\alpha}'_{\hat{\gamma}}\right)'$ for the rest of procedures.

2. Mean Absolute Deviation (MAD)

$$\mathrm{MAD} = \sqrt{\frac{1}{\mathrm{nval}}\sum_{i=1}^{\mathrm{nval}}\left|y_{\mathrm{nval,i}} - \mathrm{E}\left[y_{\mathrm{val,i}} \mid \mathbf{z}_{\mathrm{val,i}}, \left(\mathbf{y}_{\mathrm{est}}, \mathbf{Z}_{\mathrm{est}}\right), \hat{\mathbf{M}}\right]\right|}$$

3. Logarithm of the density of the predictive distribution

$$\text{LPRED} = \log \text{f}\left(\mathbf{y}_{\text{val}} \mid \mathbf{Z}_{\text{val}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}\right)$$

where $\text{f}\left(\mathbf{y}_{\text{val}} \mid \mathbf{Z}_{\text{val}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}\right)$ denotes the predictive density of $\mathbf{y}_{\text{val}}$ that is equal to $N_{nval}\left(\mathbf{Z}_{val}^{'}\hat{\boldsymbol{\mu}}_{\hat{\gamma}}, \hat{\sigma}_{\hat{\gamma}}^{2}\left(I_{nval} + \mathbf{Z}_{val}^{'}\hat{\boldsymbol{\Sigma}}_{\hat{\gamma}}\mathbf{Z}_{val}\right)\right)$ in the case of *Stepwise* procedure, $T_{nval}\left(\mathbf{Z}_{val}^{'}\hat{\boldsymbol{\mu}}_{\hat{\gamma}}, \hat{\sigma}_{\hat{\gamma}}^{2}\left(I_{nval} + \mathbf{Z}_{val}^{'}\hat{\boldsymbol{\Sigma}}_{\hat{\gamma}}\mathbf{Z}_{val}\right), nest-1\right)$ in the case of the *Most Probable* and *Median* procedures and $\sum_{\gamma \in \hat{\Gamma}} \pi\left(\text{M}_{\gamma} \left| \mathbf{y}_{est}, \mathbf{Z}_{est,\gamma}\right.\right) T_{nval}\left(\mathbf{Z}_{val}^{'}\hat{\boldsymbol{\mu}}_{\gamma}, \hat{\sigma}_{\gamma}^{2}\left(I_{nval} + \mathbf{Z}_{val}^{'}\hat{\boldsymbol{\Sigma}}_{\gamma}\mathbf{Z}_{val}\right), nest-1\right)$ in the case of the *Mixture* procedure. $\text{T}_{\text{p}}(\mu,\Sigma,\nu)$ denotes the p-dimensional Student t-distribution with mean vector μ, scale matrix Σ and ν degrees of freedom. In the case of the *Lasso* and *Ridge* procedures, $\text{f}\left(\mathbf{y}_{\text{val}} \mid \mathbf{Z}_{\text{val}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}\right)$ is calculated by means of composition sampling.

4. Percentage Mean Absolute Deviation (PMAD)

$$\text{PMAD} = \frac{1}{\text{nval}} \sum_{\text{i=1}}^{\text{nval}} \frac{\left|\text{P}_{\text{val,i}} - \text{Median}\left[\text{P}_{\text{val,i}} \mid \mathbf{z}_{\text{val,i}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}\right]\right|}{\left|\text{P}_{\text{val,i}}\right|}$$

where $\text{Median}\left[\text{P}_{\text{val,i}} \mid \mathbf{z}_{\text{val,i}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}\right] = \exp\left(\mathbf{z}_{\hat{\gamma},\text{val,i}}^{'}\hat{\boldsymbol{\mu}}_{\hat{\gamma}}\right)$ is the median of the posterior distribution of the transaction prices in all the compared procedures with the sole exception of the Bayesian *Mixture* procedure, in which this median is calculated by simulation of the predictive distribution of $\text{y}_{\text{val,i}} \mid \mathbf{z}_{\text{val,i}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}$ [4].

5. Empirical Coverage of the 100(1-α) predictive interval (COVY)

$$\text{COVY}(1\text{-}\alpha) = \frac{1}{\text{nval}} \sum_{\text{i=1}}^{\text{nval}} \text{I}_{\left[\hat{\text{y}}_{\text{val,i},\frac{1-\alpha}{2}}, \hat{\text{y}}_{\text{val,i,I}-\frac{1-\alpha}{2}}\right]}\left(\text{y}_{\text{val,i}}\right) \quad 0<\alpha<0.5$$

is the empirical coverage of the 100(1-α)% predictive interval of the elements of $\mathbf{y}_{\text{val}}$, where $\text{I}_{\text{A}}(\text{y})$ denotes the indicator function of a set A and $\hat{\text{y}}_{\text{val,i},\alpha}$ is the α-quantile of $\text{y}_{\text{val,i}} \mid \mathbf{z}_{\text{val,i}}, \left(\mathbf{y}_{\text{est}}, \mathbf{Z}_{\text{est}}\right), \hat{\mathbf{M}}$ and $\alpha = 0.01$ or 0.05.

[4] The number of simulations was 1,000

6. Gneiting and Raftery Loss of the 100(1-α) predictive interval (LOSS R-G)

$$\text{LOSS R-G}(1\text{-}\alpha) = \frac{1}{\text{nval}}\left(\sum_{i=1}^{\text{nval}}\left(\hat{y}_{\text{val},i,1-\frac{1-\alpha}{2}} - \hat{y}_{\text{val},i,\frac{1-\alpha}{2}}\right) + \right.$$

$$\left. + \frac{2}{\alpha}\left(\sum_{i=1}^{\text{nval}}\left(\hat{y}_{\text{nval},i,\frac{1-\alpha}{2}} - y_{\text{nval},i}\right) I_{\left(-\infty,\hat{y}_{\text{nval},i,\frac{1-\alpha}{2}}\right]}\left(y_{\text{nval},i}\right) + \sum_{i=1}^{\text{nval}}\left(y_{\text{nval},i} - \hat{y}_{\text{nval},i,1-\frac{1-\alpha}{2}}\right) I_{\left(\hat{y}^{(s)}_{\text{nval},i,1-\frac{1-\alpha}{2}},\infty\right]}\left(y_{\text{nval},i}\right)\right)\right)$$

and $\alpha = 0.01, 0.05$.

The PMAD, RMSE and MAD criteria assess the point predictive performance of the model selected using the housing prices (PMAD) and the logarithm of these prices (RMSE and MAD). Likewise, the COVY(0.95) and COVY(0.99) criteria evaluate the performance of predictive intervals of 95% and 99% in terms of empirical coverages, while LOSS R-G(0.95) and LOSS R-G(0.99), which denote the proper losses proposed by Gneiting and Raftery (2007), evaluate the behavior of these intervals in terms of their length and the size of the coverage errors. Finally, the LPRED criterion evaluates the goodness of fit to the data of the model selected in terms of the predictive density.

TABLES

Table 1: Characteristics of the dwellings

	Price (thousands of €)	Living area (m^2)	Building age (years)
Mean	110.09	79.23	37.81
Deviation	7.27	39.28	20.07
Median	90.22	68.03	42.52
Minimum	25.85	36.97	1.42
Maximum	514.01	361.41	113.3
Skewness	1.69	2.98	-0.023
Kurtosis	3.56	12.5	0.04

Table 2: Correlations between the estimations of the coefficients $\{\beta_k(u_i,v_i); i = 1...,N; k = 0,1,2\}$ obtained by each procedure

Intercept	Lasso	Median	Most Probable	Mixture	Ridge	Stepwise
Lasso	1.000	0.966	0.922	0.986	0.995	0.920
Median		1.000	0.910	0.988	0.968	0.907
Most Probable			1.000	0.935	0.897	0.999
Mixture				1.000	0.984	0.933
Ridge					1.000	0.895
Stepwise						1.000

Living Area	Lasso	Median	Most Probable	Mixture	Ridge	Stepwise
Lasso	1.000	0.966	0.922	0.986	0.995	0.920
Median		1.000	0.910	0.988	0.968	0.907
Most Probable			1.000	0.935	0.897	0.999
Mixture				1.000	0.984	0.933
Ridge					1.000	0.895
Stepwise						1.000

Building Age	Lasso	Median	Most Probable	Mixture	Ridge	Stepwise
Lasso	1.000	0.922	0.715	0.944	0.986	0.716
Median		1.000	0.747	0.967	0.917	0.749
Most Probable			1.000	0.812	0.655	0.999
Mixture				1.000	0.937	0.814
Ridge					1.000	0.657
Stepwise						1.000

Table 3: Intra-sample predictive analysis of the estimation and model selection procedures (the best performance for each criterion is in bold)

	Procedures						
Criteria	**Constant**	**Stepwise**	**Ridge**	**Lasso**	**Most Probable**	**Median**	**Mixture**
RMSE	0.377	0.301	0.293	**0.292**	0.301	0.300	0.293
MAD	0.302	0.234	0.227	**0.225**	0.234	0.232	0.227
LPRED	-563.05	-290.61	-295.80	-293.61	-290.61	-288.38	**-269.89**
COV95	97.08	97.32	97.87	97.63	97.32	97.32	97.48
COV99	99.21	99.05	99.29	99.37	99.05	99.05	99.13
PMAD	0.312	0.240	0.233	**0.230**	0.240	0.238	0.233
LOSS G-R(0.95)	1.744	1.493	1.490	1.477	1.496	1.499	**1.461**
LOSS G-R(0.99)	2.208	1.935	1.952	1.919	1.941	1.910	**1.902**

Table 4: Out-of-sample predictive analysis of the estimation and model selection procedures (the best performance for each criterion is in bold)

	Procedures						
Criteria	**Constant**	**Stepwise**	**Ridge**	**Lasso**	**Most Probable**	**Median**	**Mixture**
RMSE	0.375	0.339	0.330	**0.327**	0.332	0.332	0.332
MAD	0.302	0.260	0.252	**0.253**	0.253	0.253	0.253
LPRED	-117.23	-89.18	-103.20	-98.45	-81.14	-81.22	**-80.69**
COV95	97.39	96.27	94.59	94.78	96.64	96.64	96.64
COV99	99.25	98.51	98.51	98.51	98.88	98.88	98.88
PMAD	0.312	0.274	0.265	0.265	0.264	0.264	**0.263**
LOSS G-R(0.95)	1.730	1.723	1.622	**1.599**	1.679	1.681	1.673
LOSS G-R(0.99)	2.522	2.442	2.248	**2.195**	2.356	2.352	2.362

FIGURES

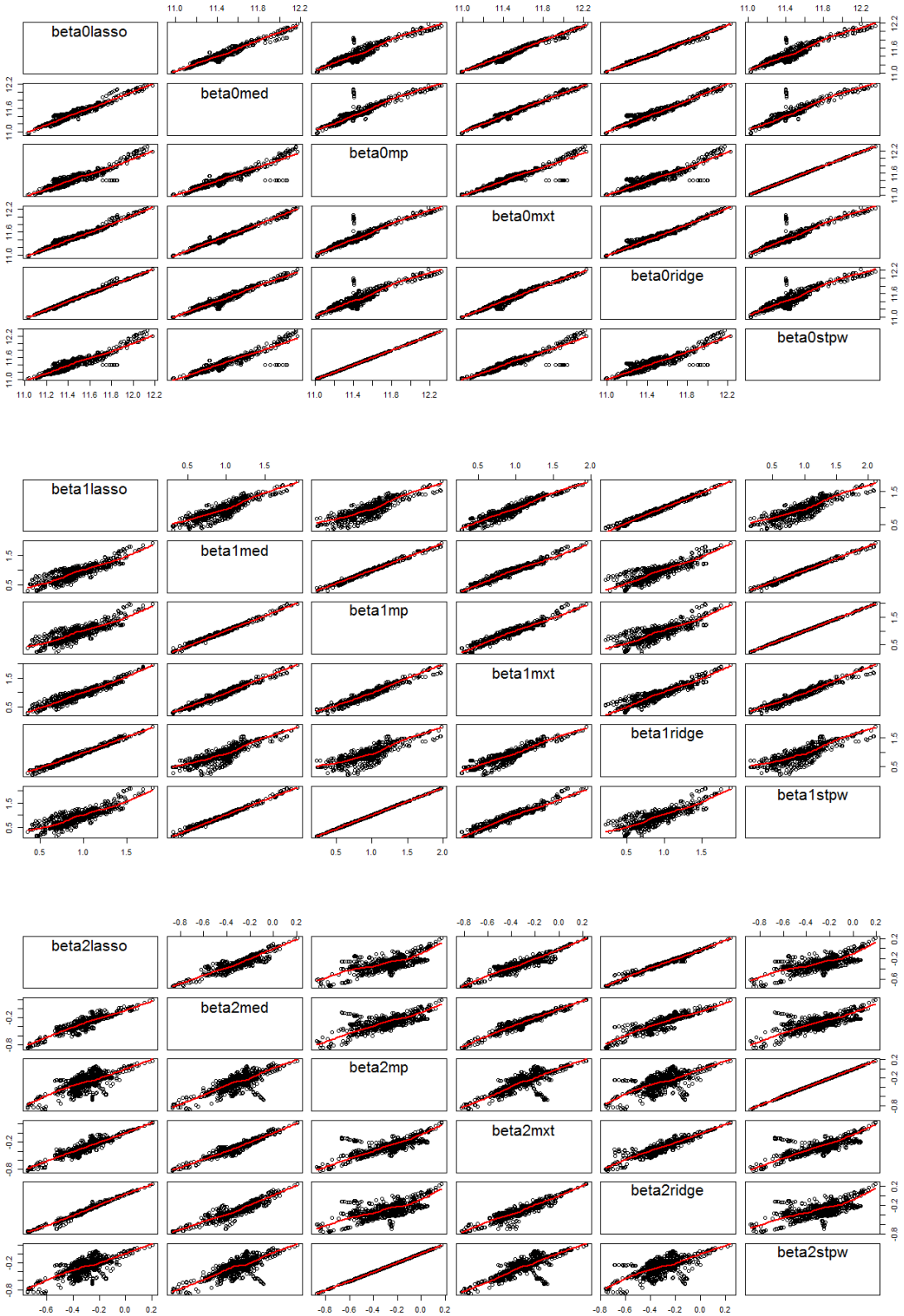

Figure 1: Scatter plot matrix of the regression coefficients together with a non-parametric regression line estimated for the *Lasso, Median, Most Probable, Mixture, Ridge* and *Stepwise* procedures

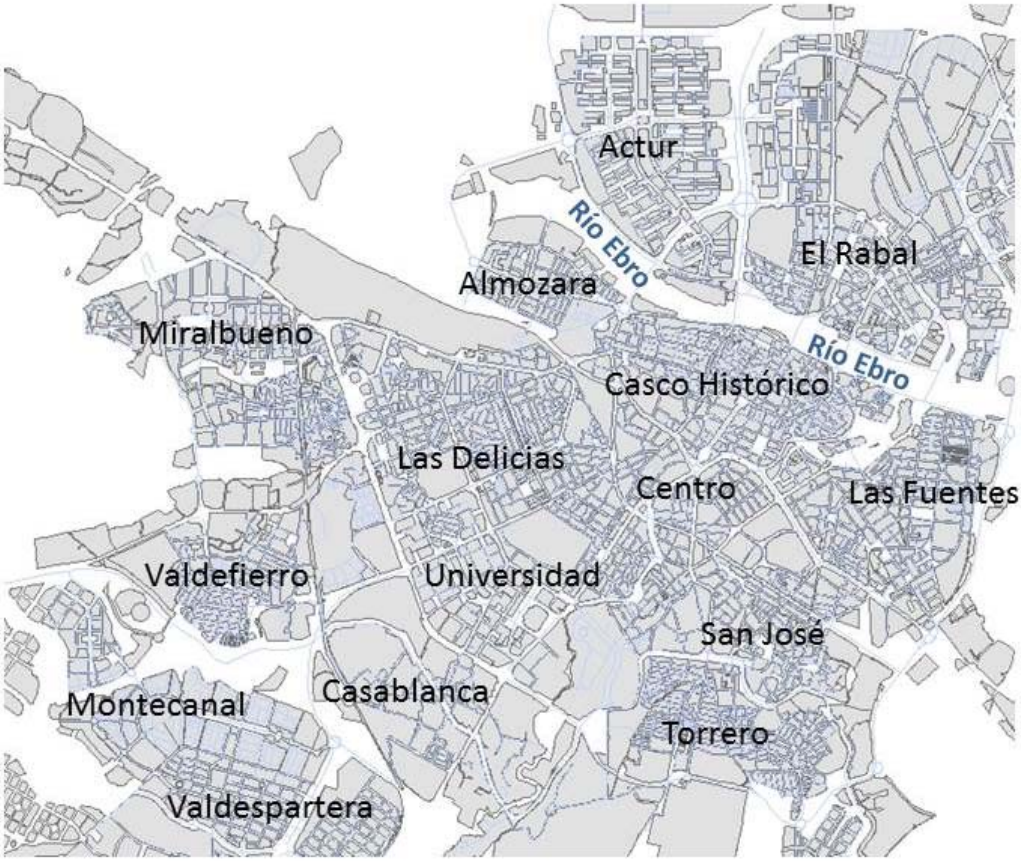

Figure 2: Map of Zaragoza with superimposed neighbourhoods

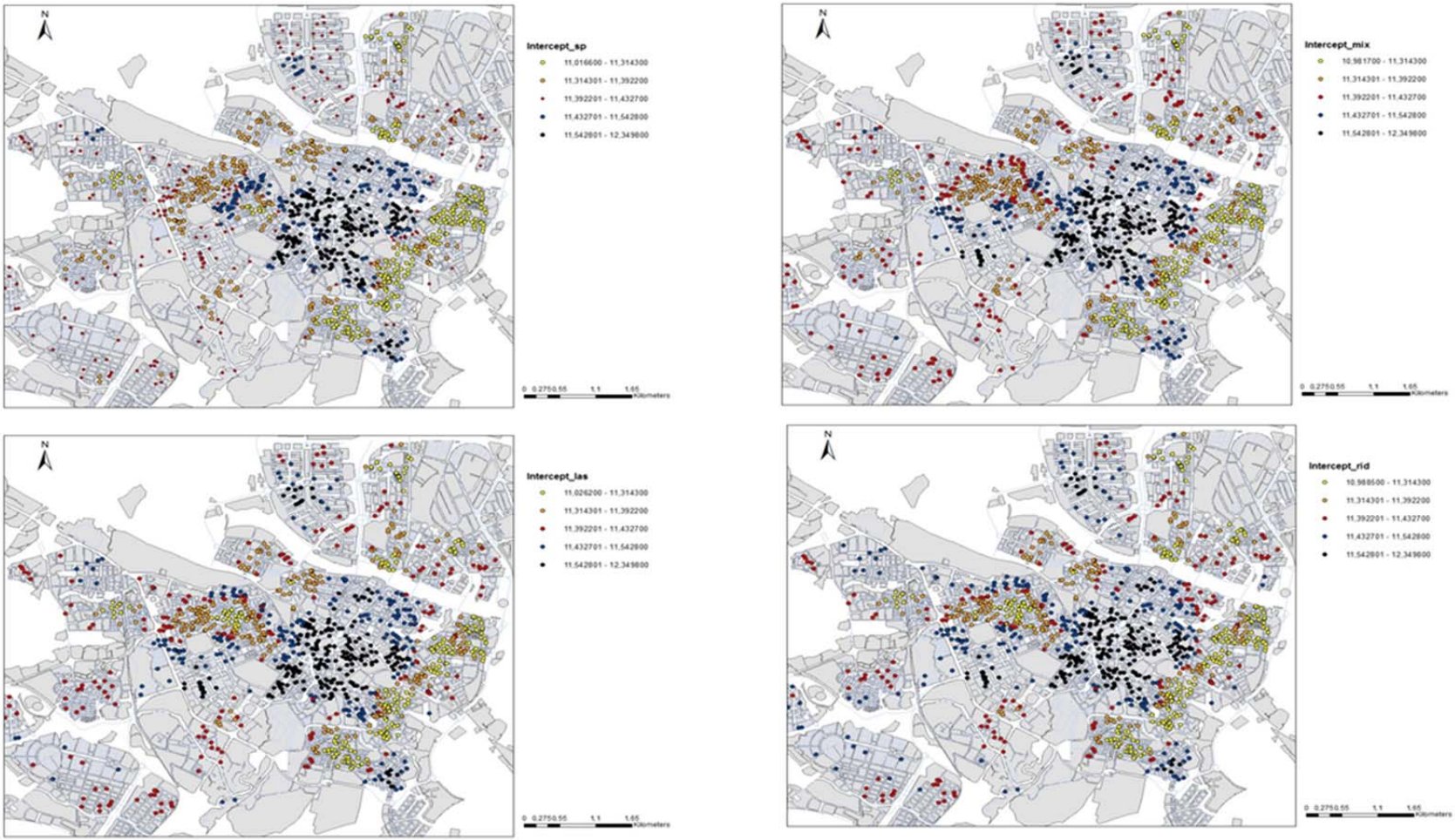

Figure 3: Estimations of the intercept (β_0)
(left to right, top to bottom: *Stepwise*, *Mixture*, *Lasso* and *Ridge* procedures)

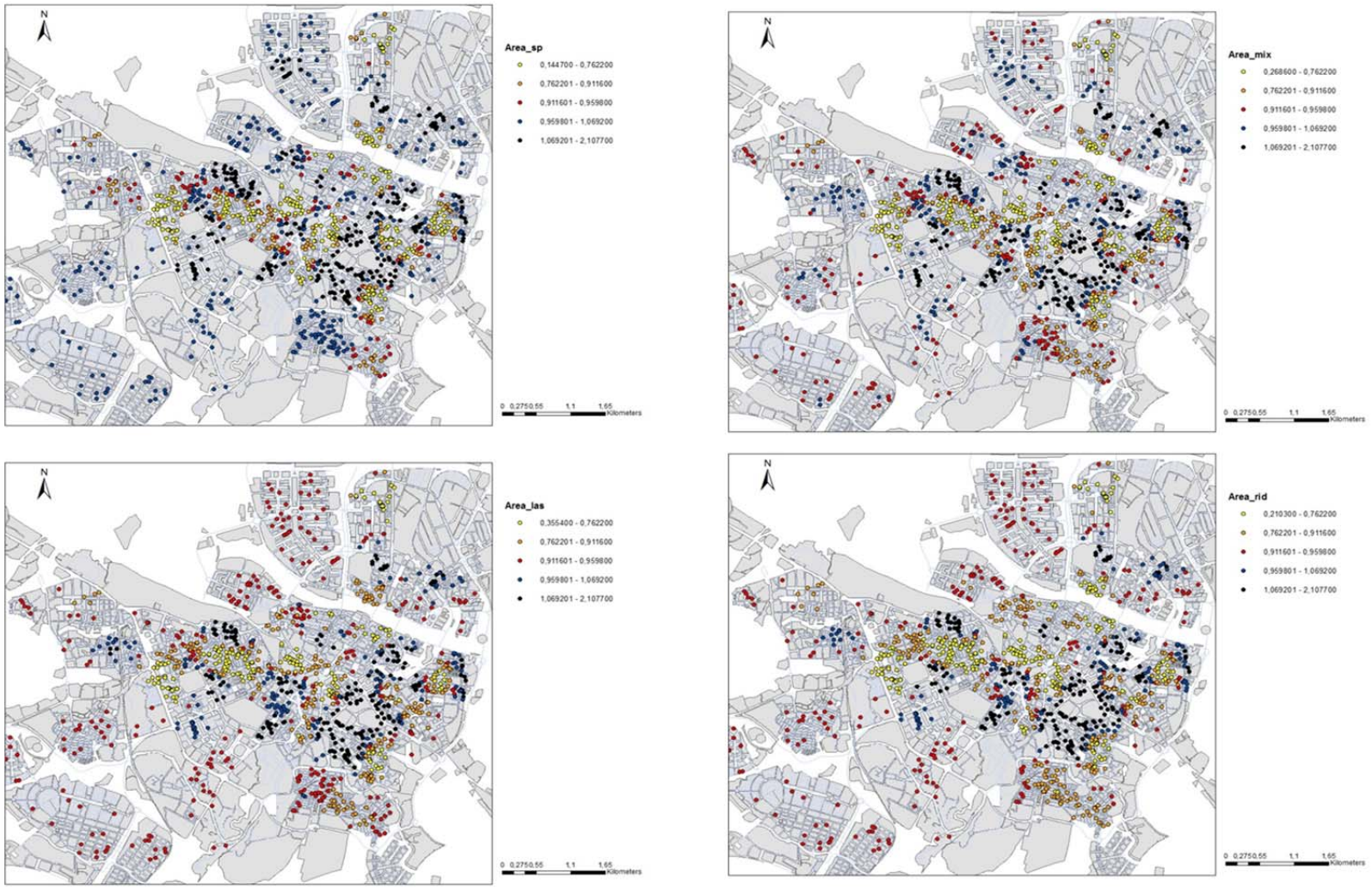

Figure 4: Estimations of the coefficients of the living area (β_1)
(left to right, top to bottom: *Stepwise*, *Mixture*, *Lasso* and *Ridge* procedures)

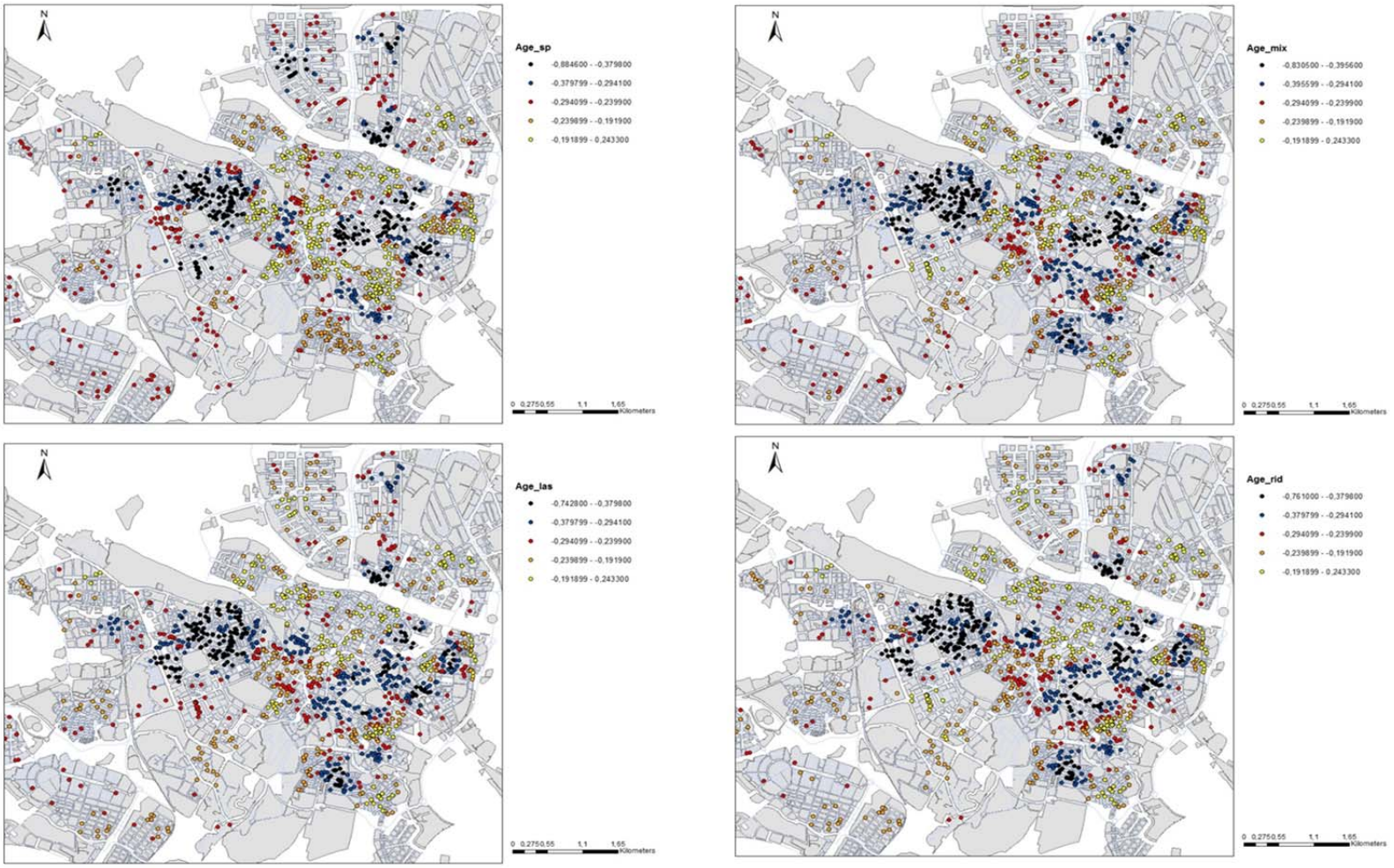

Figure 5: Estimations of the coefficients of the building age (β_2)
(left to right, top to bottom: *Stepwise*, *Mixture*, *Lasso* and *Ridge* procedures)

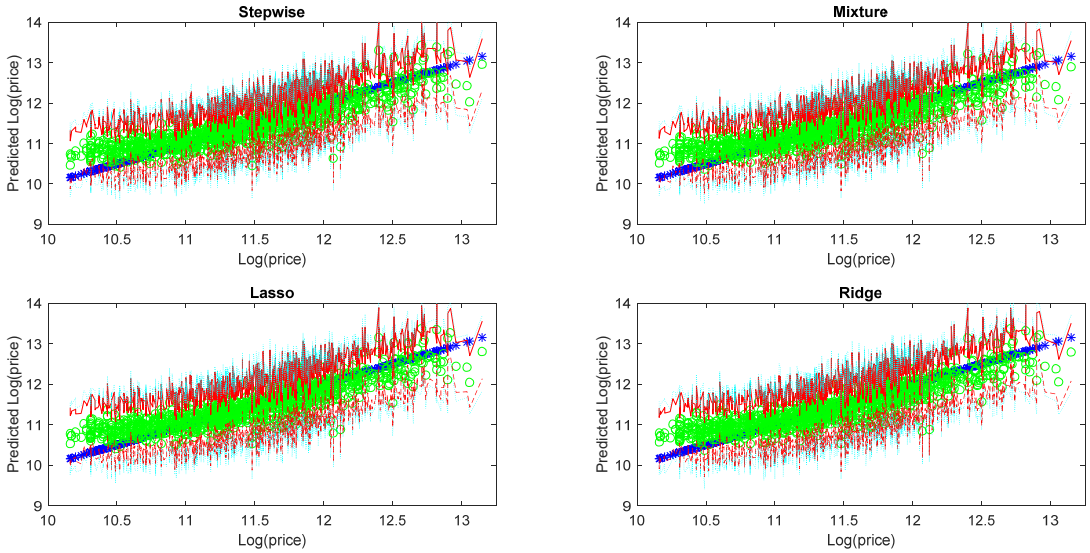

Figure 6: In-sample predictions of the logarithm of the price together with the 95% and 99% limits of the Confidence/Bayesian credible intervals obtained by the *Stepwise*, *Mixture*, *Lasso* and *Ridge* procedures (in the blue diagonal line)

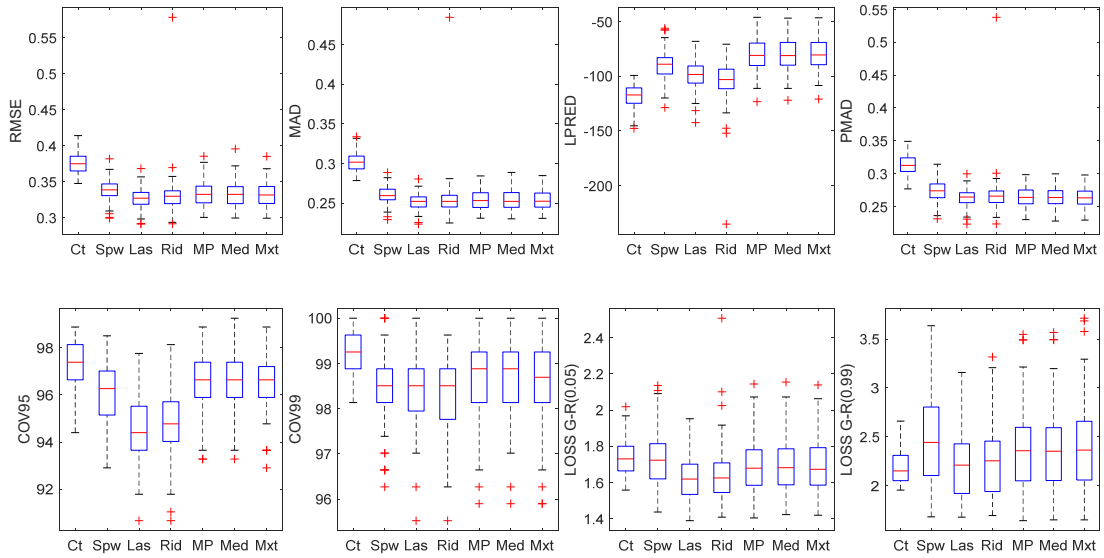

Figure 7: Boxplot of the out-of-sample values of the predictive criteria for the *Constant* (Ct), *Stepwise* (Spw), *Lasso* (Lass), *Ridge* (Rid), *Most Probable* (MP), *Median* (Med) and *Mixture* (Mxt) procedures