ISSN 1415-4765

## TEXTO PARA DISCUSSÃO Nº 672

# ESTIMATION OF A WEIGHTS MATRIX FOR DETERMINING SPATIAL EFFECTS

Elcyon Caiado Rocha Lima*
Paulo Brígido Rocha Macedo**

Rio de Janeiro, setembro de 1999

* Diretoria de Estudos e Políticas Macroeconômicas do IPEA and USU.
** Cedeplar/Face/UFMG.

*O IPEA é uma fundação pública vinculado ao Ministério do Planejamento, Orçamento e Gestão cujas finalidades são: auxiliar o ministro na elaboração e no acompanhamento da política econômica; e prover atividades de pesquisa econômica aplicada nas áreas fiscal, financeira, externa e de desenvolvimento setorial.*

**TEXTO PARA DISCUSSÃO** tem o objetivo de divulgar resultados de estudos desenvolvidos direta ou indiretamente pelo IPEA, bem como trabalhos considerados de relevância para disseminação pelo Instituto, para informar profissionais especializados e colher sugestões.

**ISSN 1415-4765**

## SERVIÇO EDITORIAL

**Rio de Janeiro – RJ**
Av. Presidente Antônio Carlos, 51 – 14º andar – CEP 20020-010
Telefax: (021) 220-5533
E-mail: editrj@ipea.gov.br

**Brasília – DF**
SBS Q. 1 Bl. J, Ed. BNDES – 10º andar – CEP 70076-900
Telefax: (061) 315-5314
E-mail: editbsb@ipea.gov.br

# SUMÁRIO

# RESUMO

A existência de efeitos de "transbordamento", como o impacto do preço de uma unidade residencial no preço de seus vizinhos adjacentes, caracteriza a chamada "dependência espacial". Uma forma de se levar em conta a dependência espacial é especificar modelos de defasagem espacial nos quais se supõe que uma variável espacialmente defasada explica, pelo menos parcialmente, a variação da variável dependente original. A maioria dos estudos fixa *a priori* os parâmetros utilizados na construção da matriz de pesos espaciais que serve de operador da defasagem espacial.

Em contraste, este trabalho não pressupõe qualquer valor *a priori* para os parâmetros da matriz de pesos espaciais na estimação de efeitos de transbordamento. Nós adotamos uma abordagem de máxima verossimilhança clássica e um procedimento bayesiano, *Sampling–Importance–Resampling (SIR)*, para estimar os pesos da matriz e a significância da dependência espacial. Utilizamos dados de unidades residenciais da cidade de Belo Horizonte, e comparamos os resultados obtidos com o procedimento desenvolvido com aqueles derivados a partir da fixação *a priori* dos pesos espaciais. A análise mostra que a função de verossimilhança tem um pico bem definido, e o parâmetro de decaimento estimado é bastante diverso dos valores prefixados usualmente adotados na literatura empírica, como o decaimento "tudo-ou-nada" dentro da distância crítica ou o uso do "inverso da distância".

# ABSTRACT

Spatial dependence results from the existence of spillover effects such as the impact of the price of one housing unit on the price of its adjacent neighbors. One way to account for spatial dependence is to specify spatial lag models in which a spatially lagged variable is assumed to play a role in explaining the variation of the original dependent variable. Most studies use *a priori* non-sample information in the construction of the spatial weights matrix which serves as a spatial lag operator.

In contrast, this study assumes no *a priori* value for the spatial weights matrix in the estimation of spillover effects. We adopt a classical maximum likelihood approach and also a Bayesian Sampling-Importance-Resampling (SIR) procedure to estimate the weights matrix and the significance of spatial dependence. We apply the two estimation procedures to data on housing prices in the city of Belo Horizonte, Brazil, and compare the results obtained with these two techniques with the one derived by *a priori* fixing the weights. The analysis shows that the likelihood function of the weights matrix parameters has a well-defined peak, and the estimated distance-decay parameter is quite different from the standard *a priori* assumptions such as the "all-or-nothing" decay within the cut-off distance or the "inverse distance" adopted in the empirical literature.

## 1 - INTRODUCTION

Spatial dependence results from the existence of spillover effects such as the impact of the price of one housing unit on the price of its adjacent neighbors. One way to account for spatial dependence is to specify spatial lag models in which a spatially lagged variable is assumed to play a role in explaining the variation of the original dependent variable. Most studies use *a priori* non-sample information in the construction of the spatial weights matrix which serves as a spatial lag operator.

In contrast, this study assumes no *a priori* value for the spatial weights matrix in the estimation of spillover effects. We adopt a classical maximum likelihood approach and also a Bayesian Sampling-Importance-Resampling (SIR) procedure to estimate the weights matrix and the significance of spatial dependence. We apply the two estimation procedures to data on housing prices in the city of Belo Horizonte, Brazil, and compare the results obtained with these two techniques with the one derived by *a priori* fixing the weights.

The main results are: the estimated distance-decay parameter is quite different from the standard *a priori* assumptions such as the “all-or-nothing/no decay within the cut-off distance” or the “inverse distance” adopted in the empirical literature (fractionary value instead of the integer value usually used); the likelihood function of the weights matrix parameters has a well-defined peak; the Bayesian procedure allows for the introduction of *a priori* information on the range of parameters and assumes a flat prior leading to a posterior distribution not significantly different from the likelihood.

This paper is organized as follows. Section 2 reviews the analytical issue of connectivity in space. Section 3 describes the data and Section 4 presents the methodology of joint estimation of both the “parameterized” weights matrix and the spatial lag coefficient. Section 5 discusses the classical maximum likelihood estimation; Section 6 presents the Bayesian approach; and Section 7 develops the application of SIR and presents its results.

## 2 - CONNECTIVITY IN SPACE

The study of the spatial pattern of geographically identifiable phenomena has been subject to increasing interest in the social sciences since the early 1970s. Special statistical methods were first developed for geography, and then expanded to other social sciences including economics, to examine whether the presence of a phenomenon in one area or location makes its presence in a neighboring area more or less likely. If the likelihood changes with proximity, the phenomenon is said to exhibit spatial autocorrelation.

As Anselin (1988) points out, both spatial autocorrelation and spatial heterogeneity are specific spatial aspects of data in regional science to which standard econometric methods do not apply: "At first sight spatial autocorrelation may seem similar to the more familiar time-wise dependence encountered in econometric tests for serial correlation but standard econometric results do not carry over in a straightforward way to spatial dependence in cross-sectional samples. This is primarily a result of the multidirectional nature of dependence in space." Regarding spatial heterogeneity, although the lack of structural stability of behavioral relationships over space can be solved in many instances by standard econometric techniques such as random coefficients regressions, there are situations in which those methods are not applicable. For example, the presence of spatial dependence in the error structure requires that the interaction between the spatial units must be taken into account.

The first formal treatment of spatial autocorrelation was by Moran (1948) with the introduction of the idea of binary conntiguity. The underlying structure is defined by 0-1 values, with the value 1 assigned to spatial units having a common border (in the case of spatial areal units), or within a critical cut-off distance (in the case of point pattern spatial units). Cliff and Ord (1973) present a more general approach to express the interaction between two spatial areal units by using a combination of distance measures (inverse distance, or negative exponentials of distance) and a measure of the length of their common border. The formal expression is as follows:

$$w_{ij} = [d_{ij}]^{-\lambda}.[\beta_{ij}]^{\delta}, \tag{1}$$

where $d_{ij}$ stands for the distance between spatial unit $i$ and $j$, $\beta_{ij}$ denotes the proportion of the interior boundary of unit $i$ in contact with unit $j$, and $\lambda$ and $\delta$ are paramaters. One distinctive feature of Cliff and Ord's approach, as opposed to Moran's binary contiguity, is the assymmetry of the resulting weights in the fomer case. Spatial areal units such as counties are typically suited to have their interaction expressed by expression (1): both the distance between their centers and the relative importance of their common border are taken into account. Within the context of Cliff and Ord's approach, the notion of contiguity of spatial units having a a point pattern geographic distribution (such as cities in a urban hierarchy or housing units in a city) is related only to the distance between any two of them.[1]

The existence of spatial association is represented by relating a variable to its spatially lagged counterpart. This relationship is constructed as a linear combination of the observations in the system. The econometric interpretation is straightforward: the linear spatial association is actually a special case of a system of a simultaneous linear equations problem. Each equation in the system is expressed as:

$$y_i = \sum_j \beta_{ij} y_j \,, \ \forall\ i, \tag{2}$$

The identification of the model parameters requires the imposition of at least some constraints. The introduction of a spatially lagged variable is a typical approach

1 Another way of analyzing contiguity is to construct a map of polygons from the original point pattern spatial units. This would allow the ususal more general weights matrix to be implemented.

adopted in spatial analysis to reduce the estimation problem to that of determining one “representative” parameter ρ:

$$y_i = \rho\Sigma_j w_{ij} y_j \tag{3}$$

where $w_{ij}$ as in equation (1) is a measure of the spatial association between the units *i* and *j* in the system.

## 3 - DATA

The database analyzed has price and characteristic information for a sample of Belo Horizonte residential apartments lying within a spatial region of approximately 16 square kilometers. The apartments were included in a market survey of residential prices conducted for the Belo Horizonte municipal government in October 1995 by the Instituto de Pesquisas Econômicas e Administrativas (IPEAD) of the Universidade Federal de Minas Gerais. The apartments’ characteristics were drawn from the city’s property tax data files which include variables such as apartment area (square meters), age, availability of garage space, local topography, and the level of public services such as piped water, electricity, and garbage collection. Topography is fairly homogeneous for the region studied, with a uniform index assigned to all apartments by city tax assessors, and this characteristic does not affect their relative market value according to realtors. The region is also well-provided with city services, and there is a homogeneous overall index of their availability. For this study, therefore, the sources of price variation, are the area of the housing unit in square meters, its age, and the availability of a garage space. The average distance between any two housing units in the sample is nearly 2.5 km, and the maximum distance is 6.4 km.

## 4 - THE GENERAL SPATIAL MODEL AND THE LIKELIHOOD FUNCTION

$$Y = \rho\, W\, Y + X\beta + \varepsilon, \tag{4}$$

$$\varepsilon = \delta\, W\, \varepsilon + \mu$$

$$\varepsilon \sim N(0,\ \sigma^2 I)$$

$$\mu \sim \mathrm{N}(0, \Omega),$$

where β is a *k* x 1 vector of parameters associated with exogenous (not lagged dependent) variables *X*, which is an *n* x *k* matrix, ρ is the coefficient of the spatially lagged dependent variable, and λ is the coefficient in a spatial autoregressive struture for the disturbance ε.

The spatial weights matrix $W$ has entries depending on the distance between the spatial units and a distance-decay parameter. Each element $w_{ij}$ of the matrix $W$ is set as follows:

$w_{ij} = 1/(d_{ij})^{\lambda}$,  if $i \neq j$ and $d_{ij}$ ó $\tau$,

$w_{ij} = 0$,  if $i \neq j$ and $d_{ij} > \tau$, or if $i = j$.

where $d_{ij}$ = Euclidian distance between spatial units $i$ and $j$; $\lambda$ = distance-decay parameter;

$\tau$ = critical cut-off distance parameter (its value is set to be lesser than the highest value of distance between any two units observed in the sample).

Let [2]: $A = I - \rho W$

$$B = I - \delta W$$

Then the model above can be represented by

$$AY = X\beta + \varepsilon$$

$$B\varepsilon = \mu$$

$$E(\mu\mu') = \Omega$$

Define $v = \Omega^{-1/2}\mu$ (a homocedastic random disturbance), then

$$AY = X\beta + B^{-1}\Omega^{1/2}v$$

Therefore,

$$v = \Omega^{-1/2}B\,(AY - X\beta) = f\,(Y, X, \theta).$$

Since $v$ cannot be observed, the likelihood function has to be based on $Y$. The Jacobian of this transformation:

$$J = \det\,(\partial v/\partial Y) = |\Omega^{-1/2}B\,A|$$

*The Range of ρ as a function of λ (the distance decay parameter) in the Spatial Autoregressive Model*

2 The development below follows Anselin (1988).

The specification test in Table 2 of Section 5 has strong evidence supporting the non-rejection of hypothesis $B = I$. Therefore, the development below will focus on a Jacobian which includes only matrix $A$.

$$A = I - \rho\, W$$

$$|A| > 0$$

$$|A| = -\rho^{n}\, |\, W - (1/\rho)\, \mathrm{I}\, | = -\rho^{n}\ |W^{*}|$$

$$\text{where}\ \ W^{*} = W - (1/\rho)\, \mathrm{I}$$

It is worth to bear in mind that $W$ is an inverse distance matrix and therefore $W$ and $W^*$ are symmetric matrices (and they have real eigenvalues). The relationship between the eigenvalues of both matrix is:

$$1/\rho + w^* = w$$

where:

$w^*$ - eigenvalue of $W^*$ ;
$w$ - eigenvalue of W.

$$|A| = -\rho^{n}\, |\, W - (1/\rho)\, I\, | = -\rho^{n} \prod_{i=1}^{n} w_i^{*} = -\rho^{n} \prod_{i=1}^{n} (w_i - 1/\rho) = \prod_{i=1}^{n} (1 - \rho w_i)$$

If $1 - \rho w_i > 0$, $\forall$ i, then $|A| > 0$. To learn about the range of $\rho$ we have to know what is the range of $w_i$.

In general examples of weights matrices discussed in the spatial econometrics literature refer to systems of areal units for which the notion of nearest neighbor is associated with the share of a common border. In dealing with systems of areal units, many authors work with the so called "standardized weights matrices" which yield to well-behaved partial derivatives of the jacobian of the model w.r.t. the parameters to be estimated. Typically this is done by requiring the summation of entries in the weights matrix, corresponding to spatial units sharing a common border with a given unit, be equal to one. Ord (1975) argues that "... to lend a natural interpretation to $\rho$, the scaling $\Sigma w_{ij} = 1$ may be used for each location, where the sums are over either $i$ or $j$. The scaling implies that $\rho < 1$." Doreian (1980) examines applications of standardized weights matrices in his discussion about the way the search procedure introduced by Ord simplifies maximum likelihood methods applied to spatial models, and Anselin (1988) considers examples on spatial econometric models also refering to systems of areal units for which the weights matrices are normalized according to Ord's scaling suggestion.

This work deals with a system of spatial units distributed as points in a urban area in which distance, instead of common border, is the criterion to assess the potential interaction between any two units. Ord's scaling procedure does not apply here because it distorts the interpretation associated with a distance decay process. In order to reduce the range of possible values of the parameter ρ while keeping the metrics of the spatial arrrangement among the units of the system, the procedure described below uses the maximum of sums of absolute values of row elements as the convenient definition of norm to scale down the inverse distance spatial weights matrix *W*.

## *Lemma*

*Let A be an* n *x* n *matrix, and let* $\|A\|$ *be the norm of matrix* A *defined as the maximum of sums of absolute values of row elements. If* δ *is any characteristic root of A then* $|\delta| \leq \|A\|$.

Proof: This is a well know lemma of linear algebra and we skip the proof.

*Proposition 1.*

*Let W be an* n *x* n *symmetric inverse distance weights matrix " normalized" by having each of its elements divided by the maximum of sums of absolute values of row elements, and let* $w_i$ *be any characteristic root of* W. *Then* $|w_i| \leq 1$ *and max* $\{w_i\} > 0$.

Proof: The normalized matrix has $\|W\|=1$ and therefore by the Lemma mentioned above $|w_i| \leq \|W\| = 1$. W has trace equal to zero (all diagonal entries are set to zero by construction) and a non-zero determinant and, therefore, at least one positive and one negative characteristic root.

*Corollary:* $1 > \rho > -1 \Rightarrow |A| > 0$.

Since we have proved that $|w_i| \leq 1$ we can determine with some precision the range of ρ in an identified model. Let $w_{\max} = \max\{w_i\}$, $\rho_{\max} = 1/w_{\max}$, $w_{\min} = \min\{w_i\}$ and $\rho_{\min} = -1/|w_{\min}|$

$0 < w_{\max} \leq 1 \Rightarrow \rho_{\max} \geq 1, \quad 0 > w_{\min} \geq -1 \Rightarrow \rho_{\min} \leq -1$

therefore, $1 > \rho > -1 \Rightarrow \rho_{\min} < \rho < \rho_{\max} \Rightarrow |A| > 0$.

*Proposition 2. The requirement of a cut-off (critical) distance.*

*If there is no critical cut-off distance, the vector of parameters* $(\beta'; \lambda; \rho) = (\sum_{i=1}^{n} y_i$, $0, \ldots 0; 0; -1)$ *(where the first entry of* $\beta$ *is the equation's intercept) gives a* *"perfect fit" to equation* y = ρW(λ) y + Xβ + ε. *Therefore without a cut-off distance there are no degrees of freedom to estimate the equation.*

Proof:

Consider the spatial lag model

$$y = \rho W(\lambda)y + X\beta + \varepsilon \tag{8}$$

If $\rho = -1$ and $\lambda = 0$, it follows that

$$y_i = -y_1 - y_2 \,.....\, - y_{i-1} - y_{i+1} \,....\, - y_n + X_i\beta + \varepsilon_i \tag{9}$$

If $\beta = (\sum_{i=1}^{n} y_i, 0, ...0)$ then $X_i\beta = \sum_{i=1}^{n} y_i$ and (9) can be written as

$$y_i = -y_1 - y_2 \,.....\, - y_{i-1} - y_{i+1} .... - y_n + \sum_{i=1}^{n} y_i + \varepsilon_i \,.$$

Therefore, $(\beta'; \lambda; \rho) = (\sum_{i=1}^{n} y_i, 0, ...0; 0 ; -1)$ in equation (8) implies

$$y_i = y_i + \varepsilon_i \quad \Rightarrow \varepsilon_i = 0 \text{ (the perfect fit).}$$

The optimizing procedure does not deliver this degenerate solution if the inverse distance weights matrix has some off-diagonal entries equal to zero, that means, it has a critical cut-off distance beyond which the assumption of no spatial interaction holds.

The predetermination of a critical cut-off distance amounts to set an *a priori value of* $\lambda$ to an infinitely large value (∞) to units with no spatial interaction between them (zero entries in the matrix). Once that restriction is assumed to hold, the optimizing algorithm proceeds to estimate the *sample based value of* $\lambda$, which describes the decaying (with distance) spatial interaction between any two units within the critical distance, as well as the parameter $\rho$ and the vector of parameters $\beta$.

## 5 - THE CLASSICAL ESTIMATION OF DISTANCE DECAY AND SPATIAL AUTOCORRELATION PARAMETERS (λ AND ρ)

Specification tests, whose results are summarized in Table 2, support the assumptions of a mixed regressive-spatial-autoregressive model with a homocedastic structure of errors. This amounts to have

$\delta = 0$ and $\varepsilon \sim N(0, \sigma^2 I)$ in equation (4). The log-likelihood function of this mixed regressive-spatial-autoregressive model, $L(Y; \rho, \lambda, \sigma, \beta, X)$, is given by:

$$L(Y; \rho, \lambda, \sigma, \beta, X) = -(T/2) \ln \pi - (T/2) \ln \sigma^2 + \ln |A| - \\ - (1/2\,\sigma^2)(AY - X\beta)'(AY - X\beta)$$

where: $A = I - \rho.\, W(\lambda)$ = matrix with known elements if $\rho$ and $\lambda$ are given;
$W(\lambda)$ = matrix of weights that are a function of $\lambda$;
$T$ = number of observations;
$Y$ = vector of observations on the dependent variable ($T$ x 1 vector);
$X = T$ x $K$ matrix of the observations on the "$k$" exogenous variables;

The log-likelihood function presented above can be concentrated with respect to $\beta$ and $\sigma^2$ . The first order conditions for the maximization of the above log-likelihood gives:

$$b = (X'X)^{-1}X'Ay \qquad (*)$$

(where $b$ is the estimated value of $\beta$ given that $\rho$ and $\lambda$ are fixed at it's optimal values)

Let, $b_0 \equiv (X'X)^{-1}X'y$ ; $b_L \equiv (X'X)^{-1}X'Wy$ .

Then, $b = (X'X)^{-1}X'y - \rho(X'X)^{-1}X'Wy = b_0 - \rho b_L$.

Given $\rho$ and $\lambda$ , we can define two set of residuals: $e_0 \equiv y - X\,b_0$, $e_L \equiv Wy - X\,b_L$.

The estimate for the error variance $\sigma^2$ , considering the first order conditions for the maximization of the above log-likelihood — given the two definitions of residuals and the optimal values $\rho$ and $\lambda$ — satisfies the following expression:

$$\sigma^2 = (1/N)\,(e_0 - \rho e_L)'\,(e_0 - \rho e_L) \qquad (**)$$

Therefore (*) and (**) yields to the following concentrated likelihood:

$$L_C = C - (T/2) \ln [(1/N)(e_0 - \rho e_L)'\,(e_0 - \rho e_L)] + \ln | I - \rho\,.\,W(\lambda)\,|,$$

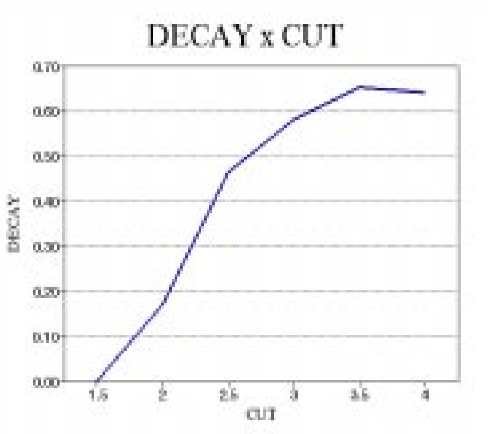
where C is a constant. The expression above is a nonlinear function of two parameters, $\rho$ and $\lambda$, and numerical techniques are applicable. The steps to optimize the likelihood are as follows:

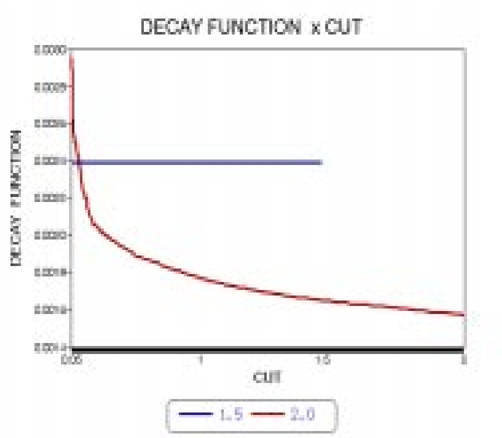
1) set initial values to $\lambda$ — which amounts to set $W(\lambda)$ — and $\rho$ in the appropriate range;
2) maximize the concentrated likelihood with respect to $\rho$ and $\lambda$ using a numeric technique (Powel - using the software GQOPT developed by Quandt );
3) given $\rho$ and $\lambda$, compute $b = b_0 - \rho b_L$ and $\sigma^2 = (1/N)(e_0 - \rho e_L)'\,(e_0 - \rho e_L)$

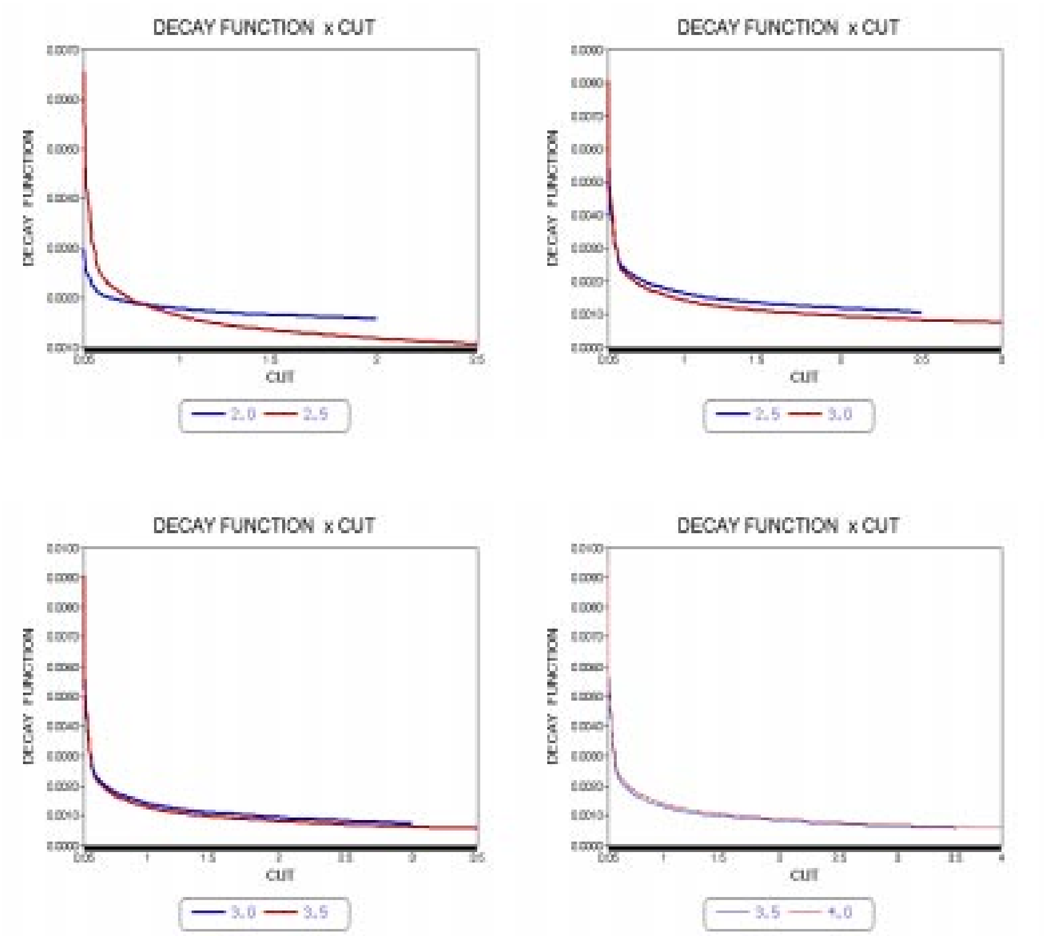
Table 1 presents the results of the estimated models using the log-likelihood function above for different assumptions of the critical cut-off distance (column cut-off distance is given in kilometers). The dependent variable is log(price), the exogenous variables are log(area), log(age), and a dummy variable for garage. The parameters estimated are the ones correspondent to these variables, the decay parameter (λ), and the coefficient correspondent to the spatially lagged dependent variable (ρ).

Table 1
**Estimated Models**
**Dependent Variable = log(Price) of the Residential Unit**

| Cut-off Distance | Stat. | Decay | Spatial Lag Coefficient | Constant | log(Area) | log(Age) | Garage |
|---|---|---|---|---|---|---|---|
| 1.5 | Coeff. | 1.0045E-14 | 0.00239 | 5.97780 | 0.98318 | -0.05922 | 0.19918 |
| | SD | - | (0.00051) | (0.39439) | (0.07369) | (0.03905) | (0.07044) |
| 2.0 | Coeff. | 0.17057 | 0.00177 | 6.06338 | 0.95687 | -0.05498 | 0.16839 |
| | SD | (0.37407) | (0.00041) | (0.40319) | (0.07675) | (0.04047) | (0.07147) |
| 2.5 | Coeff. | 0.46601 | 0.00162 | 5.89751 | 097336 | -0.04019 | 0.15406 |
| | SD | (0.21217) | (0.00035) | (0.39402) | (0.07395) | (0.03956) | (0.07047) |
| 3.0 | Coeff. | 0.58179 | 0.00141 | 5.84228 | 0.98084 | -0.03534 | 0.16440 |
| | SD | (0.19650) | (0.00036) | (0.40998) | (0.07660) | (0.04093) | (0.07269) |
| 3.5 | Coeff. | 0.65413 | 0.00128 | 5.76306 | 0.99572 | -0.03458 | 0.17323 |
| | SD | (0.19997) | (0.00039) | (0.42791) | (0.07808) | (0.04204) | (0.07444) |
| 4.0 | Coeff. | 0.64162 | 0.00138 | 5.66229 | 1.00005 | -0.03283 | 0.17839 |
| | SD | (0.19782) | (0.00046) | (0.44546) | (0.07828) | (0.04221) | (0.07479) |






*Model's specification test*

The general spatial model presented in expression (4) reproduced below represents situations where observations are available for a cross-section of spatial units, and spatial dependence may exist regarding both the dependent variable and the error terms.

$$Y = \rho\, W\, Y + X\beta + \varepsilon, \qquad (4)$$

$$\varepsilon = \delta\, W\, \varepsilon + \mu$$

$$\varepsilon \ \sim N(0,\ \sigma^2\, I)$$

$$\mu \sim N(0,\, \Omega),$$

Manipulation of the expression above leads to the following alternative representation:

$$Y = (\rho + \delta - \rho\delta W)WY + X\beta - \delta WX\beta - + \mu,$$

which can also be written as:

$$Y = (\rho + \delta)WY + X\beta - \delta WX\beta - \rho\delta W^2 Y + \mu \qquad (4')$$

White's test has a combined null hypothesis of correct specification and homocedasticity of the error structure. The spatial lag specification adopted in this work ($\rho \neq 0, \delta = 0$) is nested in expression (4') above:

$$Y = \rho WY + X\beta + \mu;$$

therefore, the non-rejection of the null hypothesis (White's test) for this specification provides evidence of no spatial dependence in the error structure. Table 2 shows the results of both White's and Jarque-Bera's tests for a number of critical cut-off distances. The null hypotheses of homocedasticity/no-misspecification and normality are not rejected in all cases.

Table 2
**White's and Normality Tests**

| CUT | White's Heteroskedasticiy and Specification Test | | | | Jarque - Bera's Normality Test | |
|---|---|---|---|---|---|---|
| | F(8,44) | | Qui-Squared(8) | | Qui-Squared(2) | |
| | Statistic | P-Value | Statistic | P-Value | Statistic | P-Value |
| 1,5 | 0,170771 | 0,9937310 | 1,596057 | 0,990996 | 0,500389 | 0,778649 |
| 2,0 | 0,530738 | 0,8269010 | 4,664293 | 0,792780 | 0,901241 | 0,637232 |
| 2,5 | 0,846079 | 0,5679520 | 7,066122 | 0,529516 | 0,575469 | 0,749961 |
| 3,0 | 0,955095 | 0,4826540 | 7,841871 | 0,449067 | 0,636687 | 0,727353 |
| 3,5 | 0,745856 | 0,6510200 | 6,329057 | 0,610424 | 0,643670 | 0,724818 |
| 4,0 | 0,755071 | 0,6432780 | 6,397811 | 0,602763 | 0,748944 | 0,687652 |

The justification for the standard procedure of pre-setting the weights matrix (that means, for a given specification have its parameter(s) predetermined) comes from the assumption that the way it describes the connectivity among the spatial units in the system is known a priori. In terms of the estimation procedure, this amounts to have more degrees of freedom as compared to an analysys which has both the model and the spatial structure determined by the data. In contrast, the advantage of the latter approach is not having the validity of the estimates depending on the extent to which the spatial structure is correctly reflected in the weights. The estimated value(s) of the weights matrix parameter(s) may as well convey relevant information about the spatial process being analyzed.

The distance-based weights matrix used here depends only on the decay parameter $\lambda$ ($0 \leq \lambda < \infty$) which affects the estimated spatial model (for a critical cut-off distance equals to 2.5 km) in the way illustrated by Table 3 below. The estimated decay parameter $\lambda = 0.46601$ is nearly half a way between the "all-or-nothing" predefined value $\lambda = 0$ (every neighbor within the cut-off distance is equally important) and the inverse distance predefined value $\lambda = 1$ (the relative importance of neighbors within the cut-off distance is inversely proportional to it).

Table 3
**Predefined Versus Estimated Decay Parameter**
**Dependent Variable = log(Price) of the Residential Unit**

| Cut-off Distance | Stat. | Decay | Spatial Lag Coeff. | Constant | log(Area) | log(Age) | Garage |
|---|---|---|---|---|---|---|---|
| 2.5 km | Coeff. | 0 (Predefined) | 0.00143 | 5.90702 | 0.98236 | -0.04929 | 0.14192 |
| | SD | - | (0.00034) | (0.40464) | (0.07576) | (0.03998) | (0.07218) |
| 2.5 km | Coeff. | 0.46601 | 0.00162 | 5.89751 | 0.97336 | -0.04019 | 0.15406 |
| | SD | (0.21217) | (0.00035) | (0.39402) | (0.07395) | (0.03956) | (0.07047) |
| 2.5 km | Coeff. | 1 (Predefined) | 0.00084 | 5.94571 | 0.99538 | -0.02765 | 0.16934 |
| | SD | - | (0.00033) | (0.41991) | (0,07840) | (0.04214) | 0.07462 |
| 2.5 km | Coeff. | 2 (Predefined) | 0.00004 | 6.02088 | 1.02814 | -0.03476 | 0.16907 |
| | SD | - | (0.00002) | (0.45313) | (0.08384) | (0.04593) | (0.08056) |

*Generating the empirical distribution of parameters under the null adopted hypothesis by using bootstrap*

The generation of artificial data sets based on the parameter estimates obtained with the original sample allows for a better understanding of the sampling distribution of the maximum likelihood estimator (MLE) used here. This resampling method, known as bootstrap, uses frequently the assumption of independently and identically distributed errors regarding the stochastic component of the presumed data-generating process. If this stochastic component is assumed to have a known distribution, random numbers drawn from that distribution allow for the generation of a set of "new samples."

Assume that the data-generating process is modeled after equation (4) reproduced below:

$$Y = \rho WY + X\beta + \varepsilon,$$

$$\varepsilon \sim N\,(0, \sigma^2 I)$$

where $\theta' = [\beta, \rho, \lambda, \sigma]'$ a $(k + 3)$ x 1 is the vector of parameters in which $\beta$ is a $k$ x 1 vector of the coefficients of the exogenous variables, $\rho$, $\lambda$, and $\sigma$ are scalars representing, respectively, the coefficient of the spatially lagged dependent variable, the distance decay parameter in the spatial matrix weights, and the variance of the structure of errors. The distance decay parameter affects each element $w_{ij}$ of $W$ according to the relationship $w_{ij} = 1/(d_{ij})^{\lambda}$, where $d_{ij}$ is the distance between any two units i and j in the system.

The bootstrap approximates the distribution of $\hat{\theta} - \theta$, where $\hat{\theta}$ stands for the estimates of the vector of parameters $[\beta, \rho, \lambda, \sigma]'$, by an empirical distribution derived from the data. First, estimates of these parameters are obtained from the original data set by using the estimator (MLE in this case) whose sampling properties are the objective of analysis. Next, estimates of the unobservable errors $\varepsilon$ are generated drawing $T$ times ($T$ = number of observations) with replacement from the normalized residuals. The non-rejection of the null hypothesis of the normality of the residuals (Table 2) justifies the choice of the normal distribution as the "urn" from which the random numbers are drawn. This set of "drawn errors" generates a set of pseudo data $Y^*$ which is used then to estimate a new set of parameters $\hat{\theta}$.

The procedure above is replicated to simulate 289 additional data sets generated using MLE to estimate the parameters for each of those samples. Table 4 reports the results of the bootstrap simulation for the parameters ρ and λ: they are very close the values adopted as the null hypothesis (λ = 0.466, ρ = 0.0162, standard deviations of 0.21217 and 0.00035, respectively).

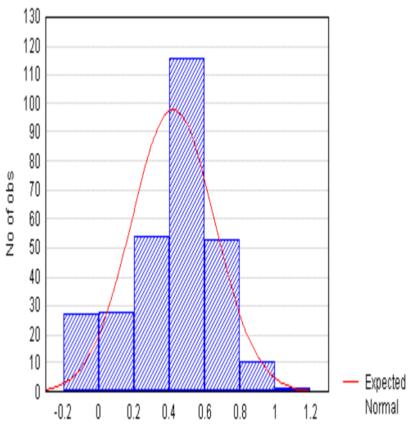
Table 4
**Bootstrap Simulated Results**

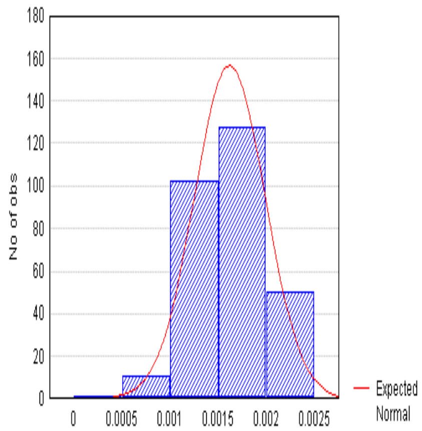
(critical cut-off distance in km = 2.5)

| Parameter | λ | ρ |
|---|---|---|
| mean | 0.4206864 | 0.0016167 |
| Standard Deviation | 0.2347235 | 0.0003672 |

*The null hypotheses about the values are: 0,466 (λ), 0,0162 (ρ), 0,21217(sd λ) e 0,00035 (sd ρ). Simulation number = 289*

Figure 1
Histograms With Simulated Results



## 6 - THE CONCENTRATED LIKELIHOOD AND THE INTEGRATION WITH RESPECT TO β AND σ WITH A FLAT PRIOR FOR THESE COEFFICIENTS

The likelihood of the mixed regressive-spatial-autoregressive model, $\ell(Y^*; \rho, \lambda, \sigma, \beta, X)$, is given by:

$$\ell(Y^*; \rho, \lambda, \sigma, \beta, X) = (2\pi)^{-T/2} \cdot \sigma^{-T} \cdot |A| \cdot e^{-(1/2\sigma^2)(Y-X\beta)'(Y-X\beta)} \tag{1}$$

where:
$A = I - \rho \cdot W(\lambda)$ = matrix with known elements if ρ and λ are given;
$W(\lambda)$ = matrix of weights that are a function of λ;
$T$ = number of observations;
$Y^*$ = vector of observations on the dependent variable ($T$ x 1 vector);
$Y = AY^* = T$ x 1 vector;
$X = T$ x $K$ matrix of the observations on the "$k$" exogenous variables.

Let:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\sigma}^2 = [(Y - X\hat{\beta})'(Y - X\hat{\beta})] / T$$

Then

$$\ell(Y^*; \rho, \lambda, \sigma, \beta, X) = (2\pi)^{-T/2} \cdot \sigma^{-T} \cdot |A| \cdot e^{-(1/2\hat{\sigma}^2)[T\hat{\sigma}^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]} \tag{2}$$

We assume that our prior p.d.f. for β and σ are independent from other coefficients of the model and that our information about β and σ is diffuse or vague. Given our assumptions we can take the elements of β and log σ to be independently and uniformly distributed [as suggested by Zellner (1971)]; that is the joint prior for β and σ, $P(\beta, \sigma)$, is

$$P(\beta, \sigma) \propto 1/\sigma, \qquad -\infty < \beta_i < \infty \quad i = 1, 2, ..., K; \quad 0 < \sigma < \infty$$

If we multiply $\ell(Y^*; \rho, \lambda, \sigma, \beta, X)$ in (2) by $P(\beta, \sigma)$ we obtain:

$$f(\beta, \sigma / \rho, \lambda, X, Y^*) \propto \sigma^{-T-1} \cdot |A| \cdot e^{-(1/2\hat{\sigma}^2)[T\hat{\sigma}^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]} \tag{3}$$

If we integrate the function above with respect to β, we obtain:

$$g(\sigma / \rho, \lambda, X, Y^*) \propto \sigma^{-(T+1)} \cdot |A| \cdot e^{-(T\hat{\sigma}^2/2\hat{\sigma}^2)}$$

which is in a form of an inverted gamma p.d.f. If we integrate the function above with respect to σ, we obtain:

$$m(Y^*; \rho, \lambda, X) \propto |A| \,.\, 2^{(T-2)/2}\, \Gamma(T/2) \,.\, (\hat{\sigma}^2)^{-T/2} \propto |A| \,.\, (\hat{\sigma}^2)^{-T/2}$$

(where: Γ(.) = gamma function)

or

$$\ln m(Y^*; \rho, \lambda, X) = C + \ln |A| - (T/2) \ln \hat{\sigma}^2 \qquad (4)$$

(where: $C$ = constant)

This last expression is exactly the concentrated likelihood used by Anselin (1988) to estimate ρ given λ. From a Bayesian point of view it can be interpreted as the distribution of $Y^*$, given ρ, λ and $X$, when we have diffuse information about β and σ expressed by a uniform and independent distribution for β and log σ.

*The Priors for ρ and λ*

Let "$N$" be the norm of matrix $W$ defined as "the maximum value of a set of values in which each element is the sum of the elements of a line of matrix $W$". We suppose we have diffuse information for λ in the interval (0, ∞) and for ρ in the interval $[-1/N(\lambda), 1/N(\lambda)]$ [$W(\lambda)$ is always normalized dividing all its elements by $N(\lambda)$]. That is, we know the range for ρ given λ and, therefore, the distributions for ρ and λ are no longer independent.

If we assume that log λ and ρ are uniformly distributed, in their respective ranges, than the joint prior p.d.f. for them is given by

$$p(\rho, \lambda) \propto N(\lambda) / \lambda, \qquad 0 < \lambda < \infty;$$

$$-1/N(\lambda) < \rho < 1 / N(\lambda).$$

Therefore, the posterior for ρ, λ , $n(\rho, \lambda / X, Y^*)$, is:

$$n(\rho, \lambda / X, Y^*) \propto [N(\lambda) / \lambda] \,.\, |A| \,.\, (\hat{\sigma}^2)^{-T/2}, \qquad 0 < \lambda < \infty \qquad (5)$$

$$-1/N(\lambda) < \rho < 1 / N(\lambda)$$

If we integrate the joint p. d. f. for β and σ given in equation (3) with respect to σ, we obtain:

$$k(\beta / \rho, \lambda, X, Y^*) \propto |A| \,.\, \{T\hat{\sigma}^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})\}^{-T/2}$$

The posterior for $\beta$, $r(\beta/X, Y^*)$, is:

$$r\,(\beta/X, Y^*) = \iint k(\beta/\,\rho, \lambda, X, Y^*)\; n\,(\rho, \lambda\,/X, Y^*)\; d\rho d\lambda \tag{6}$$

$$r(\beta/X, Y^*) \propto \iint \{T\hat{\sigma}^2 + (\beta - \hat{\beta})'\, X'\, X\,(\beta - \hat{\beta})\}^{-T/2}\,[N(\lambda) / \lambda]\,.\,|A|^2\,.\,(\hat{\sigma}^2)^{-T/2}\, d\rho d\lambda$$

The next section describes how we use the Sampling-Importance-Resampling (SIR) method — introduced by Rubin (1988), to obtain a concentrated likelihood [$m(Y^*; \rho, \lambda, X)$ in equation (4)] sample, a sample for the posterior of $\rho$ and $\lambda$ [$n(\rho, \lambda/X, Y^*)$ in equation (5)], and how we can use SIR to obtain the mean of the posterior distribution of $\beta$ [$r$ $(\beta/X, Y^*)$ as described in equation (6)]. The integrations involved in the determination of the posteriors of $\rho$ and $\lambda$ and the posterior of $\beta$ can also be achieved by MCMC (Makov Chain Monte Carlo) methods like the Metropolis-Hasting algorithm [Hasting (1970)], and a special case of the single-component-Metropolis-Hasting algorithm known as Gibbs Sampler [Geman & Geman (1984)] that became popular after the articles of Gelfand & Smith (1990) and Gelfand *et alii* (1990).

## 7 - SAMPLING-IMPORTANCE-RESAMPLING (SIR): SAMPLES FOR THE CONCENTRATED LIKELIHOOD AND FOR THE POSTERIOR OF ρ AND λ

Since the joint prior pdf for $\rho$ and $\lambda$ is not proper it cannot be used as an importance function. Therefore, a gamma distribution with parameters 100 e 200 was used for $\lambda$ and, given $\lambda$, a uniform distribution with range in $-1/N(\lambda) < \rho < 1/N(\lambda)$ was used for $\rho$. The parameters of the gamma distribution were chosen based on the maximum likelihood estimate for $\lambda$. Since we have an almost flat prior for $\lambda$ and $\rho$ we expect the shape of the posterior to be closed to the shape of the likelihood. The mean of the chosen gamma distribution is 0.5 and the variance is 0.0025.

Let $q(\varphi)$ [$\varphi = (\lambda, \rho)$] be the importance function for $\lambda$ and $\rho$. The SIR method is as follows:

*i*) Generate draws $\varphi_1, \varphi_2, ..., \varphi_n$ from $q(\varphi)$. Each draw is generate by taking a draw for $\lambda$ from Ga(100,200) and, given de value of $\lambda$, taking a draw for $\rho$ from the uniform distribution $U\,(-1/N(\lambda), 1/N(\lambda))$. We have chosen n=100,000.

*ii*) Resample $\varphi_i$, $i = 1,..., n_r$ with probability $\pi_i$, and

$$\pi_i = [p(\varphi_i)\; m(Y^*; \varphi_i, X)/\; q(\varphi_i)]/\sum_{j=1}^{n}\;[p(\varphi_j)\; m(Y^*; \varphi_j, X)/q(\varphi_j)]$$

(if we want a sample of the posterior of $\varphi$)

or

$$\pi_i = [m(Y^*; \varphi_i, X) / \sum_{j=1}^{n} [m(Y^*; \varphi_j, X)]$$

(if we want a sample of the likelihood of φ)

where: $p(.)$ is the prior distribution for λ and ρ and $m(.)$ is the concentrated likelihood for λ and ρ (both were defined in the last section). We have chosen $n_r$ = 2,000.

It can be shown that the sample generated by the resample is a sample of the posterior or the likelihood of φ depending on the $\pi_i$ selected.

Our interest is to use SIR to obtain an approximation for the posterior and likelihood distribution of β, λ and ρ.

The posterior mean of β and φ are obtained by the following integrations,

$$E(\beta/X, Y^*) = \int_\beta \beta\, r(\beta/X, Y^*) d\beta = \int_\varphi \{\int_\beta \beta\, k(\beta/\varphi, X, Y^*)\, d\beta\}\, n(\varphi//X, Y^*)\, d\varphi =$$

$$= \int_\varphi E(\beta/\varphi, X, Y^*)\, n(\varphi/X, Y^*)\, d\varphi$$

$$E(\varphi/ X, Y^*) = \int_\varphi \varphi\, n(\varphi/X, Y^*)\, d\varphi$$

Using SIR they can be approximated by

$$E(\beta/X, Y^*) = \sum_{j=1}^{nr} E(\beta_i / \varphi_i) . f_i \quad \text{and} \quad E(\varphi/ X, Y^*) = \sum_{j=1}^{nr} \varphi_i . f_i$$

where $f_i$ is the relative frequency of the draw $\varphi_i$ in the resample.

Similarly the marginal posterior of β, using SIR, can be approximated by

$$r(\beta/X, Y^*) = \sum_{j=1}^{nr} k(\beta/\varphi, X, Y^*) . f_i$$

*Empirical Analysis*

Tables 5 and 6 present the results of the resample for the likelihood and posterior of λ, ρ, and Figure1 shows the histograms of the posterior and likelihood samples of this two parameters. The values shown in Table 5 provide evidence that the SIR procedure is quite successful in obtaining a sample for the likelihood: the values of the mode of λ and ρ are very close to the estimated values obtained by maximum likelihood, and there is a large number of points being selected by the resample. Results of Table 6 refer to the posterior sample and are also encouraging: as expected — since our joint prior for λ and ρ are relatively flat, the values obtained for the posterior are very similar to those presented for the likelihood. The histograms of Figure 2 show that the shapes of the likelihood and the posterior are, as expected, very similar. Finally, Table 7 presents basic statistics of the likelihood and posterior distributions obtained with the SIR procedure (3rd through 6th column), and reproduces data from Table 4 (1st and 2nd columns) regarding the bootstrap simulated results computed under the null hypothesis of classical MLE estimates.

Table 5
**Likelihood Sample**

| Hyperparameter | λ | Number of Points | ρ | Number of Points | Constant | Area | Age | Garage |
|---|---|---|---|---|---|---|---|---|
| 25% | 0.477488 | 187 | 0.00132391 | 240 | 5.88324 | 0.961539 | -0.0401399 | 0.154794 |
| 50% | 0.499271 | 385 | 0.00159293 | 375 | 5.89811 | 0.973093 | -0.0373774 | 0.156884 |
| 75% | 0.521873 | 582 | 0.00182816 | 502 | 5.91511 | 0.986307 | -0.0349617 | 0.159275 |
| Mode | 0.443874 | 9 | 0.00164931 | 408 | 5.89455 | 0.970324 | -0.0367984 | 0.156383 |
| Estimated Value (Max. Likelihood) | 0.466000 | - | 0.00162000 | - | 5.89800 | 0.973 | -0.0400000 | 0.154000 |

*OBS: Likelihood Sample.*

Table 6
**Posterior Sample**

| Hyperparameter | λ | Number of Points | ρ | Number of Points | Constant | Area | Age | Garage |
|---|---|---|---|---|---|---|---|---|
| 25% | 0.471142 | 290 | 0.00136491 | 391 | 5.88366 | 0.962332 | -0.0411516 | 0.153550 |
| 50% | 0.494048 | 642 | 0.00159348 | 681 | 5.89846 | 0.973613 | -0.0387170 | 0.155897 |
| 75% | 0.518521 | 1012 | 0.00182770 | 969 | 5.91287 | 0.984986 | -0.0364059 | 0.158038 |
| Mode | 0.440467 | 16 | 0.00146116 | 512 | 5.90035 | 0.975256 | -0.0392041 | 0.156927 |
| Estimated Value (Max. Likelihood) | 0.466000 | - | 0.00162000 | - | 5.89800 | 0.973000 | -0.0400000 | 0.154000 |

*OBS: Posterior Sample.*

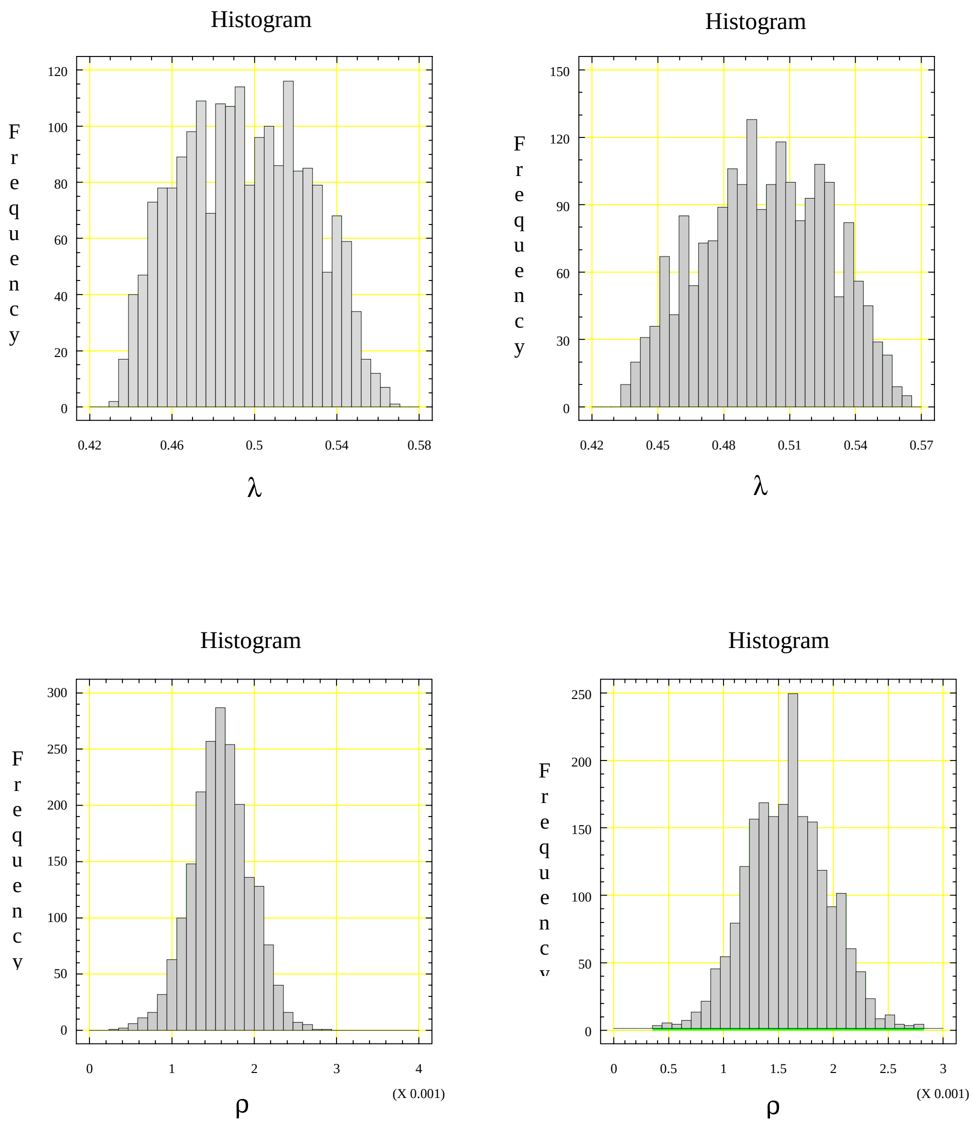
Table 7
**Boostrap Classical MLE and SIR**

| Hyperparameter | Bootstrap Simulated Results | | SIR | | | |
|---|---|---|---|---|---|---|
| | | | Posterior | | Likelihood | |
| | λ | ρ | λ | ρ | λ | ρ |
| mean | 0.4206864 | 0.0016167 | 0.4951869 | 0.0015945 | 0.4988072 | 0.0015825 |
| mode | - | - | 0.4404670 | 0.0014612 | 0.4438740 | 0.0016493 |
| Standard Deviation | 0.2347235 | 0.0003672 | 0.0302537 | 0.0003573 | 0.0287179 | 0.0003664 |

*The null hypotheses about the values are: 0,466 (λ), 0,0162 (ρ), 0,21217(sd λ) e 0,00035 (sd ρ).*
*Simulation number = 289.*
*Importance Functions: λ ~ Ga(100,200) e ρ ~ U (-1,1).*

Figure 2
Histograms of the Posterior and Likelihood Samples

Posterior Sample                                    Likelihood Sample




## 8 - CONCLUSIONS

This study adopts a classical maximum likelihood approach and also a Bayesian Sampling-Importance-Resampling (SIR) procedure to estimate the weights matrix and the significance of spatial dependence. It appplies the two estimation procedures to data on housing prices in the city of Belo Horizonte, Brazil, and compares the results obtained with these two techniques with the one derived by *a priori* fixing the weights.

The main results are: the estimated distance-decay parameter is quite different from the standard *a priori* assumptions such as the "all-or-nothing/no decay within the cut-off distance" or the "inverse distance" adopted in the empirical literature (fractionary value instead of the integer value usually used); the likelihood function of the weights matrix parameters has a well-defined peak; the Bayesian procedure allows for the introduction of *a priori* information on the range of parameters and assumes a flat prior leading to a posterior distribution not significantly different from the likelihood.

## BIBLIOGRAPHY

ANSELIN, L. *What is special about spatial data? Alternative perspectives on spatial data analysis*. Santa Barbara, CA: University of California, National Center for Geographic Information & Analysis, Geography Department, 1989, (Technical Paper, 89-4).

ANSELIN, L. *Spatial econometrics: methods and models.* The Netherlands: Kluwer Academic Publishers, 1988.

CLIFF, A., ORD, J. *Spatial autocorrelation*. London: Pion, 1973.

DOREIAN, P. Linear models with spatially distributed data, spatial disturbances or spatial effects? *Sociological Methods and Research*, v. 9, p. 29-60, 1980.

GELFAND, A. E., SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *Journal American Statist.* Ass., v. 85, p. 398-409, 1990.

GELFAND, A. E., HILLS, S. E., RACINE-POON, A., SMITH, A. F. M. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal American Statist.* Ass., v. 85, p. 972-985, 1990.

GEMAN, S., GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, p. 721-741, 1984.

GRAYBILL, F. *Matrices with applications in statistics*. Belmont, California: Wadsworth International Group, 1983.

HASTINGS, W. K. Monte Carlo sampling methods using Markov Chain and their applications. *Biometrika,* v. 57, p. 97-109, 1970.

MORAN, P. The Interpretation of statistical maps. *Journal of the Royal Statistical Society B*, v. 10, p. 243-251, 1948.

ORD, K. Estimation methods for spatial interaction. *Journal of the American Statistical Association*, v. 70, p. 126-26, 1975.

RUBIN, D. B. Using the SIR Algorithm to simulate posterior distributions. In: BERNARDO, J. M., DeGROOT, M. H., LINDLEY, D. V., SMITH, A. F. M. *Bayesian statistics,* 3ª ed. Cambridge, MA: Oxford University Press, p. 395-402, 1988.

SMITH, A. F. M., GELFAND, A. E. Bayesian statistics without tears: a sampling-resamplig perspective. *Journal American Statist.* Ass., v. 46, n. 2, p. 84-88, 1992.

ZELLNER, A. *An introduction to Bayesian inference in econometrics*. Malabar, Florida: Robert E. Krieger Publishing Company Inc., 1971.