VALOREM
Revista da Sociedade Brasileira de Engenharia de Avaliações

MASS APPRAISAL OF URBAN LAND WITH HOMOGENIZATION FACTORS: A SPATIAL MODELS-BASED APPROACH

Avaliação em massa de terrenos urbanos com fatores de homogeneização: uma abordagem baseada em modelos espaciais

Antônio Augusto Ferreira de Oliveira
http://orcid.org/0000-0003-3890-4219
Secretaria Municipal das Finanças do Município de Fortaleza (SEFIN), Fortaleza, Brasil.
augusto.oliveira@sefin.fortaleza.ce.gov.br

Sandro Ricardo Vasconcelos Bandeira
http://orcid.org/0000-0003-2730-025X
Secretaria Municipal das Finanças do Município de Fortaleza (SEFIN), Fortaleza, Brasil.
sandro.bandeira@sefin.fortaleza.ce.gov.br

Victor Jucá Távora
http://orcid.org/0009-0002-0878-640X
Secretaria Municipal das Finanças do Município de Fortaleza (SEFIN), Fortaleza, Brasil.
victor.tavora@sefin.fortaleza.ce.gov.br

ABSTRACT

This study uses spatial econometrics techniques within Fortaleza's municipality to address the real estate mass appraisal of urban lands. It utilizes spatial weight matrices, Lagrange multiplier tests, and spatial regression models to comprehend and model the spatial structure of land prices. The findings highlight the critical role of careful selection of spatial weight matrices in reducing residual autocorrelation and enhancing estimation accuracy. Moreover, employing homogenization factors based on a paradigm lot proved effective in adjusting land prices, aiding in more equitable taxation, and reflecting price variations due to locational factors. The study's specific objectives include evaluating the influence of spatial weight matrices' choice, analyzing the effectiveness of spatial models in capturing spatial dependence, and investigating the impact of price homogenization. The study also outlines methodological limitations such as reliance on georeferenced data and the assumption that land attributes uniformly impact prices across different city areas.

Keywords: Real estate mass appraisal; Homogenization factors; Spatial regression.

RESUMO

Este estudo aborda a avaliação em massa de terrenos urbanos utilizando técnicas de econometria espacial no contexto do município de Fortaleza. A metodologia emprega matrizes de pesos espaciais, testes de multiplicadores de Lagrange e modelos de regressão espacial para entender e modelar a estrutura espacial dos preços dos terrenos. Os resultados indicam que a seleção cuidadosa das matrizes de pesos espaciais é fundamental para reduzir a autocorrelação residual e melhorar a precisão das estimativas. Além disso, a aplicação de fatores de homogeneização baseados em um lote paradigma demonstrou ser uma abordagem eficaz para ajustar os preços dos terrenos, contribuindo para uma tributação mais equitativa e refletindo variações de preços devido a fatores locacionais. Os objetivos específicos do estudo incluem avaliar a influência da escolha das matrizes de pesos espaciais, analisar a eficácia dos modelos espaciais na modelagem da dependência espacial e investigar o impacto da homogeneização dos preços. Este trabalho também discute as limitações da metodologia, como a dependência de dados georreferenciados e a premissa de que os atributos dos terrenos impactam uniformemente os preços em diferentes áreas da cidade.

Palavras-Chave: Avaliação em massa; Fatores de homogeneização; Regressão espacial.

Preenchimento dos Editores

INFORMAÇÕES SOBRE O ARTIGO

Submetido em	21/03/2024
Publicado em	25/04/2024

Comitê Científico Interinstitucional
Editor-Responsável: Carlos Augusto Zilli
Avaliado pelo Sistema Double Blind Review (SEER/OJS – Versão 3)

1. INTRODUCTION

Mass real estate appraisal techniques have undergone numerous transformations in recent decades. These changes are driven by the increased availability and ease of obtaining property characteristics and price data sourced from real estate buying and selling portals, market observatories maintained by public and private entities, and massive georeferenced property registration conducted in Brazilian municipalities. Through Geographic Information Systems (GIS), other variables and data sources, including cadastral maps, satellite images, unmanned aerial vehicles (UAVs), 360º terrestrial imaging, socioeconomic data, and urban planning information, among others, can be integrated, enabling the identification of patterns, trends, and spatial relationships that impact property prices. Additionally, the advancement of algorithms in spatial econometrics, geostatistics, and, more recently, machine learning—a subfield of artificial intelligence—has significantly contributed to improving these appraisal methods.

Despite these advances and the breadth of analytical techniques available, the treatment of data by factors is still widely used in mass real estate appraisal, especially for tax purposes in preparing general property value maps (PVG) in almost all Brazilian municipalities. This is due, among other reasons, to the ease of developing and applying the method, its transparency, and its explanatory power for those subject to the tax incidence. On the other hand, the indiscriminate use of this approach, applying factors imported from the classical national literature devoid of any underlying scientific methodology, can cause severe distortions in assessed values, leading to inequities in property taxation.

Moreover, it is crucial to consider the spatial structure of property prices and attributes in global models. Ignoring this can lead to errors in the magnitude and significance of the coefficients, especially in areas with strong spatial autocorrelation. As seen below, classical regression assumptions are not met in the presence of spatial dependence and/or spatial heterogeneity. These effects are inherent to the behavior of property market prices, as evidenced in numerous studies. The urgency of addressing this issue cannot be overstated.

In this context, this study aims to propose a scientific methodology for the mass appraisal of urban land through the spatial econometric approach and subsequent extraction of homogenization factors, making it replicable to other municipalities. The specific objectives include (1) evaluating the influence of the choice of spatial weight matrices in capturing and modeling spatial dependence; (2) analyzing the effectiveness of spatial regression models (SAR, SEM, and SAC) in reducing residual autocorrelation and improving price estimate accuracy; (3) investigating the impact of land price homogenization using homogenization factors derived from the most appropriate spatial model, and evaluating their representativeness and applicability in different regions of the municipality.

This article is organized as follows: in the Literature Review section, we analyze the primary studies on mass real estate appraisals with a spatial approach in Brazil and South America. Next, in the Theoretical Framework, we detail the theories and spatial econometric models that underpin the research, including the definitions of key concepts such as spatial dependence, spatial heterogeneity, spatial weight matrices, detection of spatial autocorrelation, and other related concepts. The Methodology section describes the sample used, the methods for specifying and choosing the spatial weight matrices, and the flowchart for analyzing the most appropriate spatial model according to the methodology of Anselin and Rey (2014). The Results are presented in the fourth section, followed by the Discussion, in which we interpret the findings in light of the reviewed literature. Finally, in the Conclusion, we summarize the main findings, discuss the study's limitations, and suggest directions for future research.

2. LITERATURE REVIEW

Dantas et al. (2003), in a pioneering study on spatial econometrics applied to property appraisal in Brazil, demonstrated the necessity of considering the spatial autocorrelation present in real estate market data. The study sample comprised 59 market data of apartments located in four neighborhoods in the central-western region of Recife. The authors stated that not considering the spatial dependence in the data can lead to biased, inconsistent, and inefficient results. Moreover, they identified that spatial autocorrelation in the residuals is related to (i) measurement errors caused by artificial divisions of geographical units, such as neighborhood boundaries that do not necessarily coincide with the actual dimension of the observed phenomenon, and (ii) the omission of relevant locational variables (model specification error).

Dantas (2008) applied spatial econometric models to analyze the housing market in Recife using a sample of 232 apartments financed by CAIXA, collected between June 2000 and June 2002, distributed across 36 neighborhoods. The spatial weight matrix used in the study was based on the inverse distance between buildings. The influence between properties was weighted according to this metric, and the matrix was row-standardized to facilitate the interpretation of weights as a weighted average of the neighbors' values. The study concluded that there is a strong presence of spatial dependence in property prices, demonstrating the need to use spatial econometric models to obtain more accurate and consistent estimates. The analysis showed that locational and structural characteristics and neighborhood spatial interaction effects influence property prices.

The article "Engineering Evaluations Based on GAMLSS Models" by Florêncio (2010) used a sample of 2,109 urban land observations in Aracaju, covering 2005, 2006, and 2007. The study's main objective was to estimate the hedonic price function using the class of GAMLSS (Generalized Additive Models for Location, Scale, and Shape) regression models. According to graphical and numerical residual analyses and Akaike and Schwarz criteria, the results showed that GAMLSS models provided a superior fit compared to traditional classical linear regression and generalized linear models (GLM). The study concluded that the GAMLSS model class is more appropriate for estimating the parameters of the hedonic price function due to its flexibility and ability to fit a broad family of distributions for the response variable.

Campos and Almeida (2018) used hierarchical models that considered spatial effects on a sample of 9,682 apartment launches in São Paulo between January 1985 and July 2012. The authors concluded that property sale prices are influenced both by intrinsic property characteristics and by neighborhood and adjacency effects. According to them, spatial hierarchical models allow for a more robust analysis of property price variations, requiring control for the spatial dependence of the data and the spatial heterogeneity of amenities among the city's districts and their spillover effects.

Bandeira (2020) compared results from the classical regression model with those of a spatial model using land plots in Fortaleza as the study object. A sample of 914 data points was used for model estimation, and another validation sample was used for 423 observations. The methodology followed by the author indicated the spatial lag model as the most suitable, using a spatial weight matrix weighted by the inverse of the squared Euclidean distance between observations. The author concluded that, although the estimated unit values were similar, there were specification errors in the classical model.

The article "Mass Appraisal Using Generalized Additive Models and Geostatistics" (OLIVEIRA; OLIVEIRA; BANDEIRA, 2020) used a sample of 1,924 urban land data from Fortaleza. The main objective was to develop a methodology for mass property appraisal using Generalized Additive Models (GAM) combined with geostatistics techniques to improve the accuracy of appraisals for tax purposes. The results confirmed hypotheses about real estate market price behavior, such as the increase in unit value of land located in gated communities, with available water networks, and located on paved roads. However, incorporating land coordinates into the GAM model was insufficient to eliminate the spatial autocorrelation of residuals, requiring the application of a spherical variogram and ordinary kriging interpolation of the residuals. It was concluded that the proposed methodology is promising but should be used cautiously due to potential measurement errors in market data collection.

Zilli and Hochheim (2023) used global spatial regression models (CAR) and geographically weighted regression models (GWR) to address spatial dependence and heterogeneity in mass property appraisal. The sample consisted of 225 market data of urban properties in the Centro, Agronômica, and Trindade neighborhoods in Florianópolis, Santa Catarina, Brazil, collected between March and April 2020. The main conclusions indicated that the GWR model outperforms the conditional autoregressive (CAR) model in terms of accuracy and ability to consider spatial dependence and heterogeneity, standing out as a viable alternative for establishing a fair and equitable basis for property tax calculation.

Oliveira et al. (2023) made a significant contribution to the field of mass property appraisal by analyzing the selection of spatial weight matrices in spatial econometric models. They aimed to capture the maximum observed spatial dependence, thereby enhancing the quality of final appraisals. Their study, which utilized a sample of 3,580 house price data points in Aracaju, revealed the spatial dependence of prices and residuals from the classical linear regression model. The authors' key finding was that distance-based spatial weight matrices lead to better model fits, as they more effectively incorporate spatial dependence and yield superior results in AIC and root mean square error (RMSE) criteria.

In South America, the work of Cerino et al. (2022) stands out, where spatial econometrics was also used to develop a scientific methodology to replace the traditional homogenization tables with factors used in Argentina since 1939. The sample consisted of 1,033 land data collected in the Province of Córdoba through the local Real Estate Market Observatory (OMI) from 2018 to 2020, including transfer tax data. Outliers were eliminated, especially in the case of under declarations of this tax, through local Moran index analysis. The chosen spatial weight matrix was of the inverse Euclidean distance type between observations with a contagion radius of up to 500m. Lagrange multiplier tests confirmed the existence of spatial autocorrelation in the residuals and the dependent variable, leading to the choice of a combined spatial autocorrelation model (Spatial Autoregressive Combined – SAC). The marginal effects of each variable of interest on the observed unit price were estimated with the composition (sum) of direct and indirect effects, the latter arising from the spatial dependence structure itself, as will be detailed in the following sections of this article. Thus, the total effect calculated a unique homogenization factor relative to a paradigm lot with 19m frontage and 562m² area. Finally, the authors used the proposed methodology to redesign the mass appraisal of land value throughout the Province of Córdoba.

3. THEORETICAL FRAMEWORK

3.1. SPATIAL REGRESSION AS AN ALTERNATIVE TO CLASSICAL REGRESSION IN MASS APPRAISALS

Like many other economic and social phenomena, the distribution of prices in the real estate market is influenced by spatial proximity effects. In mass real estate appraisal, these effects cannot be ignored, considering that the main objective is to assess many properties within a geographical context that presents significant spatial interactions. Although hedonic price models have been successful since Rosen's (1974) work, which attempts to explain an individual's choice by seeking maximum satisfaction from a particular "bundle of attributes" present in a property influencing their purchase decision, this paradigm of an isolated and atomized economic agent (GOLGHER, 2015), which considers only the individual and intrinsic characteristics of each property, must be overcome when studying mass appraisals.

Property prices are highly influenced by neighborhood aspects and economic externalities, both positive and negative, that propagate beyond the traditional boundaries of study units considered in classical models. Neighborhood aspects relate to the concept of spatial dependence (or spatial autocorrelation), and externalities relate to the idea of spatial heterogeneity, both definitions being species of the genus "spatial effects" according to Almeida (2012).

The existence of spatial dependence observed in the real estate market has its first explanation in the well-known "First Law of Geography" (TOBLER, 1970), according to which geographically closer individuals tend to behave more similarly than those who are further apart (ALMEIDA, 2012). This can occur in both the observed property prices and the explanatory variables, as well as in the residuals resulting from the application of classical regression models. The dependence on the prices of geographically proximate properties and their intrinsic characteristics can be intuitively explained because these properties share the same urban infrastructure and access to services, have similar amenities, are subject to the same urban planning parameters, and have the same topography, among other common characteristics.

The stochastic errors become spatially autocorrelated if this spatial dependence is not incorporated into classical regression models. For example, suppose important locational explanatory variables are omitted in the classical regression model (specification error) or are insufficient to explain the observed price behavior. In that case, the prices of properties in more affluent regions may exhibit positive residuals correlated with those of other properties in similar and nearby regions. At the same time, properties in poorer areas would show negative residuals that correlate with other properties in adjacent regions. This is quite evident with the elaboration of residual maps and the use of clustering techniques or the application of the LISA (Local Indicators of Spatial Association) index (ANSELIN, 1995). With spatial dependence, the assumption of exogeneity is violated, meaning that stochastic errors cannot be correlated with the explanatory variables (GOLGHER, 2015).

$$E[\epsilon \mid X] = 0 \quad (1)$$

It is worth noting that this violation results in a harmful and undesirable effect in mass appraisals for tax purposes: fiscal regressivity, in which properties with lower assessed values end up paying a proportionally higher property tax than those with higher values.

Spatial heterogeneity can lead to heteroscedasticity, where the variance of errors varies according to geographical location. Its concept is linked to Goodchild's Second Law of Geography (ALMEIDA, 2012; GOODCHILD, 2004), which can be understood as the presence of a specific stability in the behavior of variables in a context of uneven spatial structure due to historical economic development of the locations or due to externalities. In other words, in the real estate market context, the estimators of explanatory variables for price behavior in a particular region may vary in sign and intensity from one locality to another. For instance, proximity to the beach in coastal cities may add more value to properties in those areas than to those farther from the coast; the effect of the "corner lot" variable may be more significant in commercial areas than in residential ones; land plots with suitable dimensions for real estate development may have a considerable premium in upscale regions with a shortage of vacant lots compared to the other areas where there is an abundance of them. Moreover, positive externalities, such as the construction of a shopping center, the revitalization of an urban park, or an increase in the exploitation index due to changes in urban legislation, impact not only the properties directly affected but also spill over their effects beyond (spillover). Conversely, the increase in crime in an area or air and noise pollution caused by opening a new expressway near residential areas are examples of negative externalities whose spatial effects need to be known in advance in modeling. However, Golgher (2015) emphasizes that non-spatial econometric techniques can also address heteroscedasticity.

3.2. SPATIAL WEIGHT MATRIX

Through the methodology developed by Anselin (1988), the spatial econometrics approach indicates that a spatial weight matrix (**W**) must be defined beforehand. The **W** matrix will be included in the regression model to capture the spatial interdependence between the various geographic units of analysis (in this case, between the sample points of property locations).

Basically, spatial weight matrices come in binary neighborhood (or binary contiguity) and distance matrices. Contiguity matrices can consider neighboring observations as those that only share borders (rook) or those that share vertices (queen). These matrices equally weigh the values of the found neighbors without considering the distances between them. Additionally, contiguity matrices can be expanded to include higher orders, i.e., not only immediate neighbors but neighbors of neighbors, allowing for a more comprehensive analysis of spatial interactions at different levels of proximity. In property appraisals, represented by points (and not polygons), contiguity matrices are constructed by polygons generated by some tessellation process, such as Voronoi or Delaunay triangulation.

Distance matrices, on the other hand, weigh the influence of neighbors based on fixed distance or "k-nearest neighbors." In the first case, a maximum contagion distance is defined, beyond which there is no neighborhood relationship; for observations within this distance, the neighborhood relationship is weighted by the inverse distance or higher-order negative exponentials of this distance. In the second case, weights are defined for a fixed number of neighbors, and there is the possibility of weighting them by negative exponentials.

According to Anselin (1988), a matrix that best represents the spatial dependence of the studied phenomenon should be chosen. According to Corrado and Fingleton (2012), the various criteria for constructing the **W** matrix directly impact the model results. Defining them appropriately is a sine qua non condition for a good spatial model, as will be seen later in the "Results" section. An inadequate **W** matrix can leave unexplained residual autocorrelation, violating the regression model's assumptions (KOBYLIŃSKA, 2021).

3.3. MEASUREMENT OF SPATIAL AUTOCORRELATION

3.3.1. Global Moran's I Index

The Moran's I index is the primary indicator of global spatial autocorrelation used in spatial regression models. The Moran's I index is given by:

$$I = \frac{n \sum_i \sum_j \left(w_{ij} \, (x_i - \bar{x})(x_j - \bar{x}) \right)}{\left(\sum_i \sum_j w_{ij} \right) \left(\sum_i (x_i - \bar{x})^2 \right)} \qquad (2)$$

Where *n* is the number of observations; w_{ij} is the element of the spatial weight matrix **W**; x_i and x_j are the values of the observations at location i and location j; $\bar{x}$ is the mean of the observations; $\sum_i \sum_j w_{ij}$ is the sum of all spatial weights, which, in the case of **W** being row-standardized, is equal to *n*.

Positive values of Moran's I indicate positive autocorrelation, while negative values indicate negative autocorrelation. Values close to zero indicate the absence of autocorrelation. A better interpretation of the index can be obtained through the Moran scatterplot, which is a scatter plot where the abscissa represents the values of the variable xxx, normalized about the mean:

$$z_i = x_i - \bar{x} \tag{3}$$

and the value of this lagged variable represents the ordinate, that is, the weighting of the values of this same variable for the neighbors of x_i by the spatial weight matrix **W**, also normalized:

$$Wz_i = \sum_j w_{ij}(x_j - \bar{x}) \tag{4}$$

The slope of the regression line fitted to the Moran scatterplot corresponds to the value of the index itself (Figure 1). In the first and second quadrants, we observe positive spatial autocorrelation, as points with high values have spatial contributions from neighbors also with high values (1st quadrant). Similarly, low-value points have neighbors with low values (2nd quadrant). In the third and fourth quadrants, we observe negative spatial autocorrelation, as points with high values have spatial contributions from neighbors with low values (3rd quadrant). Similarly, low-value points have neighbors with high values (4th quadrant).

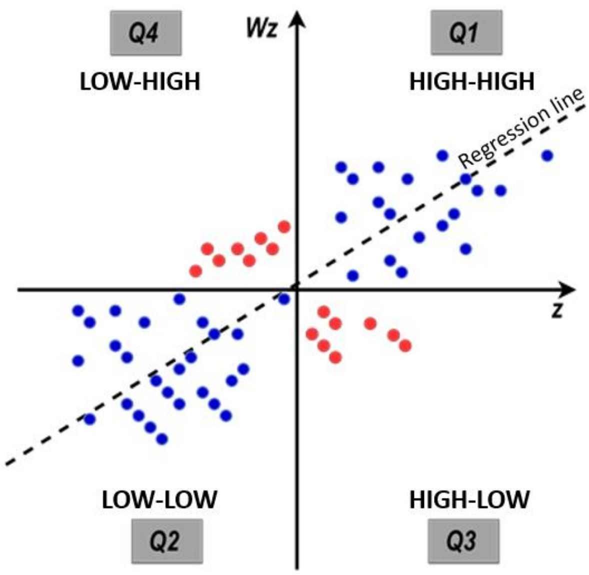

Figure 1 - Moran's I Scatterplot

After calculating Moran's I index, the statistical significance of the observed spatial autocorrelation must be evaluated. This is achieved through an iterative process in which the observations are permuted while keeping the spatial weight matrix unchanged. The purpose of each permutation is to create a spatial distribution of the data such that the initially observed spatial autocorrelation is eliminated. With their respective calculated indices, the various permutations generate a reference distribution of Moran's I statistics under the null hypothesis of no spatial autocorrelation. By comparing the observed statistic (the calculated index value) with the null distribution, one can derive a statistical significance assessment of the observed spatial autocorrelation.

3.4. SPATIAL MODELS

According to Dantas (2003), the spatial regression model aims to formally analyze spatial interaction in cross-sectional data models in studies involving geographic units. Given the expected spatial correlation between the dependent variable (unit price), its explanatory variables, and/or its residuals, this methodology has been widely used in real estate market analysis. An appropriate spatial model incorporates such effects into the regression equation for each situation.

3.4.1. Spatial Lag Model (SAR)

The spatial lag model, also known as the Spatial Autoregressive Model (SAR) and by the acronym SLM, standing for Spatial Lag Model, serves a specific purpose. The SAR model is used to correct distortions when the dependent variable has spatial autocorrelation and follows the structure shown in equation (5) in its matrix form:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\cdot\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I_n}) \tag{5}$$

Where $\mathbf{y}$ is the vector of values of the dependent variable; ρ is the spatial autoregressive coefficient; $\mathbf{W}$ is the spatial weight matrix; $\mathbf{X}$ is the matrix of explanatory variables; $\boldsymbol{\beta}$ is the vector of coefficients of the explanatory variables, and $\boldsymbol{\varepsilon}$ is the vector of error terms, normally distributed with a mean of zero and variance σ^2. Rearranging equation (5), we have:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\,(\mathbf{X}\cdot\boldsymbol{\beta} + \boldsymbol{\varepsilon}), \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I_n}) \tag{6}$$

Where $\mathbf{I}$ is the identity matrix.

3.4.2. Spatial Error Model (SEM)

The spatial error model (Spatial Error Model – SEM) is used to correct distortions when there is spatial autocorrelation of the errors and follows the structure shown in equation (7):

$$\boldsymbol{y} = \mathbf{X}\cdot\boldsymbol{\beta} + \mathbf{u}\,, \;\; \mathbf{u} = \lambda\mathbf{W}\cdot\mathbf{u} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I_n}) \tag{7}$$

Where $\boldsymbol{y}$ is the vector of values of the dependent variable; λ is the spatial autocorrelation coefficient of the errors; $\mathbf{W}$ is the spatial weight matrix; $\mathbf{X}$ is the matrix of explanatory variables; $\boldsymbol{\beta}$ is the vector of coefficients of the explanatory variables; $\mathbf{u}$ is the vector of error terms, which is spatially autocorrelated, and $\boldsymbol{\varepsilon}$ is the vector of error terms, normally distributed with a mean of zero and variance σ^2. The equation (7) can also be written as:

$$\boldsymbol{y} = \mathbf{X}\cdot\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\cdot\boldsymbol{\varepsilon} \tag{8}$$

Where $\mathbf{I}$ is the identity matrix.

3.4.3. Spatial Autocorrelation Combined Model (SAC)

The spatial autocorrelation combined model, SAC, combines elements of the spatial lag model (SAR) and the spatial error model (SEM). It is also known as the Kelejian-Prucha Model, named after researchers Harry Kelejian and Ingmar Prucha, who significantly contributed to developing and applying this type of spatial model. Additionally, it is referred to as SARMA (Spatial Autoregressive Moving Average Model) or SARAR (Spatial Autoregressive and Autoregressive Error Model). The representation of this model is given by equation (9):

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\cdot\boldsymbol{\beta} + \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}\,, \qquad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I_n}) \tag{9}$$

Rearranging equation (9), we have:

$$\boldsymbol{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\cdot\mathbf{X}\cdot\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\cdot\boldsymbol{\varepsilon} \tag{10}$$

3.4.4. Lagrange Multipliers (LM)

The global Moran's I index does not indicate which spatial model is the most suitable. For this purpose, we need to calculate the LM tests of Lagrange multipliers, which refer to the spatially lagged dependent variable and the spatially lagged error, or both (combined). The LM tests are hypothesis tests that follow a chi-squared distribution with one degree of freedom. The null hypothesis for the LM (error) test is the absence of spatial autocorrelation in the error term (λ = 0). For the LM (lag) test, it is the absence of spatial lag in the dependent variable (ρ = 0).

According to the expressions below, both tests can be calculated using the residuals from the multivariate linear regression with OLS estimation.

$$\mathrm{LM}_{\lambda}(\mathrm{erro}) = \frac{\left[\frac{\frac{\mathbf{e}'\mathbf{W}_e}{(\hat{\sigma}_{ML})^2}}{n}\right]}{[tr(\mathbf{W}^2 + \mathbf{W}'\mathbf{W})]} \approx X^2 \tag{11}$$

$$\mathrm{LM}_{\rho}(\mathrm{defasagem}) = \frac{\left[\frac{e'\mathbf{W}_e}{(\hat{\sigma}_{ML})^2}\right]^2}{\left[(\mathbf{W}\,\mathbf{X}\,\boldsymbol{\beta})'\frac{(\mathbf{MW}\,\mathbf{X}\,\boldsymbol{\beta})}{(\hat{\sigma}_{ML})^2} + tr(\mathbf{W}^2 + \mathbf{W}'\mathbf{W})\right]} \approx X^2 \tag{12}$$

Where $\mathbf{e}$ is the vector of residuals obtained by the ordinary least squares (OLS) method; **W** is the spatial weight matrix (**W′** is the transposed matrix); $\hat{\sigma}_{ML} = \frac{e' \cdot e}{n}$ is the maximum likelihood estimate of the variance of the model **y** = **Xβ** + **ε**; *n* is the number of sample data points; *tr* is the trace operator of the matrix; $\boldsymbol{y}$ is the observation vector of the dependent variable; $\boldsymbol{X}$ is the observation matrix of the independent variables; $\boldsymbol{\beta}$ is the parameter vector of the independent variables (OLS model); $\mathbf{M} = \mathbf{I} - \mathbf{X} \cdot (\mathbf{X}' \cdot \mathbf{X}) - \mathbf{I} \cdot \mathbf{X}'$, where $\mathbf{I}$ is the identity matrix.

Anselin and Rey (2014) comment that the classic LM_{ρ} (lag) test is also sensitive to the presence of spatial autocorrelation in the errors ($\lambda \neq 0$). Similarly, the LM_{λ} test is sensitive to the presence of spatial autocorrelation in the dependent variable itself ($\rho \neq 0$). Thus, these authors suggest performing new tests (robust LM^*_{ρ} and LM^*_{λ}) that corrects the behavior. As seen in section 4.2.2, the robust tests will be used to distinguish the best spatial econometric model when both classic tests, LM_{ρ} and LM_{λ}, are significant.

3.4.5. Model Selection Criteria for Spatial Regression

Once spatial dependence in the model has been identified, the appropriate spatial regression model (error or lag) must be chosen to address the problem. Anselin and Rey (2014) established a practical methodology based on various studies to help decide which model will perform better in most situations. The tests are conducted using the residuals obtained from the classical model and the predicted values for the dependent variable, both associated with the chosen spatial weight matrices **W**.

According to the authors, the first step is to analyze the LM_{ρ} and LM_{λ} statistics, which follow a chi-squared distribution, where the null hypothesis indicates the absence of spatial autocorrelation. If the null hypothesis of no spatial dependence is accepted for both the error and the dependent variable, the classical regression model with OLS estimates should be retained.

On the other hand, if only one of the statistics is significant, the model indicated by the significant test should be used: the Spatial Error Model (SEM) for the LM_{λ} (error) statistic, or the Spatial Lag Model (SAR) for the LM_{ρ} (lag) statistic. Finally, suppose both statistics are significant, indicating the simultaneous presence of spatial autocorrelation, the robust tests LM^*_{ρ} and LM^*_{λ} are compared. The model with the significant test, or if both are significant, the model with the higher test statistic value, should be chosen. Figure 2 summarizes the process to be followed.

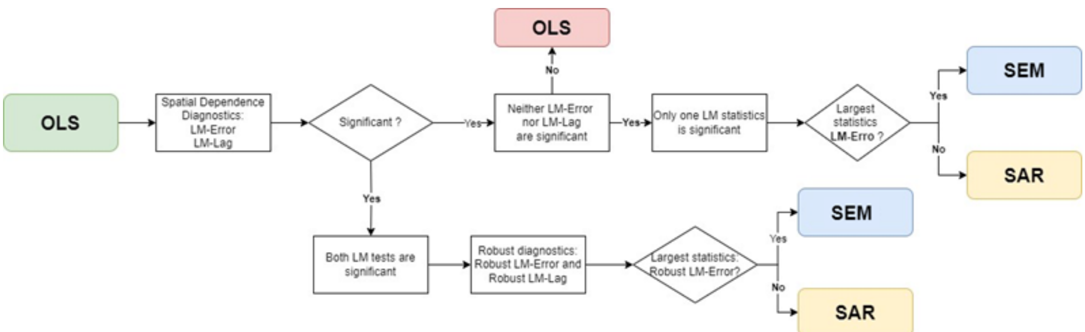

Figure 2 - Schematic Flowchart for Selecting the Spatial Regression Model (with adaptations Anselin and Rey, 2014).

3.4.6. Marginal Contributions of Explanatory Variables in Spatial Models

To fully understand the methodology proposed in this article, it is essential to comprehend how each explanatory variable contributes to forming the dependent variable values. In other words, in the present context, it is of interest to determine how certain variables, such as land area, property frontage, the number of frontages, the exploitation index, the presence of urban infrastructure, among others, influence the formation of the observed unit price of the land while keeping all other variables fixed (*ceteris paribus* condition).

It's important to understand that not all spatial models can be interpreted using the same approach as classical regression models. In classical regression, we calculate simple partial derivatives to analyze how the dependent variable is affected. This approach assumes that each observation is independent, so a change in an explanatory variable for one observation affects only the dependent variable for that specific observation. However, with spatial dependence, this isn't the case. A change in one observation also affects others and feeds back into the original observation (GOLGHER, 2015).

Consider a SAR model given by equation (6). The derivative of y with respect to an independent variable x_k is given by:

$$\frac{\partial y}{\partial x_k} = (\mathbf{I} - \rho\mathbf{W})^{-1}\beta_k \quad (13)$$

At the outset, it is observed that the partial derivative has a matrix that multiplies the coefficient β_k, ruling out the possibility of simply interpreting β_k as the marginal contribution of the variable x_k to y. This matrix is known as the matrix of spatial impacts (ANSELIN, 1988) and propagates the spatial effects that arise from the neighborhood network established by the chosen spatial weights matrix (**W**). According to Golgher and Voss (2016), to obtain the effects of a variable x_k on y in a SAR or SAC model, one must divide these effects into two:

- Direct effects: represent the expected average change in the dependent variable for all observations in a specific region due to a one-unit increase in one particular explanatory variable in that same region. They are calculated by the average of the sum of the elements on the main diagonal of the matrix of spatial impacts ($(\mathbf{I} - \rho\mathbf{W})^{-1}$).
- Indirect effects: represent the changes in the dependent variable of a specific region resulting from a one-unit increase in an explanatory variable in another region. The average of the elements off the diagonal of the matrix of spatial impacts calculates them.

In the context of this work, direct effects are the expected average changes in the price of a property due to a one-unit change in a specific characteristic of that same property, such as an increase of 1m² in the land area, an increase in frontage, etc. Indirect effects are the changes in the price of a property due to a one-unit change in a specific characteristic of properties in other locations. For example, a significant improvement in the infrastructure of a nearby area may increase the value of properties in adjacent areas. The sum of direct and indirect effects will be used to infer the homogenization factors of the explanatory variables of interest in this work, as seen in the following sections.

4. METODOLOGY

4.1. DATA

The sample consisted of 4,426 plots, with 18m² to 20,000m², geographically distributed randomly throughout the municipality's territorial area, covering the 121 neighborhoods, collected between 2019 and 2024. The sample data comes from three sources: (i) market offers collected from the Fortaleza Real Estate Observatory (OMI) (SEFIN, 2024); (ii) base values for ITBI calculations; and (iii) completed transactions. 76.2% of the data are listing prices, 12.9% are transactions, and 10.9% are ITBI guides. Data in the exact geographic location were eliminated, retaining only the most recent information. Extremely discrepant outliers were also removed during classical regression modeling, keeping the sample for all models.

It is observed that the highest unit prices are in the northern region of the municipality, particularly in the more touristic neighborhoods such as Meireles and Aldeota. In an initial exploratory data analysis, a positive spatial autocorrelation of unit prices in this region was observed using the spatial weights matrix with a queen contiguity criterion and applying the Local Spatial Association Index (LISA). This indicates that the unit price of a plot in this area is positively influenced by the high unit prices of its neighbors. This information, shown in red points in Figure 4, was statistically significant, as indicated by the map in Figure 3.

Properties generally have the lowest unit prices in the peripheral regions south of the municipality, where Bom Jardim, Siqueira, and Jangurussu neighborhoods are located. In this region, a positive spatial autocorrelation was also observed, indicating that the unit price of a plot tends to be influenced by the low prices of neighboring plots. This relationship, statistically significant (Figure 3), is highlighted by the blue points on the map in Figure 4.

Regions with gray points on both maps indicate statistically insignificant spatial autocorrelation values. This means that plots in these locations do not tend to experience significant influences from the prices of neighboring plots.

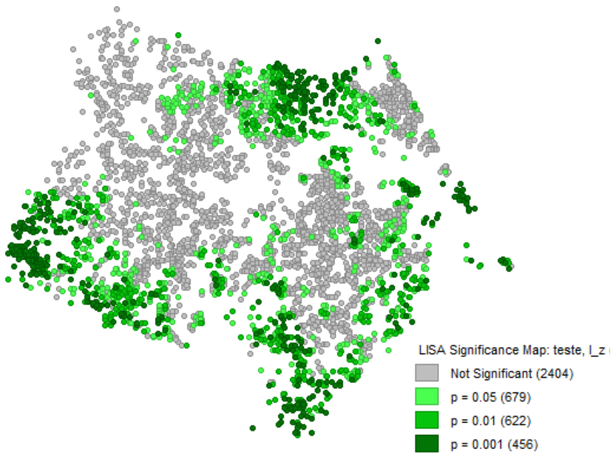

Figure 1 - LISA Significance Map

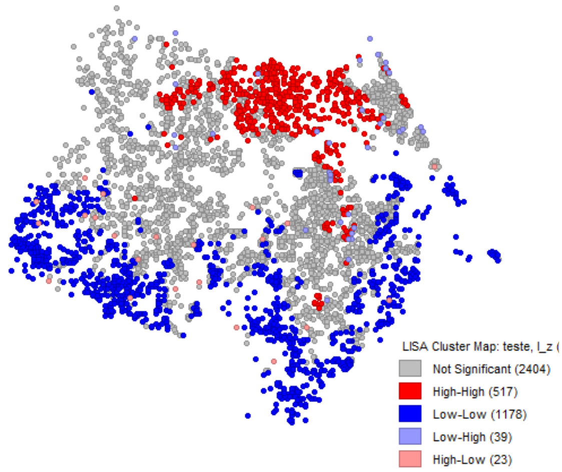

Figure 2 - LISA Clustering Map

4.2. LAND UNIT PRICES MODELING

4.2.1. Classic Model

The initial model was estimated using the ordinary least squares method with the natural logarithm of the unit price as the dependent variable. In total, 28 explanatory variables, including a second-degree trend polynomial, were used to capture the observed gradients and trends in large-scale prices. Many explanatory variables can be considered spatial proxies obtained through geoprocessing operations in geographic information systems (GIS). The generic regression model is shown in equation (14).

$$\begin{aligned}\ln(Punit) = {} & \beta_0 + \beta_1(x) + \beta_2(y) + \beta_3(x^2) + \beta_4(xy) + \beta_5(lotcond) + \beta_6(numfr) + \beta_7(esq) \\ & + \beta_8(f\_inc) + \beta_9(test) + \beta_{10}(\ln[area]) + \beta_{11}(pvacp) + \beta_{12}(gp) + \beta_{13}(\ln[renda]) \\ & + \beta_{14}(dv) + \beta_{15}(\ln[dscom]) + \beta_{16}(acasamed) + \beta_{17}(iamaxeq) + \beta_{18}(pzpa1) \\ & + \beta_{19}(distpv) + \beta_{20}(\ln[distbm]) + \beta_{21}(\ln[distvp]) + \beta_{22}(a2020) + \beta_{23}(a2021) \\ & + \beta_{24}(a2022) + \beta_{25}(a2023) + \beta_{26}(a2024) + \beta_{27}(tr) + \beta_{28}(of) + \varepsilon\end{aligned} \tag{14}$$

Where *Punit* is the dependent variable, unit price, in R$/m²; *x*, *y*, x^2, and *xy* are continuous quantitative variables, where *x* and *y* represent the UTM coordinates (EPSG 31984) of the plot centroids (divided by 10,000) and normalized relative to the center of the municipality; they form the trend polynomial based on the data's location; *lotcond* is a dichotomous variable representing the plot's location, indicating whether it belongs to a gated community (1) or not (0); *numfr* is a discrete quantitative variable indicating the number of frontages the plot has; *esq* is a dichotomous variable representing the plot's location, indicating whether the property is on a corner (1) or not (0); *f_inc* is a dichotomous variable indicating whether the plot is eligible for incorporation, i.e., if it has a land area greater than or equal to 750m²; *test* is a quantitative variable representing the length of the main frontage in meters (m); *area* is a quantitative variable considering the total area of the plot, in m²; *pvacp* is a dichotomous variable indicating whether the plot has concrete, cobblestone, or asphalt paving; *gp* is a dichotomous variable indicating whether there is a stormwater gallery on the plot (1) or not (0); *renda* is a macro location variable, represented by the average income of the head of the household, in minimum wages, adjusted to a trend surface constructed by kriging, based on average income data of the household head in the census sector, provided by the Brazilian Institute of Geography and Statistics (IBGE) census (2010); *dv* is a continuous proxy variable representing the density of verticalization, using the concentration of vertical condominiums with elevators obtained by kernel interpolation with a quartic function of 200m radius relative to the plot centroid; *dscom* is a proxy variable representing the density of commercialization in the surroundings (500m radius from the plot centroid), weighted by the commercial built-up area obtained by quartic kernel; *acasamed* is a proxy variable representing interpolation by inverse distance weighting (IDW) within a 500m radius of the median built-up area of residential houses; *iamaxeq* is a quantitative proxy variable representing the maximum equivalent utilization index; *pzpa1* is a quantitative proxy variable representing the percentage of preservation area in the water resources protection zone (ZPA1); *distpv* is a quantitative and continuous variable corresponding to the Euclidean distance from the plot to the nearest prime location; *distbm* is a quantitative proxy variable measuring the distance (m) between the centroids of the surveyed plots and Av. Beira Mar; *distvp* is a quantitative proxy variable representing the distance (m) from the plot centroid to the nearest main road (expressway, arterial I, arterial II, and collector); *a2020*, *a2021*, *a2022*, *a2023*, and *a2024* represent binary encoding to indicate the time of data collection; *tr* and *of* are binary encoding variables indicating the 3 data sources; β_i are the model parameters to be estimated, and ε are the random errors.

4.2.2. Spatial Models

After estimating the classical model, where only the explanatory variables significant at 10% were retained, 16 types of spatial weight matrices were tested, including (i) six contiguity matrices (rook of 1st, 2nd, and 3rd orders and queen of 1st, 2nd, and 3rd orders), (ii) three k-nearest neighbors matrices (with 5, 10, and 20 nearest neighbors), and (iii) seven distance matrices (inverse distance and inverse distance squared limited to 1,320m, and inverse distance with limitations of 500m, 800m, 1,500m, 2,000m, and 3,000m).

Lagrange multipliers tests were calculated using the classical regression model for each listed matrix to determine the most suitable spatial regression model according to Anselin's (2005) criterion. Subsequently, estimates were made for the most appropriate spatial models (SAR or SER), followed by calculating the global Moran's I index to assess the matrix's ability to capture all spatial dependence without leaving any residual autocorrelation in the errors. In cases where more than one model showed no autocorrelation, the one with the lowest Akaike Information Criterion (AIC) was chosen. Figure 5 illustrates the described methodology.

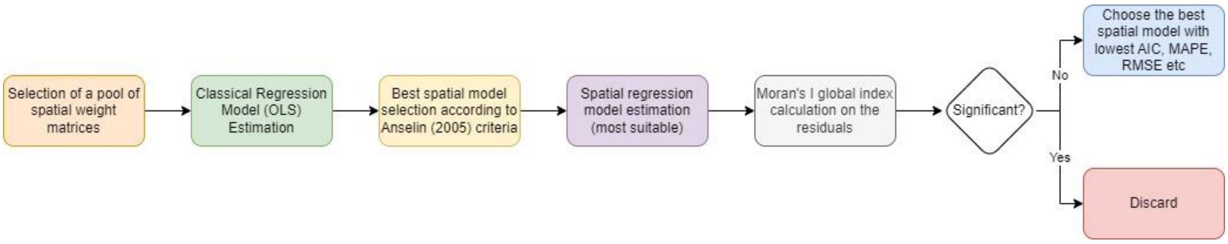

Figure 5 - Flowchart of the methodology for selecting the most appropriate spatial model.

4.3. HOMOGENIZATION FACTORS EXTRACTION

With the resulting equation from the most appropriate spatial model and lowest AIC, the next step is determining the weights for each explanatory variable. To do this, the characteristics of the benchmark plot are projected for each independent variable.

For establishing the benchmark plot, the characteristics of the most representative plot in the municipality were defined, considering the median values from the entire dataset of approximately 404,000 plots. Descriptive statistics of the cadastral base of territorial properties were analyzed for each variable of interest. However, it is essential to note that the variables *testada* and *area* did not follow this methodology. Instead, values of 12m and 400m² were established, approximately corresponding to the cadastral base's third quartile (Q3). The main reason for these exceptions is that plots with only 7m of frontage and 187m² of area, which are the respective median values, are less attractive and consequently less valued in the municipal market. Using these values would introduce biases that could impair the analysis of the factors.

Considering the choice of the spatial lag model, the influence of an attribute's variation on the dependent variable unit price was determined by the sum of direct and indirect effects, as discussed in section 3.4.6. The total effect was interpreted as a coefficient of a classical regression model (CERINO et al., 2022). Finally, the homogenization factor for each attribute was calculated, and its respective homogenization tables were prepared.

5. RESULTS AND DISCUSSION

Table 1 shows the models recommended for the various tested matrices based on the robust LM and LM* tests. Attention is drawn to the recommendation against using spatial models in the case of higher-order contiguity matrices (rook 2, rook 3, queen 2, and queen 3), which appears inconsistent given the presence of spatial autocorrelation in the observed prices as confirmed by the exploratory spatial analysis.

Table 1 - Results of Lagrange Multiplier Tests for Each Spatial Weight Matrix.

Weight Matrix	LM_λ	p-value	LM_ρ	p-value	LM_λ^*	p-value	LM_ρ^*	p-value	Recommended model
Rook 1	348.60	0.00%	352.94	0.00%	28.06	0.00%	32.40	0.00%	SAR
Rook 2	0.00	99.05%	1.12	28.90%	0.69	40.50%	1.82	17.76%	OLS
Rook 3	0.06	80.15%	0.79	37.51%	0.98	32.18%	1.71	19.16%	OLS
Queen 1	375.98	0.00%	373.87	0.00%	35.26	0.00%	33.15	0.00%	SER
Queen 2	0.17	68.44%	1.83	17.64%	0.28	59.98%	1.94	16.39%	OLS
Queen 3	0.94	33.17%	0.03	86.36%	1.83	17.63%	0.92	33.85%	OLS
k-NN 5	406.45	0.00%	365.08	0.00%	59.92	0.00%	18.55	0.00%	SER
k-NN 10	580.61	0.00%	491.55	0.00%	122.10	0.00%	33.04	0.00%	SER
k-NN 20	758.82	0.00%	535.58	0.00%	259.57	0.00%	36.33	0.00%	SER
1/d (1320m)	416.50	0.00%	362.51	0.00%	83.06	0.00%	29.07	0.00%	SER
1/d² (1320m)	295.82	0.00%	301.84	0.00%	18.45	0.00%	24.47	0.00%	SAR
1/d (500m)	0.11	73.49%	6.15	1.31%	2.06	15.07%	8.10	0.44%	SAR
1/d (800m)	0.54	46.30%	5.54	1.86%	0.72	39.67%	5.72	1.68%	SAR
1/d (1500m)	404.04	0.00%	341.19	0.00%	89.94	0.00%	27.08	0.00%	SER
1/d (2000m)	352.31	0.00%	280.47	0.00%	94.61	0.00%	22.77	0.00%	SER
1/d (3000m)	259.71	0.00%	194.25	0.00%	81.54	0.00%	16.08	0.01%	SER

The global Moran's I index was calculated on the new residuals for each recommended model. Based on this statistic, models that did not show statistical significance for spatial autocorrelation were selected, as shown in Table 2. Only the lag models with spatial weight matrices using inverse distance with contagion up to 500m and 800m met this criterion. The performance indicators used for the final selection were MAE, MAPE, RMSE, coefficient of determination (R^2), and AIC.

Table 2 – Performance Metrics and Global Moran's I Index Results for the Spatial Regression Residuals

Modelo	Moran	p-value	MAE	MAPE	RMSE	R²	AIC
SAR - Matriz Rook 1	0.07	0.00	322.48	30.41	606.73	0.66	65,198.45
SAR - Matriz 1/d (500m)	0.00	0.28	331.25	31.52	618.75	0.64	65,361.70
SAR - Matriz 1/d (800m)	0.00	0.36	331.11	31.53	618.93	0.64	65,364.16
SAR - Matriz 1/d² (1320m)	0.10	0.00	321.73	30.29	605.54	0.66	65,182.09
SER - Matriz Queen 1	0.15	0.00	332.17	31.56	621.64	0.64	65,400.54
SER - Matriz k-NN (5)	0.16	0.00	332.18	31.53	621.40	0.64	65,397.27
SER - Matriz k-NN (10)	0.13	0.00	332.54	31.62	621.18	0.64	65,394.38
SER - Matriz k-NN (20)	0.11	0.00	333.61	31.81	621.95	0.64	65,404.65
SER - Matriz 1/d (1320m)	0.09	0.00	334.47	32.00	621.38	0.64	65,397.02
SER - Matriz 1/d (1500m)	0.09	0.00	334.95	32.05	621.63	0.64	65,400.39
SER - Matriz 1/d (2000m)	0.07	0.00	336.12	32.20	621.89	0.64	65,403.92
SER - Matriz 1/d (3000m)	0.05	0.00	336.96	32.39	621.03	0.64	65,392.39

Figure 6 shows the Moran's I index almost as a horizontal line (index calculated close to zero). This implies that the residuals of the SAR spatial regression are randomly dispersed (absence of spatial autocorrelation), indicating that the chosen matrix has managed to capture all the spatial dependence observed in the prices.

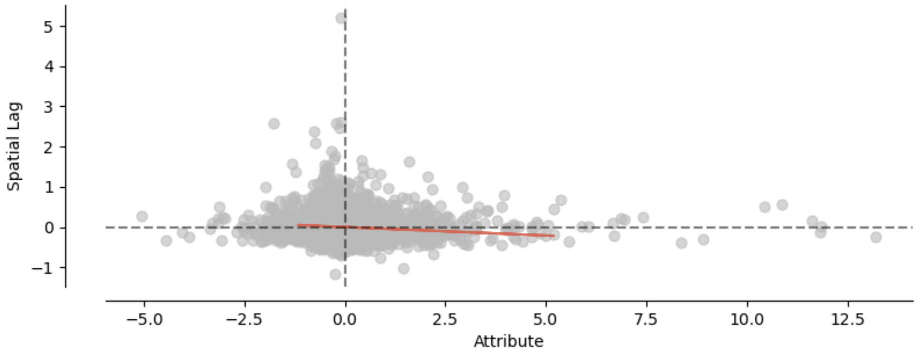

Figure 6 - Scatter Plot of Moran's I Spread for the SAR Model Residuals Using the 1/d (500m) Matrix

In analyzing the parameters, we calculated the direct and indirect effects, as shown in Table 3. The indirect effects, which reflect a property's unit price's influence on neighboring properties' prices, accounted for 5.04% of the total effect. The ratio between the indirect effect and the total effect determined this percentage.

It is essential to highlight that the direct effect corresponds to the coefficient value associated with the specific attribute. These coefficients for both the spatial lag and the classical models are detailed in Table 4. The values were significant and logical, considering the expected positive or negative signs. For example, attributes related to distance, such as Avenida Beira Mar (a major tourist road in the Meireles neighborhood), prime locations, and main roads, exhibited negative signs, as anticipated. This is because plots closer to these areas tend to be more valuable. Similarly, the area attribute also showed a negative sign, indicating that larger plots generally have lower unit prices, a phenomenon explained by the law of diminishing marginal utility. Although this relationship does not uniformly apply to all area ranges studied, it is generally valid in large samples. The significance of the spatial autoregressive coefficient at 5% (ρ) is also noteworthy.

Another attribute with a negative sign is *pzpa1*, which indicates the percentage of the plot located in a preservation area where construction is not allowed. As expected, the unit price of the property tends to increase as the proportion of this preservation area decreases. Other variables, such as frontage, corner location, commercial density, and paving, exhibited positive signs, demonstrating a direct relationship between these explanatory variables and the unit price of the property.

Table 3 – SAR Model Effects Using 1/d (500m) Matrix

Variável	**Effects**		
	Direct	**Indirect**	**Total**
lotcond	0.5400	0.0287	0.5687
numfr	0.0537	0.0029	0.0566
esq	0.0456	0.0024	0.0481
test	0.0023	0.0001	0.0024
area	– 0.0878	– 0.0047	– 0.0924
pvacp	0.1022	0.0054	0.1076
gp	0.0991	0.0053	0.1043
iamaxeq	0.0928	0.0049	0.0977
pzpa1	– 0.2610	– 0.0139	– 0.2748

Table 4 – Comparison of Parameter Estimates Between the OLS and the SAR Model Using the 1/d (500m) Matrix and Total Effect

	Classic (OLS)	**SAR (matrix 1/d 500m)**	**Spillover effect incorporation**
Variável	**Coeficient**	**Coeficient**	**Total effect**
CONSTANT	7.99313	7.65341	
x_	0.03312	0.03357	0.03536

y_	0.10751	0.10788	0.11364
x2_	-0.22905	-0.22859	-0.24078
xy_	-0.27392	-0.27304	-0.28761
lotcond	0.54341	0.53991	0.56870
numfr	0.05318	0.05370	0.05656
esq	0.04661	0.04564	0.04807
f_inc	0.07903	0.07753	0.08167
test	0.00230	0.00230	0.00242
ln_area	-0.08828	-0.08774	-0.09242
pvacp	0.10200	0.10216	0.10761
gp	0.09916	0.09903	0.10432
ln_renda	0.18767	0.18798	0.19800
dv	0.72383	0.72500	0.76366
ln_dscom	0.11313	0.11331	0.11935
acasamed	0.85567	0.85200	0.89744
iamaxeq	0.09269	0.09275	0.09770
pzpa1	-0.25977	-0.2609	-0.27481
distpv	-0.00018	-0.00018	-0.00019
ln_distbm	-0.18718	-0.18691	-0.19688
ln_distvp	-0.10777	-0.10774	-0.11349
a2020	0.07249	0.07162	0.07544
a2021	0.10779	0.10867	0.11447
a2022	0.19005	0.19022	0.20036
a2023	0.29844	0.29677	0.31260
a2024	0.30775	0.30593	0.32224
tr	0.10026	0.10140	0.10680
of	0.27999	0.28044	0.29540
ρ		0.05063	

In determining factors for attributes with dichotomous or discrete quantitative variables, each possible scenario receives a calculated and distinct factor. For the esquina attribute, for instance, assuming the standard case where the property is mid-block with a corner variable value of zero, the factor is calculated using the equation (15).

$$f_{esq(1)} = e^{(total\ eff.)[appraised\ -\ paradigm]} = e^{0.0481[1-0]} = 1.05 \tag{15}$$

Since the dependent variable is transformed using the natural logarithm, for dichotomous explanatory variables, the total effect directly indicates the percentage variation the dependent variable undergoes due to changes in the explanatory attribute. In this case, it is expected that the unit price of a corner property is, on average, 5% higher than that of a non-corner property. For continuous variables, the factor for each attribute is calculated by applying the equation with values from the assessed lot and the paradigm lot. In the specific case of the area attribute, which was transformed using the natural logarithm, the calculation follows equation (16), illustrating the factor calculation for a lot with an area of 1,125m², simplified by the fact that the natural logarithm appears in the exponent of Euler's number.

$$f_{area(1.125m^2)} = e^{(total\ eff.)[\ln(appraised) - \ln(paradigm)]} = \left(\frac{appraised}{paradigm}\right)^{total\ eff.} = \left(\frac{1125}{400}\right)^{-0.0924} = 0.91 \tag{16}$$

For the other continuous variables, which were not transformed, such as *test* and *pzpa1*, the factor is calculated according to equation (17), which demonstrates the calculation of the test factor for a lot with a 30m frontage.

$$f_{test(30m)} = e^{(total\ eff.)[(appraised) - (\text{paradigm})]} = e^{(0.00242)[(30) - (12)]} = 1.04 \tag{17}$$

6. CONCLUSIONS

This study presented an innovative methodological approach for the mass appraisal of urban land, utilizing spatial econometrics techniques, and demonstrated its applicability and effectiveness in the context of the Municipality of Fortaleza. Incorporating spatial weight matrices, Lagrange multiplier tests, and spatial regression models proved crucial for understanding and capturing the spatial structure of the observed land prices. The use of different types of spatial weight matrices and the subsequent selection of the most appropriate spatial model significantly impacted minimizing residual autocorrelation, as evidenced by applying the global Moran's I index, but only for some preselected spatial weight matrices. This highlights the importance of carefully selecting the spatial weight matrix to ensure the accuracy of mass appraisal models.

The study also emphasized the importance of extracting homogenization factors using a more rigorous scientific approach. By extracting these factors from the SAR spatial model, considering both direct and indirect spatial effects, it was possible to provide more refined adjustments to the assessed land values, considering not only the individual characteristics of each lot but also the impact of their spatial attributes. Using "homogenization factors" for lots in property appraisal through spatial regression proves a viable methodology for simplifying mass appraisals and developing homogenization tables by scientific methods. The estimated factors represent a price variation concerning an adopted paradigm lot that is representative of the existing lots in the municipality. This approach adjusts land prices more fairly and allows for more equitable taxation, better reflecting price variations due to locational and contextual factors.

However, this study has some limitations despite the methodological advances and significant practical implications. Although it is an established method for mass property appraisal, it is essential to note that applying factors for homogenization or projection of values based on attribute variation about a paradigm lot in a large municipality like Fortaleza relies on a solid premise. This premise is that the variation in the studied attributes impacts the unit price in the same way for different regions of the city and different economic classes, which may not be the case. Additionally, the dependence on the availability and quality of georeferenced data and the subjectivity in choosing spatial weight matrices can influence the results.

Finally, it is recommended that future research explore the application of these techniques in other municipalities and contexts, adapting the methodology to different data configurations and urban characteristics. This would further validate the proposed method and contribute to a deeper understanding of the applicability and limitations of econometric models.

REFERÊNCIAS

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

______. **Local Indicators of Spatial Association—LISA**. Geographical Analysis, v. 27, n. 2, p. 93-115, 1995.

______; REY, Sergio J. **Modern Spatial Econometrics in Practice**: a guide to geoda, geodaspace and pysal. Chicago: Geoda Press LLC, 2014. 394 p.

ALMEIDA, Eduardo. **Econometria espacial aplicada**. Campinas, SP: Editora Alínea, 2012.

BANDEIRA, Sandro Ricardo Vasconcelos. **Regressão Espacial e Avaliação de Terrenos**: um estudo de caso para a cidade de Fortaleza-ce. 2020. 69 f. Dissertação (Mestrado) - Curso de Economia, Faculdade de Economia, Administração, Atuária e Contabilidade, Universidade Federal do Ceará, Fortaleza, 2019.

CAMPOS, Rodger Barros Antunes; ALMEIDA, Eduardo Simões de. Decomposição Espacial nos Preços de Imóveis Residenciais no Município de São Paulo. **Estudos Econômicos** (São Paulo), vol. 48, n. 1, p. 5-38, jan.-mar. 2018.

CERINO, Rocío Mariel; CARRANZA, Juan Pablo; PIUMETTO, Mario; BULLANO, Emilia; MONZANI, Federico; CÓRDOBA, Mariano. **Homogeneización del valor de la tierra urbana mediante técnicas de econometría espacial en valuaciones masivas automatizadas**. 2022. Disponível em: < https://rdu.unc.edu.ar/handle/11086/28446 >. Acesso em: 2 jul. 2024.

CORRADO, Luisa e FINGLETON, Bernard. Where is the economics in spatial econometrics**? Journal of Regional Science**, v. 52, n. 2, p. 210–239, 2012.

DANTAS, R. A.; ROCHA, F. J. S.; MAGALHAES, A. M.; LIMA, R. C. **Uma nova metodologia para avaliação de imóveis utilizando regressão espacial**. 2003.

DANTAS, Rubens A.. **Modelos Espaciais Aplicados ao Mercado Habitacional**. Tese de doutorado, Universidade Federal de Pernambuco, 2008.

FLORENCIO, Lutemberg de Araújo. **Engenharia de Avaliações com Base em Modelos Gamlss**. Dissertação (Mestrado em Estatística) — Universidade Federal de Pernambuco, Recife, 2010. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/6227/1/arquivo616_1.pdf>. Acesso em: 03 jul. 2024.

GOODCHILD, Michael. **The Validity and Usefulness of Laws in Geographic Information Science and Geography**. In: ASSOCIATION OF AMERICAN GEOGRAPHERS, 94., 2004, Washington, DC. Anais [...] . Oxford: Blackwell Publishing, 2004. v. 2, p. 300-303.

GOLGHER, André Braz. **Introdução à Econometria Espacial**. Jundiaí: Paco Editorial, 2015. 384 p.

______; VOSS, P. R. **How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects**. Spat Demogr, v. 4, p. 175–205, 2016.

KOBYLIŃSKA, Katarzyna. The Application of Spatial Autoregressive Models for Analyzing the Influence of Spatial Factors on Real Estate Prices and Values. **Real Estate Management and Valuation**, [S.l.], v. 29, n. 4, p. 23-35, dez. 2021. DOI: 10.2478/remav-2021-0027.

OLIVEIRA, Antônio Augusto Ferreira de; OLIVEIRA, Terezinha Ferreira de; BANDEIRA, Sandro Ricardo Vasconcelos. **Avaliação em massa com modelo aditivos generalizados e geoestatística.** In: Congresso de Cadastro Multifinalitário e Gestão Territorial, 14., 2020, Florianópolis. Anais [...]. Florianópolis: Cobrac, 2020. Disponível em: https://www.ocs.cobrac.ufsc.br/index.php/cobrac/cobrac2020/paper/view/854/358. Acesso em: 03 jul. 2024.

OLIVEIRA, A. A. F., PRADO, J. F., BANDEIRA, S. R. V, SOUZA, H. A., BRITO, J. A., NOBERTO, L. V. V. **Análise da dependência espacial na avaliação em massa de casas em Aracaju, Brasil**. 2023.
Disponível em: < https://revistavalorem.com/index.php/home/article/view/9 >. Acesso em: 3 jul. 2024.

PELLI SISTEMAS. **Sisdea**. Versão 1.62.6.2.

REY, Sergio J.; ANSELIN, Luc**. PySAL**: A Python Library of Spatial Analytical Methods. The Review of Regional Studies, v. 37, n. 1, p. 5-27, 2007.

ROSEN, Sherwin. Hedonic prices and implicit markets: product differentiation in pure competition. **Journal of Political Economy**, Chicago, v. 82, n. 1, p. 34-55, 1974.

SEFIN (org.). **Infraestrutura de dados espaciais da SEFIN**: preço de mercado de terrenos. Preço de mercado de terrenos. 2024. Elaborado por Secretaria das Finanças do Município de Fortaleza. Disponível em: https://ide.sefin.fortaleza.ce.gov.br/. Acesso em: 01 jul. 2024.

ZILLI, Carlos Augusto; HOCHHEIM, Norberto. **Avaliação em massa de imóveis para fins de imposto predial: abordagem utilizando regressão espacial global e local**. Valorem: Revista da Sociedade Brasileira de Engenharia de Avaliações, v. 2, n. 2, p. 10-24, 2023. Disponível em: http://orcid.org/0000-0002-4926-1186. Acesso em: 24 jul. 2024.

APPENDIX

Table 5 – Equations for area, frontage, and pzpa1 factors the factors

Area factor	Frontage factor	Pzpa1 Factor
$\left(\frac{A_{app}}{400}\right)^{-0,0924}$	$e^{(0,00242)[\ (\mathrm{test}_{app})-\ (12)]}$	$e^{-0,27481[\ (\mathrm{pzpa1}_{app})-\ (0)]}$
Where: A_{app} is the area of the appraised property (m²).	Where: $test_{app}$ is the front of the appraised property (m)	Where: $pzpa1_{app}$ is the proportion of preservation area of the assessed property

Table 6 - Paving, corner, storm drain and condominium factor

Value	Paving Factor (pvacp)	Corner Factor	Stormwater Gallery Factor	Condominium Factor
0 (absence)	0.87	1.00	0.89	1.00
1 (presence)	1.00	1.04	1.00	1.57

Table 7 – Nº of frontages factor

Nº of frontages	Factor
1	1.00
2	1.05
3	1.10
4	1.16
5	1.21
6 or more	1.27

Table 81 – Floor area ratio factor

Legal land utilization index	Factor
From 0 to 1	0.72
From 1 to 2	0.85
From 2 to 3	1.00
3	1.18