

## ANÁLISE DA DEPENDÊNCIA ESPACIAL NA AVALIAÇÃO EM MASSA DE CASAS EM ARACAJU, BRASIL

*Analysis of spatial dependence in the mass appraisal of houses in Aracaju, Brazil*

**Antônio Augusto Ferreira de Oliveira**

<http://orcid.org/0000-0003-3890-4219> 

Secretaria das Finanças do Município de Fortaleza (SEFIN), Fortaleza, Brasil  
augusto.oliveira@sefin.fortaleza.ce.gov.br

**João Freire Prado**

<http://orcid.org/0000-0000-0000-0000> 

Secretaria Municipal da Fazenda de Aracaju (SEMFAZ), Aracaju, Brasil  
joao.prado@aracaju.se.gov.br

**Sandro Ricardo Vasconcelos Bandeira**

<http://orcid.org/0000-0003-2730-025X> 

Secretaria das Finanças do Município de Fortaleza (SEFIN), Fortaleza, Brasil  
sandro.bandeira@sefin.fortaleza.ce.gov.br

**Héverton Alves de Souza**

<http://orcid.org/0000-0000-0000-0000> 

Secretaria Municipal da Fazenda de Aracaju (SEMFAZ), Aracaju, Brasil  
heverton.souza@aracaju.se.gov.br

**Jamille Almeida Brito**

<http://orcid.org/0000-0002-5208-896X> 

Secretaria Municipal da Fazenda de Aracaju (SEMFAZ), Aracaju, Brasil  
jamille.brito@aracaju.se.gov.br

**Luan Victor Vasconcelos Noberto**

<http://orcid.org/0000-0002-9810-6780> 

Secretaria das Finanças do Município de Fortaleza (SEFIN), Fortaleza, Brasil  
luan.noberto@sefin.fortaleza.ce.gov.br

### RESUMO

Este trabalho traz à baila o dilema de escolha de matrizes de ponderação espacial nos modelos econométricos espaciais aplicados à avaliação em massa de imóveis, de tal forma a capturar a máxima dependência espacial observada, da mesma forma que garanta qualidade das avaliações finais e testes de robustez/sensibilidade do modelo final frente a essa escolha. Partiu-se de uma amostra de 3.580 dados de preços de casas no Município de Aracaju, todas georreferenciadas, o que permitiu uma análise exploratória onde se evidenciou a dependência espacial daqueles preços, bem como dos resíduos do modelo de regressão linear clássica. Ficou evidenciado que as matrizes de ponderação espacial baseadas na distância promovem melhor ajustes nos modelos econométricos espaciais, seja pela melhor incorporação da dependência espacial, seja pelo melhor resultados nos critérios de AIC e raiz do erro quadrático médio.

**Palavras-Chave:** Avaliação em massa; Modelos econométricos espaciais; Matrizes de ponderação espacial.

### ABSTRACT

This article brings up the dilemma of choosing different spatial weighting matrices in spatial econometric models applied to the mass appraisal of real estate, in such a way as to capture the maximum observed spatial dependence, in the same way as to guarantee the quality, sensitivity and robustness of the final appraisals. We started with a sample of 3,580 data on house prices in the Municipality of Aracaju, which allowed an exploratory analysis where the spatial dependence of those prices was evidenced, as well as on the residuals of the classical linear regression model. It was evidenced that the spatial weighting matrices based on distance promote better adjustments in the spatial econometric models, either by better incorporation of spatial dependence, or by better results in the AIC criteria and root mean square error.

**Keywords:** Mass appraisal of real estate; Spatial econometric models; Spatial weighting matrices.

Preenchimento dos Editores

### INFORMAÇÕES SOBRE O ARTIGO

Submetido em 15/04/2022  
Publicado em 15/06/2022

Comitê Científico Interinstitucional  
Editor-Responsável: Carlos Augusto Zilli  
(SEER/OJS – Versão 3)



## 1. INTRODUÇÃO

Todo processo estocástico espacial está sujeito à “Primeira Lei da Geografia” de Waldo Tobler: “todas as coisas estão relacionadas com todas as outras, mas coisas próximas estão mais relacionadas do que coisas distantes”. O comportamento dos preços no mercado imobiliário não foge a essa regra, sendo fato inconteste a autocorrelação espacial deles. A NBR 14.653-2:2011, no seu anexo C, recomenda a incorporação dos efeitos de dependência espacial no próprio modelo clássico de regressão linear, “com o objetivo de assegurar as estimativas características de não tendenciosidade, eficiência e consistência” (item C.2.2). Neste mesmo anexo normativo, encontra-se o destaque da utilidade dos modelos com tratamentos de dados por regressão espacial quando aplicados às “avaliações em massa plantas de valores genéricos, estudos de velocidades de vendas e de demandas habitacionais, entre outros”. Este trabalho cuida da avaliação em massa para casas isoladas da cidade de Aracaju no enfoque da econometria espacial.

O conceito formal de autocorrelação espacial, bem como as duas abordagens mais comuns de tratamento dela, serão apresentados a partir da seção 2.3. Por ora, cumpre observar que a abordagem econométrica espacial se baseia na incorporação da dependência espacial no próprio modelo de regressão clássico. Na prática, essa incorporação se dá por meio das matrizes de ponderação espacial. A definição destas está sob o arbítrio do pesquisador/avaliador e, a partir desta escolha, fundamentam-se todos os demais testes estatísticos para a escolha do melhor modelo econométrico espacial. Tão importante é a importância desta escolha, que Almeida (2012), citando Leenders (2002), afirma que “qualquer modelo espacial é inútil se a matriz de pesos for mal especificada”. Dubin (1988), citado por Golgher (2015), afirma que diversas matrizes foram especificadas para o estudo do mercado imobiliário, onde se observou diferentes resultados do modelo final, conforme o tipo escolhido, indicando, desta forma, a alta sensibilidade do modelo a esta escolha.

A partir do procedimento determinado por Baumont (2004), citado por Almeida (2012), este trabalho objetiva analisar as possíveis especificações de matrizes de ponderação espacial que capturem a maior dependência espacial possível observada nos preços das casas de Aracaju, bem como a sensibilidade das previsões dos modelos finais frente à essa escolha.

## 2. REFERENCIAL TEÓRICO

### 2.1. DIAGNÓSTICO DA AUTOCORRELAÇÃO ESPACIAL

A autocorrelação espacial pode ser entendida pelo fenômeno de que o valor da variável de interesse, que no caso de avaliação de imóveis é o preço observado do imóvel, em dada localização  $i$ , tende a estar associada ao valor desta mesma variável nas localizações vizinhas  $j$ . A autocorrelação espacial pode ser expressa em termos da matriz de covariância:

$$\text{Cov}(y_i, y_j) = E(y_i y_j) - E(y_i) E(y_j) \neq 0, \text{ para } i \neq j \quad (1)$$

Segundo Almeida (2012), existem duas formas de análise da dependência espacial: i) abordagem direta da geoestatística, por meio da análise variográfica e ii) abordagem pela econometria espacial, onde se procura incorporar os efeitos daquela dependência no próprio modelo econométrico.

É cediça a lição que de a econometria espacial se baseia na análise de autocorrelação espacial realizada sobre os resíduos da regressão linear clássica com estimação dos parâmetros pelo método dos mínimos quadrados ordinários (MQO). Entretanto, não se pode olvidar a importância da análise prévia de comportamento espacial da variável dependente (e independentes) no contexto geográfico como um todo. A NBR 14653-2:2011, no anexo C, estabelece como diagnóstico da autocorrelação espacial as duas abordagens supracitadas, bem como a análise do gráfico (sic) espacial dos resíduos do modelo econométrico (alínea a). A bem da verdade, pode-se estudar a dependência espacial da própria variável dependente, bem como suas variáveis explicativas, num procedimento conhecido como análise exploratória de dados espaciais (exploratory spatial data analysis – ESDA), e não só por meio da “especialização dos resíduos do modelo” e a realização de testes estatísticos para determinar o modelo econométrico espacial mais adequado. Esta etapa é sim necessária, pois pode indicar vieses nas estimativas geradas no modelo tradicional, mas o que estamos a defender é uma exploratória espacial ampla e prévia de tal dependência, inclusive com um estudo a priori sobre qual matriz de ponderação espacial adotar.

## 2.2. ABORDAGEM GEOESTATÍSTICA - ANÁLISE VARIOGRÁFICA

Em termos simples, o semivariograma é uma função que reflete as relações espaciais (por meio de suas covariâncias) entre pares de dados espaçados por uma distância  $h$  dado pela seguinte expressão:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n \{ [Z(y_i + h) - Z(y_i)]^2 \} \quad (2)$$

onde  $n$  é o número de pares de pontos amostrais e  $y_i$  e  $(y_i + h)$  são os valores da variável de interesse separados por uma distância  $h$ .

Uma maneira de se comprovar a presença de autocorrelação espacial da variável de interesse é tentar ajustar um semivariograma teórico sobre esta. Se o ajuste for satisfatório, há o indicativo da presença de autocorrelação espacial.

Entende-se por semivariograma satisfatoriamente ajustado quando este apresente uma curva de crescimento acentuada (“arranque”), a partir do efeito pepita, constituindo um campo estruturado, até um determinado limite (alcance<sup>1</sup>), onde se estabiliza, produzindo um campo aleatório (Figura 1). A presença deste campo estruturado indica que, nessa região, as semivariâncias dos pares de dados próximos são pequenas, aumentando conforme a distância  $h$  escolhida.

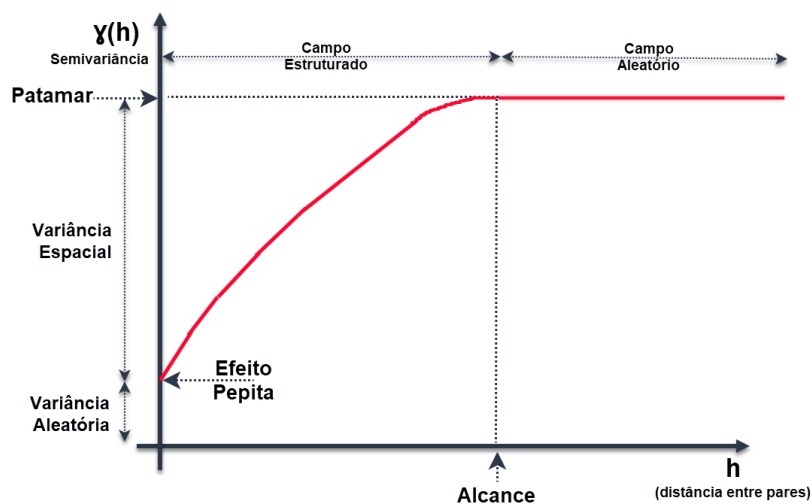


Figura 1 - Feições principais de um semivariograma (adaptado de Andriotti, 2009 e Yamamoto, 2020).

## 2.3. ABORDAGEM ECONOMÉTRICA ESPACIAL

Como já mencionado, as abordagens econométricas espaciais baseiam-se na incorporação da autocorrelação espacial envolvida no fenômeno em estudo no próprio modelo econométrico. Isso se traduz, em outras palavras, na introdução do lado direito da equação de regressão (como variável endógena) de operadores de defasagens espaciais obtidos por meio de matriz(es) de ponderação espacial, comumente representada pela letra  $W$ .

### 2.3.1. Matrizes de Ponderação Espacial

A escolha da matriz de ponderação espacial  $W$  é um dos passos mais críticos da econometria espacial (ALMEIDA, 2012; ANSELIN, 1995 apud ALMEIDA, 2012; GOLGHER, 2015, dentre outros). É uníssona, na lição destes autores, que a matriz  $W$  deve ser aquela que consiga capturar a máxima dependência espacial do fenômeno estudado, haja vista que ela será usada na obtenção das variáveis endógenas  $W_y$  e  $W_\varepsilon$  que são utilizadas nos principais modelos econométricos a serem apresentados na seção 2.4. Em linhas gerais, as matrizes de ponderação espacial podem ser divididas em dois grandes grupos: matrizes de contiguidade e matrizes de distância, conforme figura 2. Cada uma delas pode ser subdividida em outras formas de mensurar a interação entre os vizinhos.

<sup>1</sup> É a distância a partir da qual a variação média entre duas observações não é mais função da distância entre elas, indicando ausência de autocorrelação espacial.

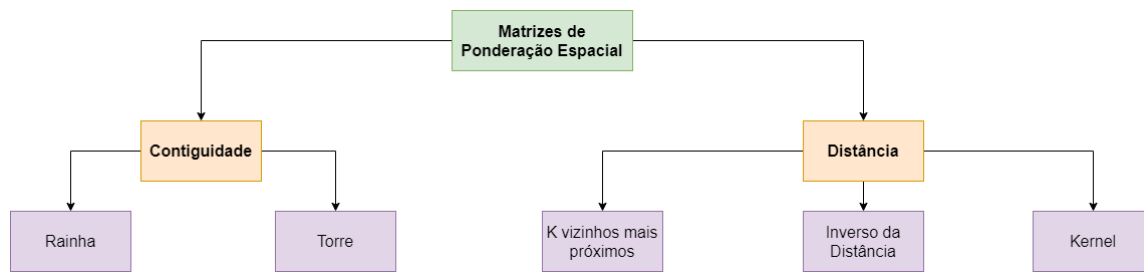


Figura 2 - Principais matrizes de ponderação espacial citadas na literatura.

Almeida (2012) estabelece um conjunto de propriedades “desejáveis” para as matrizes de ponderação espacial: i) os pesos espaciais devem ser finitos e não negativos ( $0 < w_{ij} < \infty$ ); ii) a matriz não deve possuir ilhas, devendo todo elemento ter no mínimo uma interação com pelo menos um vizinho ( $\sum w_{ij} \neq 0$ ); iii) sua diagonal principal deve ser nula, pois nenhuma localização pode interagir consigo mesma ( $w_{ii} = 0$ ) e iv) seus pesos devem ser exógenos, no sentido que estes não devem influenciar a variável dependente ( $E(w_{ij}, \varepsilon) = 0$ ).

Neste trabalho, utilizar-se-ão a matriz de contiguidade do tipo rainha obtida pela tesselação de Voronoi e as seguintes matrizes de distância euclidiana:

- 1) matriz ponderada pelo inverso da distância,  $1/d_{ij}$ ;
- 2) matriz ponderada pelo inverso da distância ao quadrado  $1/d_{ij}^2$ ;
- 3) matriz de ponderação bisquare  $(1 - d_{ij}^2/d_{MAX}^2)^2$  e;
- 4) matriz de ponderação gaussiana  $\exp(-0,5 \times d_{ij}^2/d_{MAX}^2)$ .

Todas as matrizes serão padronizadas por linha, assim, cada elemento é obtido pela razão entre o seu valor e a soma de todos os demais valores da respectiva linha a que pertence, ou seja:

$$w_{ij}^S = \frac{w_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \tag{3}$$

Outro aspecto de destaque nesse trabalho é a definição das distâncias de cut-off ( $d_{max}$  na equação 4) para cada matriz de peso espacial a ser testada. Essa é a distância a partir do qual o peso de interação entre as localizações será zero. Por exemplo, no caso da matriz ponderada pelo inverso da distância:

$$w_{ij} = \begin{cases} 1/d_{ij} & \text{se } d_{ij} \leq d_{max} \\ 0 & \text{se } d_{ij} > d_{max} \end{cases} \tag{4}$$

### 2.3.2. Índice Global de Moran I

O diagnóstico de dependência espacial pode ser obtido, inicialmente, pelo Índice global de Moran I. Ele tem como hipótese nula a independência espacial (nesse caso assume o valor zero). Valores positivos do índice indicam correlação direta e valores negativos, correlação inversa. Esse teste é realizado sobre os resíduos da regressão clássica pelo MQO<sup>2</sup>. Entretanto, ele não indica qual modelo espacial é o mais adequado (ANSELIN; REY, 2014). Formalmente a estatística desse teste é dada por:

$$I = \frac{n}{\sum_{i=1}^n (y_i - \bar{y})^2} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \tag{5}$$

onde n é o número de observações, y é a variável estudada,  $\bar{y}$  é a sua média e  $w_{ij}$  é o elemento na matriz de vizinhança correspondente ao par de elementos i e j. Deve-se observar que se a matriz for normalizada por linha, o termo  $\sum_{i=1}^n \sum_{j=1}^n w_{ij}$  é igual a 1. Impende observar, novamente, que o índice de Moran I depende da escolha prévia (e arbitrária) da matriz de ponderação espacial.

<sup>2</sup> Nada impede que se calcule do índice de Moran I sobre a variável dependente ou mesmo variáveis independentes para uma análise preliminar e exploratória das dependências espaciais.

### 2.3.3. Multiplicadores de Lagrange LM (defasagem e erro) e LM Robusto (defasagem e erro).

Como vimos, o índice global de Moran I não indica qual modelo espacial é o mais adequado. Para esse fim, devemos calcular os testes LM de multiplicadores de Lagrange, referentes à variável dependente defasada espacialmente e o erro defasado espacialmente ou ambos (misto). Os testes LM são testes de hipótese, que seguem uma distribuição qui-quadrada, com um grau de liberdade. A hipótese nula para o teste do LM (erro) é a de não existência de autocorrelação espacial no termo do erro ( $\lambda = 0$ ). Para o teste LM (defasagem), é a inexistência de defasagem espacial na variável dependente ( $\rho = 0$ ).

Os dois testes podem ser calculados pelos resíduos da regressão linear multivariada com estimativa MQO, segundo as expressões abaixo.

$$LM_{\lambda}(\text{erro}) = \frac{\left[ \frac{e'W_e}{(\hat{\sigma}_{ML})^2} \right]^2}{n \cdot [\text{tr}(W^2 + W'W)]} \approx X^2 \quad (6)$$

$$LM_{\rho}(\text{defasagem}) = \frac{\left[ \frac{e'W_e}{(\hat{\sigma}_{ML})^2} \right]^2}{\left[ (W X \beta)' \frac{(M W X \beta)}{(\hat{\sigma}_{ML})^2} + \text{tr}(W^2 + W'W) \right]} \approx X^2 \quad (7)$$

onde:

n: número de dados da amostra.

e: vetor dos resíduos obtidos pelo método dos mínimos quadrados (MQO);

W: Matriz de pesos espaciais (W' é a matriz transposta);

$\hat{\sigma}_{ML} = e'e/n$ : estimativa da máxima verossimilhança da variância do modelo  $y = X\beta + \varepsilon$ ;

tr: operador de traço da matriz;

y: vetor de observação da variável dependente;

X: matriz de observação das variáveis independentes;

$\beta$ : vetor dos parâmetros das variáveis independentes (modelo MQO);

$M = I - X(X'X)^{-1}X'$ , onde I é a matriz identidade.

Anselin (1996) (apud ANSELIN; REY, 2014) demonstrou que o teste clássico  $LM_{\rho}$  (defasagem) também é sensível à presença de autocorrelação espacial nos erros ( $\lambda \neq 0$ ). Da mesma forma, o teste  $LM_{\lambda}$  é sensível à presença de autocorrelação espacial na própria variável dependente ( $\rho \neq 0$ ). Desta feita, esse autor desenvolveu novos testes ( $LM_{\rho}$  e  $LM_{\lambda}$  robustos) que corrigem esse comportamento. Conforme se verá na seção 2.5, estes testes robustos serão usados para distinguir o melhor modelo econométrico espacial quando ambos os testes clássicos,  $LM_{\rho}$  e  $LM_{\lambda}$ , forem significantes.

## 2.4. PRINCIPAIS MODELOS ECONOMÉTRICOS ESPACIAIS

Nesta seção, apresentar-se-á os três principais e mais utilizados modelos econométricos espaciais: SEM, SAR e SAC. Oportuno observar que a NBR 14653-2:2011, no seu anexo C, item C.3.2.2, só se refere aos dois primeiros, quais sejam, aos modelos de erro espacial (SEM) e ao modelo de defasagem espacial (SAR). Não há qualquer referência ao modelo “misto” proposto pelo Kelejian-Prucha (SAC), nem outros modelos citados na literatura (que não serão abordados nesse trabalho), tais como o modelo espacial de Durbin (SDM), o de erro espacial de Durbin (SDEM) e modelo de Manski<sup>3</sup>.

### 2.4.1. Modelo de Erro Espacial (SEM)

O modelo de erro espacial ou Spatial Error Model (SEM) é expresso formalmente por:

$$y = X \cdot \beta + u \quad \text{onde } u = \lambda W_u + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n) \quad (8)$$

onde W é a matriz de vizinhança,  $W_u$  é um vetor n por 1, resultado da multiplicação da matriz de vizinhança

<sup>3</sup> Demais modelos econométricos espaciais podem ser vistos em Golgher, 2015.

com o vetor dos resíduos,  $X$  é a matriz das variáveis independentes,  $\lambda$  e  $\beta$  são os parâmetros a serem estimados e  $\varepsilon$  o vetor de resíduos estocásticos.

A equação 8 pode ser reescrita da seguinte forma após algumas transformações algébricas ( $I_n$  é a matriz identidade de ordem  $n$ ):

$$y = X \cdot \beta + (I_n - \lambda \cdot W)^{-1} \cdot \varepsilon \quad (9)$$

Como se verá, diferentemente do modelo de defasagem espacial (SAR), a interação espacial ocorre nos resíduos devido a efeitos não modelados e que não estão distribuídos aleatoriamente no espaço (ALMEIDA, 2012). Outra hipótese de autocorrelação espacial nos erros está no fato de utilização de variáveis proxies resultantes da divisão artificial das unidades geográficas da área em estudo (DANTAS, 2003). Por exemplo, a utilização da variável renda do responsável por setor censitário (dado do IBGE), ou a utilização do índice de desenvolvimento humano por bairro, sem uma suavização da destas em uma superfície, podem ocasionar erros de medida pelo efeito de transbordamento (a variável se propaga para além da sua fronteira).

Observa-se ainda pela equação 9, que os termos de dependência espacial estão no último termo da equação, dando a possibilidade de se interpretar cada coeficiente  $\beta$  da maneira “tradicional”, ou seja, como o efeito marginal do respectivo regressor sobre a variável dependente (ALMEIDA, 2012).

#### 2.4.2. Modelo de Defasagem Espacial (SAR)

O modelo de defasagem espacial ou Spatial Autoregressive Model (SAR) é dado pela equação abaixo:

$$y = \rho W_y + X \cdot \beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n) \quad (10)$$

onde  $W$  é a matriz de vizinhança,  $W_y$  é um vetor  $n$  por  $1$ , resultado da multiplicação da matriz de vizinhança com o vetor dos valores das variáveis dependentes,  $X$  é a matriz das variáveis independentes,  $\rho$  e  $\beta$  são os parâmetros a serem estimados e  $\varepsilon$  o vetor de resíduos estocásticos.

Conforme supracitado, a matriz de vizinhança tem diagonal principal zero e foi normalizada na linha. Destarte, a linha  $i$  de  $W_y$  representa a ponderação dos preços observados dos vizinhos sobre seu próprio valor. O vetor  $W_y$  também é conhecido como operador de defasagem espacial ou coeficiente autorregressivo espacial.

De plano, observa-se que se tem uma dependência de  $y$  com seus vizinhos, caracterizando a autocorrelação espacial. Fazendo um paralelo com as séries temporais, se na equação 10, não tivéssemos as variáveis exógenas representadas pela matriz  $X$ , teríamos um modelo autorregressivo onde o valor observado na posição espacial  $i$  ( $y_i$ ) se relacionaria com os valores observados de seus vizinhos ( $y_j$ ) (nas séries temporais seria a autocorrelação no tempo  $t$  com os períodos anteriores<sup>4</sup>).

#### 2.4.3. Modelo Kelejjan-Prucha (SAC)

O modelo SAC pode ser considerado um modelo misto onde há presença de dois termos de correlação espacial defasados no lado direito da equação (condição endógena): um referente à variável dependente defasada, outro envolvendo o termo de erro:

$$y = \rho W_y + X \cdot \beta + u \quad \text{onde } u = \lambda W_u + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n) \quad (11)$$

### 2.5. FLUXOGRAMA PARA ESPECIFICAÇÃO DO MODELO ECONOMETRICO ESPACIAL

Anselin e Rey (2014) propõem um conjunto de passos, baseados num conjunto de regras de decisão que foram exitosas em diversos outros trabalhos empíricos anteriores carreados por estes e outros autores, conforme fluxograma adaptado e apresentado na figura 3. Este fluxograma parte do modelo de regressão linear múltipla com estimativa dado pelo MQO, realiza os diagnósticos de dependência espacial por meio

<sup>4</sup> Diferentemente da autocorrelação temporal, onde a direção da causalidade ocorre apenas em um sentido (passado influencia o presente e futuro), a autocorrelação espacial pode ocorrer em várias direções do espaço tridimensional. Fonte: LONGLEY, et al., 2013.

dos testes de multiplicadores de Lagrange, LM-Erro e LM-lag, sejam nas suas formas clássicas, sejam nas suas formas robustas (ANSELIN et al. 1996 apud ANSELIN; REY, 2014)<sup>5</sup>.

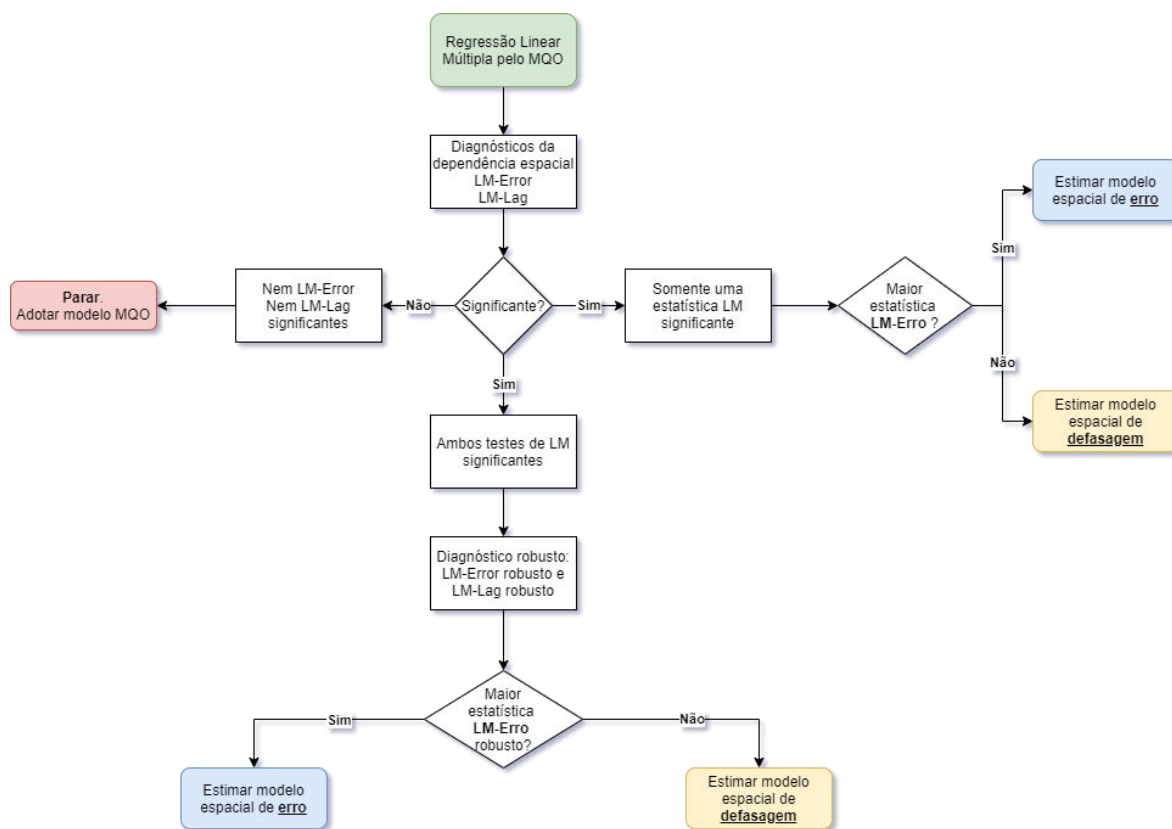


Figura 3 - Fluxograma de especificação do modelo econométrico espacial (adaptado de Anselin e Rey, 2014).

De acordo com o fluxograma proposto, depois de realizados os quatro testes de multiplicadores de Lagrange, pode-se inferir o modelo econométrico espacial mais adequado: i) caso os testes  $LM_p$  e  $LM_\lambda$  não sejam significativos, adota-se o modelo MQO; ii) caso apenas um deles seja significativo, adota-se o modelo respectivo e iii) se ambos forem significativos, realiza-se os testes robustos de defasagem e erro e escolhe-se aquela forma significativa ou com maior estatística (caso em que ambos os testes robustos foram significantes).

Não se pode esquecer que a estatística dos multiplicadores de Lagrange necessita da definição prévia da matriz de ponderação espacial, haja vista que as variáveis  $W_y$  e  $W_\epsilon$  entram nos seus cálculos. Portanto, não obstante o fluxograma bem delineado acima de como proceder na escolha do modelo econométrico mais apropriado, a escolha da matriz de ponderação espacial ainda é uma lacuna a ser preenchida pelo próprio pesquisador.

Explorando essa lacuna, este trabalho objetiva a testar a sensibilidade dos modelos econométricos quanto à escolha de variadas matrizes de ponderação espacial. É o que se verá na seção seguinte.

### 3. METODOLOGIA

#### 3.1. ÁREA DE ESTUDO

A área de estudo do presente trabalho corresponde ao município de Aracaju, capital do estado de Sergipe, que se estende por 182,163 km<sup>2</sup>, em uma área delimitada ao norte pelo rio Sergipe, ao sul pelo rio Vaza-Barris e à leste pelo oceano Atlântico. Segundo projeção do IBGE para o ano de 2020, a população estimada é de 664.908 habitantes, o que corresponde a uma densidade demográfica projetada de 3.140.65 hab./km<sup>2</sup>. Em 2010, por meio do Atlas Brasil, a Organização das Nações Unidas estimou para Aracaju IHDM de 0.770, o que posicionou, à época, o município na 55ª posição no ranking nacional.

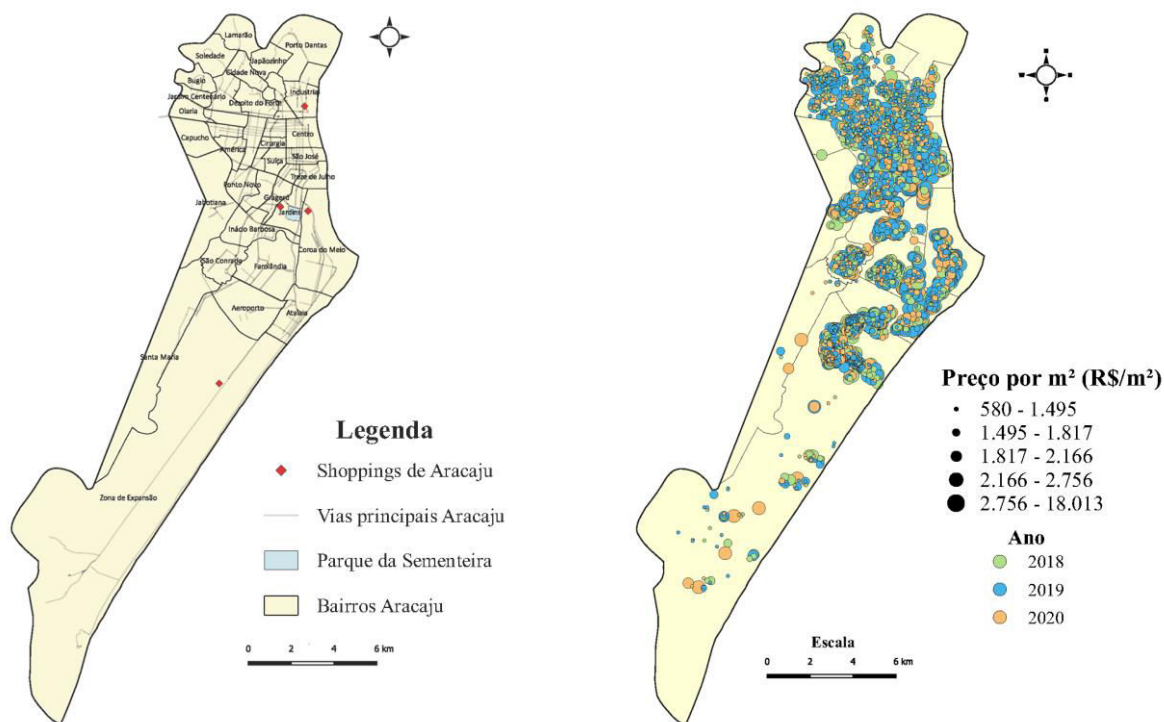
<sup>5</sup> As formas robustas dos testes LM corrigem suas altas sensibilidades à autocorrelação espacial, que podem sugerir uma forma errada de especificação (confusão da dependência entre a defasagem e erro).

Aracaju cresce predominantemente em direção ao sul, cujos bairros de adensamento urbano mais recente são a Zona de Expansão e o bairro Santa Maria. Os 37 demais bairros ocupam todo o meio norte do município. Por estimativa da Secretaria Municipal da Fazenda de Aracaju (SEMFAZ), uma área aproximada de 87,7 km<sup>2</sup> no município está densamente construída, representando um percentual de 48% da área total do território. Além disso, aquela secretaria contabiliza aproximadamente 120.056 lotes cadastrados e 243.080 unidades imobiliárias autônomas. Dentre estas, 21.655, ou aproximadamente 9%, são de unidades ainda não edificadas (terrenos). Com relação às unidades edificadas, verifica-se que 46% das tipologias são de “casas isoladas” (111.310), ao passo que 32,50% são da tipologia “apartamentos” (79.012).

Esse município impõe intensa polarização na integração regional, constituindo-se uma metrópole estadual, possuindo alta interação com o município de Nossa Senhora do Socorro, no setor Noroeste da capital (LUIZ et al., 2012). Recentemente, dois vetores de expansão da cidade tendem a modificar a configuração da cidade: o primeiro ocorre a Sudoeste, em direção a São Cristóvão, e o segundo, a Nordeste, em direção à Barra dos Coqueiros.

### 3.2. AMOSTRA DE DADOS

A amostra de dados desse trabalho é composta por 3.580 informações de preços de casas isoladas (fora de condomínios horizontais) provindas do mercado imobiliário (ofertas e transações), bem como das avaliações e declarações do imposto de transmissão inter vivos (ITBI) realizada no âmbito Secretaria Municipal da Fazenda de Aracaju (SEMFAZ). Os dados foram colhidos nos anos de 2018 a outubro de 2020. A representação espacial destes dados foi por meio do centroide do terreno de localização da casa. O mapa 1 apresenta os bairros, vias principais, shopping centers (mapa à esquerda) e a distribuição espacial da amostra (mapa à direita). Por este último mapa se observam poucos dados na região sul da cidade. Isto pois, a Zona de Expansão cresce por meio de condomínios horizontais de alto padrão, ao passo que casas de baixo padrão construtivo predominam no bairro Santa Maria.



Mapa 1 - Bairros e distribuição espacial da amostra na cidade de Aracaju.

## 4. RESULTADOS E DISCUSSÃO

### 4.1. ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS (ESDA)

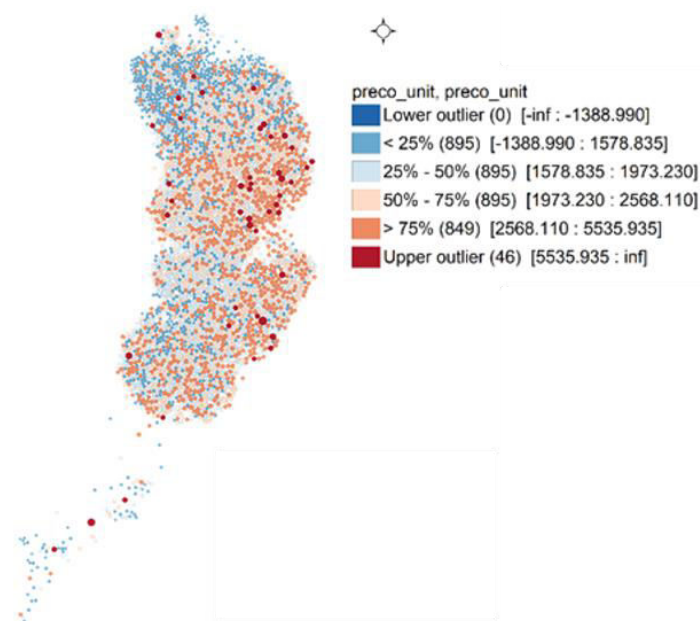
Ressaltou-se na seção 2.1 a importância da ESDA realizada não somente sobre os resíduos do modelo de regressão, mas logo no início do processo avaliatório, antes mesmo de qualquer modelagem. A ESDA



consiste num conjunto de técnicas exploratórias capazes de “identificar localizações atípicas ou outliers espaciais, detectar padrões de associação espacial e clusters além de sugerir regimes espaciais ou outras formas de heterogeneidade espacial” (ERTUR; GALLO, 2003 apud TYSZLER, 2006).

Nesse contexto, esta seção partiu de uma ESDA tendo como variável “alvo” de estudo o preço unitário das casas (R\$/m<sup>2</sup> de área construída), não obstante, como se verá, a escolha da variável dependente sendo o preço total das casas observadas. Isto pois, esta variável tem grande dispersão de valores, sendo preferível diminuir a sua amplitude pela divisão com a área edificada de cada observação para não enviesar as análises obtidas.

O mapa 2 apresenta o cartograma, o qual é um mapa em que cada ponto (centroide do lote da casa) é substituído por círculos proporcionais ao preço unitário observado, apresentando os intervalos interquartis com cores diferentes e as respectivas quantidades de dados em cada classe. No cartograma, ainda temos a representação dos “outliers espaciais” em destaque (vermelho escuro) segundo o critério de estarem fora de uma vez e meia o intervalo interquartil (em relação a Q<sub>1</sub> e Q<sub>3</sub>).



Mapa 2 - Cartograma dos preços unitários observados.

Pode-se notar uma concentração de preços unitários altos das casas a leste e na parte central da cidade, ocupados por bairros como Treze de Julho e Jardins. Considerando-se esta mancha de alto valor em laranja (dados não outliers e acima do terceiro quartil) no sentido Norte-Sul, observam-se ao norte, na fronteira com o centro comercial e histórico de Aracaju, os bairros que possuem uso misto do solo, com a presença marcante de serviço médico-hospitalares privado e especializado, junto a comércio, bares e restaurantes. Ainda nessa área, aparecem os bairros Treze de Julho e Jardins, rodeados por importantes equipamentos de lazer públicos, ressaltando-se, dentre eles, o Parque da Sementeira e Orla da Praia Formosa. Há também centros comerciais e de lazer privado, destacando-se os dois principais shopping centers da capital, Shopping Jardins e Riomar. Essa região da cidade tem sido intensamente verticalizada com edificações de alto padrão, inclusive com o lançamento de algumas tipologias de padrão luxo. Seguindo na direção Sul, ao longo da Foz do Rio Poxim, há bairros como Farolândia e Coroa do Meio que também estão em processo de renovação urbana, porém mais recente, com a presença de novos empreendimentos verticalizados de médio e alto padrão, que se valem da proximidade a serviços de educação, saúde, comércio e lazer, como a Universidade Tiradentes (Farolândia), Shopping Riomar (Coroa do Meio), Parque dos Cajueiros (Farolândia) e Orla de Atalaia, um pouco mais ao Sul.

A visualização da autocorrelação local pode ser feita por meio do gráfico de dispersão local de Moran I (Local Indicators of Spatial Association – LISA), conforme proposto por Luc Anselin (1995). Como é sabido, essa estatística e a global de Moran I necessitam da escolha de uma matriz de ponderação espacial. Para essa segunda ESDA, adotou-se como matriz de ponderação espacial, o inverso da distância entre os vizinhos num raio de até 1.251,26m, com intuito de que todos os pontos tivessem pelo menos um vizinho (minimum threshold distance).

Como a reta de regressão do gráfico 1, é positivamente inclinada, temos uma indicação de autocorrelação espacial positiva, como a grande maioria dos pontos nos quadrantes 1 (alto-alto) e 3 (baixo-baixo). Pode-se ainda se observar em vermelho os valores com estatística de Moran significativos a 5% e representando a autocorrelação espacial alto-alto (high-high) e, em azul escuro, a relação baixo-baixo (low-low). Em azul claro e laranja, estão os pontos situados no segundo e quarto quadrante com autocorrelação espacial negativa. Em cinza, estão as observações não significantes na estatística de Moran I a 5%. Pelo mapa 3, observa-se autocorrelação positiva baixo-baixo na região norte e alto-alto na região central e a leste da cidade (bairros Treze de Julho e Jardins).

Como supracitado, insights do comportamento espacial dos preços de mercado é quase como condição *sine qua non* de uma boa modelagem avaliatória em massa de imóveis.

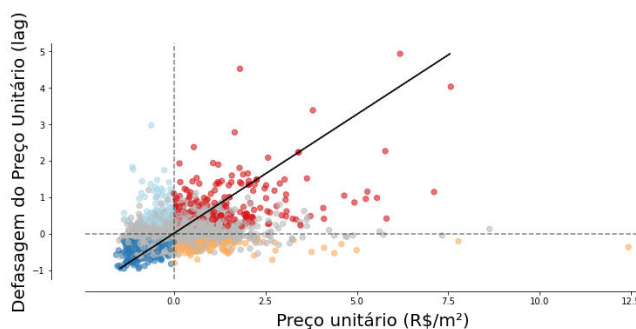
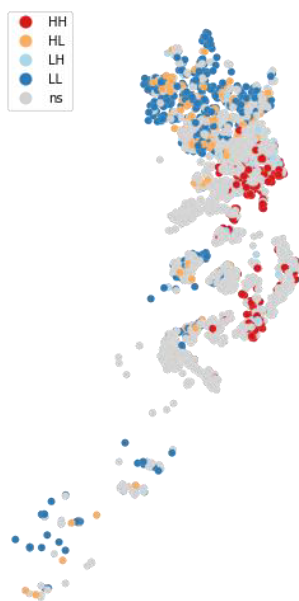


Gráfico 1 - Dispersão Local de Moran I



Mapa 3 - Mapa de Clusters LISA

#### 4.2. MODELO CLÁSSICO DE REGRESSÃO LINEAR MÚLTIPLA COM ESTIMATIVAS PELO MQO

As estimativas dos coeficientes da regressão linear múltipla pelo método dos mínimos quadrados ordinários (MQO), com variável dependente sendo o logaritmo natural do preço observado, se encontram na tabela 4 do Apêndice B. As estimativas tiveram seu sinal de acordo com o esperado e foram significantes a 5%, com exceção daquele relacionado à variável de indicação de informação colhida no ano de 2019 (a2019)<sup>6</sup>, mas, mesmo assim, significativa a 10%<sup>7</sup>. Rejeitou-se a hipótese nula de todos os coeficientes nulos ao mesmo tempo a 1% (Estatística F). Os coeficientes de determinação normal e ajustado foram praticamente iguais, representando um poder de explicação de 88,3% do logaritmo natural dos preços em função das variáveis independentes escolhidas.

<sup>6</sup> Isso indica não existir diferença significativa entre os preços de 2018 e 2019.

<sup>7</sup> A descrição das variáveis se encontra no APÊNDICE A.

Quanto à verificação de multicolinearidade, a matriz de correlação dos regressores tomados dois a dois não apresentou valores elevados. A distribuição dos resíduos se mostrou aparentemente normal (Apêndice B, figura 5).

O modelo clássico estimado pelo MQO se mostrou heterocedástico<sup>8</sup> (variância do erro não é constante), ferindo um dos seus pressupostos básicos. Essa violação implica que suas estimativas, embora lineares e não tendenciosas, não serão as melhores estimativas lineares (melhor estimador linear não tendencioso, BLUE). Desta feita, utilizou-se outro método de estimação, conhecido como mínimos quadrados generalizados (MQG), onde o erro padrão de cada um dos estimadores foi corrigido para a heterocedasticidade e autocorrelação serial (HAC). O modelo MQG produziu todas as estimativas significativas a 5%, e, como é sabido, apresenta os mesmos coeficientes das estimativas MQO<sup>9</sup>.

#### 4.3. COMPROVAÇÃO DA AUTOCORRELAÇÃO ESPACIAL DOS RESÍDUOS - ANÁLISE VARIOGRÁFICA

O primeiro variograma experimental e ajustado (exponencial) da figura 4 apresenta campo estruturado que se estende da origem até o alcance de 338 m, com efeito pepita de 0,02 e patamar bem definido de 0,04, o que nos leva a inferir a presença de autocorrelação espacial nos resíduos do modelo de regressão.

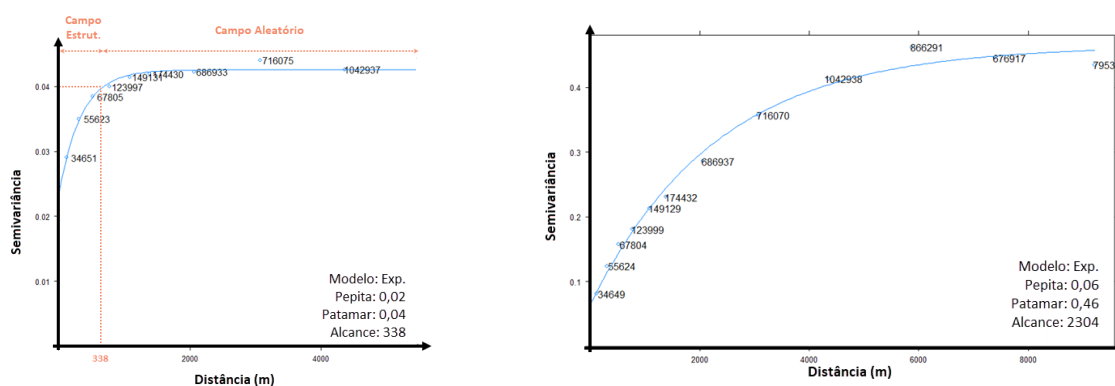


Figura 4 - Variograma experimental e teórico dos resíduos do modelo de regressão e do logaritmo natural do preço.

O segundo variograma ajustado do tipo exponencial da figura 4 foi realizado sobre o logaritmo natural dos preços e apresentou alcance de 2.304 m. Essa última distância será usada para definição da distância máxima de vizinhança para algumas matrizes de ponderação espacial (distância cut-off), conforme se verá.

#### 4.4. ESCOLHA DAS MATRIZES DE PONDERAÇÃO ESPACIAL

Atestada a dependência espacial nos resíduos, parte-se para a definição de uma matriz de ponderação espacial seguindo o critério de Baumont (2004) apud Almeida (2012) que consiste nos seguintes passos: i) estimativa do modelo de regressão linear clássico; ii) testar a autocorrelação espacial pelo índice global de Moran I para um conjunto predefinido de matrizes de ponderação espacial e iii) selecionar a matriz de pesos que promoveu a maior estatística de Moran I, desde que essa seja significativa.

O Tabela 1 apresenta as especificações das matrizes de ponderação espacial a terem suas estatísticas de Moran I calculadas. Para as matrizes de pesos, ponderadas pela distância (matrizes  $W_1$  a  $W_{12}$ ), há um valor mínimo (minimum threshold distance) para que todos os pontos tenham pelo menos um vizinho: aproximadamente 1.262m. Essa distância mínima evita que algum dado fique isolado dos demais, indicando ausência de interação espacial e representando perda de grau de liberdade na matriz de vizinhança ( $\sum w_{ij} = 0$ ). Desta feita, especificou-se a primeira matriz espacial de distâncias tendo como limite máximo (cut-off) para definição de vizinhança este valor. Em outras duas especificações de distâncias, utilizou-se o alcance determinado no variograma teórico da figura 4 para os preços observados e a distância máxima possível entre vizinhos (sem limite). Estabeleceu-se ainda critérios distintos de ponderação dessa relação: i) pelo inverso da distância; ii) pelo inverso da distância ao quadrado; iii) pela ponderação bisquare e iv) pela ponderação gaussiana. Quanto às matrizes de contiguidade, especificou-se a relação de vizinhança pela

<sup>8</sup> Foram realizados os testes formais de Breusch-Pagan, Koenker e Goldfeld-Quandt, o quais têm como hipótese nula a homocedasticidade.

<sup>9</sup> Por conta disso, preferimos utilizar a expressão "MQO" a "MQG" neste trabalho para melhor entendimento.

tesselação de Voronoi (polígonos de Thiessen) segundo o critério rainha (queen) com as seguintes ordens de vizinhança: 1, 2, 3, 5 e 10 (matrizes  $W_{13}$  a  $W_{17}$ ). Convém ressaltar, que todas as matrizes foram padronizadas por linha ( $\sum w_j = 1$ ).

Tabela 1 - Especificação das matrizes de ponderação espacial.

Espec.	Tipo	Distância cut-off (dMAX)		Espec.	Tipo	Distância cut-off (dMAX)	
$W_1$	Distância inversa (exp=1)	1.262,00		$W_{10}$	Distância (gaussiana)	2.300,00	
$W_2$	Distância inversa (exp=2)	1.262,00		$W_{11}$	Distância (bisquare)	27.272,00	
$W_3$	Distância inversa (exp=1)	2.300,00		$W_{12}$	Distância (gaussiana)	27.272,00	
$W_4$	Distância inversa (exp=2)	2.300,00		$W_{13}$	Contig. Queen 1ª ord.	-	
$W_5$	Distância inversa (exp=1)	27.272,00		$W_{14}$	Contig. Queen 2ª ord.	-	
$W_6$	Distância inversa (exp=2)	27.272,00		$W_{15}$	Contig. Queen 3ª ord.	-	
$W_7$	Distância (bisquare)	1.262,00		$W_{16}$	Contig. Queen 5ª ord.	-	
$W_8$	Distância (gaussiana)	1.262,00		$W_{17}$	Contig. Queen 10ª ord.	-	
$W_9$	Distância (bisquare)	2.300,00					

Notas: i)  $d_{ij}$  é a distância euclidiana em metros entre as observações i e j; ii)  $d_{MAX}$  é a distância limite de definição de vizinhança, ou seja, a partir daquela, não há mais relação de vizinhança.

#### 4.5. RESULTADOS DOS MODELOS ECONOMÉTRICOS ESPACIAIS

A tabela 2 apresenta os testes de Moran I e de multiplicadores de Lagrange para a dependência espacial dos resíduos de modelo de regressão, de acordo com cada especificação de matriz de ponderação espacial, conforme a tabela 1. De imediato, já se observa o pouco poder de captura de dependência espacial, refletido no baixo valor da estatística de Moran I significativa para as matrizes de contiguidade, independentemente de sua ordem (matrizes  $W_{13}$  a  $W_{17}$ ). A matriz de contiguidade de décima sétima ordem ( $W_{17}$ ) não apresentou resultado significativo de autocorrelação.

As matrizes ponderadas pelo inverso da distância ao quadrado (matrizes  $W_2$ ,  $W_4$  e  $W_6$ ) foram as que mais conseguiram capturar a dependência espacial, apresentando maiores valores da estatística de Moran I. Observa-se um pequeno decréscimo na estatística dessas tabelas com o aumento da distância de cut-off. As matrizes  $W_{11}$  e  $W_{12}$  (bisquare e gaussiana sem limite de distância) não apresentaram indicação de autocorrelação (estatística de Moran I significativa e igual a zero).

Tabela 2 - Resultados dos testes de dependência espacial de acordo com a especificação das matrizes de ponderação espacial W.

Espec.	Moran I		LM $\rho$ (SAR)		LM $\lambda$ (SEM)		Robusto LM $\lambda$ (SEM)		Robusto LM $\lambda$ (SEM)		SAC	
	Valor	Prob.	Valor	Prob.	Valor	Prob.	Valor	Prob.	Valor	Prob.	Valor	Prob.
$W_1$	0,144	0	66,462	0	303,077	0	0,674	0,412	237,29	0	303,751	0
$W_2$	0,19	0	78,877	0	205,001	0	0,001	0,981	126,125	0	205,002	0
$W_3$	0,129	0	65,605	0	296,088	0	0,045	0,833	230,528	0	296,133	0
$W_4$	0,188	0	79	0	208,131	0	0,002	0,969	129,133	0	208,133	0
$W_5$	0,085	0	35,788	0	194,368	0	1,132	0,287	159,713	0	195,501	0
$W_6$	0,183	0	71,032	0	201,086	0	0,141	0,707	130,196	0	201,228	0
$W_7$	0,061	0	20,739	0	537,405	0	0,042	0,838	516,708	0	537,447	0
$W_8$	0,052	0	10,285	0,001	617,924	0	0,405	0,525	608,043	0	618,328	0
$W_9$	0,046	0	15,15	0	875,444	0	0,106	0,745	860,4	0	875,55	0
$W_{10}$	0,037	0	19,12	0	937,351	0	2,532	0,112	920,763	0	939,883	0
$W_{11}$	0	0	16,669	0	0,126	0,723	30,723	0	14,18	0	30,849	0
$W_{12}$	0	0	0,308	0,579	0,224	0,636	23,856	0	23,772	0	24,08	0
$W_{13}$	0,094	0	24,538	0	93,457	0	0,001	0,973	68,921	0	93,458	0
$W_{14}$	0,078	0	13,382	0	149,905	0	2,012	0,156	138,535	0	151,917	0
$W_{15}$	0,057	0	1,991	0,158	141,377	0	6,721	0,01	146,108	0	148,099	0
$W_{16}$	0,033	0	4,156	0,041	99,338	0	0	0,985	95,183	0	99,339	0
$W_{17}$	-0,002	0,853	4,231	0,04	1,375	0,241	5,059	0,024	2,202	0,138	6,434	0,04

Nota: i) Moran I calculado sobre os resíduos da regressão; ii) em destaque, as matrizes com maior estatística de Moran I e indicação de que o modelo econométrico mais apropriado é o de erro espacial (ANSELIN; REY, 2014)

Seguindo o procedimento do fluxograma da figura 3 de Anselin e Rey (2014), por meio da tabela 2 observa-se que, com exceção das matrizes  $W_{11}$ ,  $W_{12}$ ,  $W_{15}$  e  $W_{17}$ , os multiplicares de Lagrange LM $\rho$  (SAR) e LM $\lambda$  (SEM) foram ambos significantes, o que nos levaria à realização destes testes na sua forma robusta. Fazendo isto, para essas matrizes, observa-se que os modelos econométricos mais apropriados são os de erro espacial (SEM). Destarte, apresenta-se na tabela 3, os resultados dos modelos econométricos testados

utilizando as matrizes de ponderação espacial que tiveram maior valor da estatística de Moran I, modelos estes tanto na forma de erro espacial (SEM), como misto (SAC). Todos os modelos apresentados tiveram melhor desempenho frente ao modelo MQO, segundo o critério de Akaike (o AIC deste foi de -1.046,39).

Percebe-se que as matrizes que tiveram maior valor para a estatística de Moran I ( $W_2$ ,  $W_4$  e  $W_6$ ) tiveram maiores AIC e erros quadráticos médio dos preços observados na escala original do que as matrizes  $W_1$  e  $W_3$ .

Tabela 3 - Modelos Econométricos ordenados pelo critério de Akaike (menor para o maior) com significância entre parênteses.

Ordem	Matriz	Especificação	Rho ( $\rho$ )	Lambda ( $\lambda$ )	AIC	Log likelihood	RMSE	RMSE (não linear)
1	$W_1$	Erro (SEM)		0,475 (0,000)	-1.306,89	678,447	0,197	120.932,01
2	$W_1$	Misto (SAC)	0,031 (0,000)	0,458 (0,000)	-1.306,74	679,369	0,197	121.092,37
3	$W_3$	Misto (SAC)	0,047 (0,000)	0,494 (0,000)	-1.303,88	677,942	0,198	121.351,80
4	$W_3$	Erro (SEM)		0,519 (0,000)	-1.302,06	676,031	0,197	120.976,83
5	$W_4$	Erro (SEM)		0,292 (0,000)	-1.248,22	649,112	0,199	122.497,22
6	$W_4$	Misto (SAC)	0,021 (0,000)	0,276 (0,000)	-1.247,80	649,9	0,199	122.746,50
7	$W_2$	Erro (SEM)		0,285 (0,000)	-1.244,36	647,182	0,199	122.700,12
8	$W_2$	Misto (SAC)	0,019 (0,000)	0,27 (0,000)	-1.243,68	647,84	0,2	122.933,65
9	$W_6$	Erro (SEM)		0,294 (0,000)	-1.242,66	646,332	0,199	122.553,38
10	$W_6$	Misto (SAC)	0,019 (0,000)	0,28 (0,000)	-1.241,85	646,925	0,2	122.775,53

## 5. CONCLUSÕES

A econometria espacial baseia-se na introdução explícita dos relacionamentos espaciais nos modelos de regressão linear múltipla. Para isso, é necessário a definição de uma matriz de ponderação espacial. A escolha errada dessa matriz pode levar a resultados não satisfatórios de incorporação da dependência espacial, bem como na qualidade de ajustes dos modelos. Este trabalho abordou questões relacionadas à escolha e à utilização de diferentes matrizes de ponderação espacial na avaliação em massa de casas isoladas no Município de Aracaju.

Para modelos espaciais aplicados à avaliação em massa, necessário se faz uma ESDA por meio de mapas de correlograma, gráficos de dispersão local de Moran I (LISA) e seus respectivos mapas de clusters, bem como o uso da geoestatística (análise variográfica) antes de qualquer procedimento de estimação de modelos. Isso dará ao avaliador/pesquisador insights suficientes para a definição da melhor matriz de ponderação espacial. Como visto, a análise variográfica pode indicar distâncias (cut-off) adequadas para montagem da matriz de pesos espaciais. A ESDA pode ser realizada sobre a própria variável dependente, variáveis independentes, e não só sobre os resíduos do modelo de regressão múltipla com estimativa pelo MQO.

As matrizes de ponderação espacial baseadas na distância promovem melhor ajustes nos modelos econométricos espaciais, seja pela melhor incorporação da dependência espacial, seja pelo melhor resultados nos critérios de AIC e raiz do erro quadrático médio. Ponderações mais elaboradas do efeito da distância como as matrizes de distância tipo bisquare e gaussiana não foram capazes de se sobressair frente às matrizes tradicionais baseadas no inverso da distância. As matrizes de contiguidade, mesmo padronizadas por linha e com diferentes ordens de vizinhança, não apresentam resultados satisfatórios para captura e incorporação da dependência espacial, não sendo recomendado a sua utilização nos modelos econométricos de avaliação em massa. Conclui-se que estas matrizes são mais indicadas em análise regionais de dados situados em polígonos e não para as avaliações de imóveis, onde estes são representados pelo centroide de seu terreno (ou edificação). Isto pois, dados pontuais requerem o

estabelecimento de relações de vizinhança do tipo tesselação de Voronoi, o que se afasta, neste caso, da primeira Lei da Geografia de Waldo Tobler.

Cabe ao avaliador/pesquisador sobepesar o trade-off de escolha entre matrizes que capturem ao máximo a dependência espacial dos preços observados com aquela que tenha melhor performance nas predições. Como visto, as matrizes que apresentaram a maior estatística de Moran I não foram aquelas com melhor performance, apesar de resultados finais serem bastante similares entre todas as matrizes ponderadas pelo inverso da distância escolhidas (seja com expoente 1 ou 2 de ajuste de peso), o que nos indica baixa sensibilidade (e boa robustez) dos modelos frente a essas escolhas.

Por fim, pela experiência prévia dos autores, e, confirmada nesse trabalho, os modelos de erro espacial são os mais adequados às avaliações em massa de imóveis. Também não se verificou diferenças significativas entre os resultados dos modelos SEM e SAR, o que nos leva a opinar, pelo princípio da parcimônia, ao modelo SEM.

## REFERÊNCIAS

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

ANDRIOTTI, José Leonardo Silva. **Fundamentos de estatística e geoestatística**. São Leopoldo: Unisinos, 2003.

ANSELIN, L., **Local indicators of spatial association – LISA**. In: Geographical Analysis, Vol. 27, No. 2 (April 1995). Disponível em: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1538-4632.1995.tb00338.x>. Acesso em: 10 out. 2020.

ANSELIN, Luc; REY, Sergio J. **Modern Spatial Econometrics in Practice: a guide to GeoDa, GeoDaSpace and PySAL**. Geoda Press LLC. Chicago, IL, 2014. 368 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2: imóveis urbanos**. Rio de Janeiro, 2011.

BRASIL. Ministério das Cidades. Gabinete do Ministro. **Portaria nº 511, de 7 de dezembro de 2009**. Diretrizes para a criação, instituição e atualização do Cadastro Territorial Multifinalitário (CTM) nos municípios brasileiros. Diário Oficial da União, Brasília, DF, 08 dez. 2009.

DANTAS, Rubens A. **Modelos Espaciais aplicados ao mercado habitacional – um estudo de caso para a cidade de Recife**. Recife, 2003 (Tese – Universidade Federal de Pernambuco – UFPE).

GOLGHER, André B. **Introdução à econometria espacial**. Jundiaí: Paco Editorial, 2015.

IBGE. **IBGE - cidades @**. 2020. Disponível em: <<http://www.ibge.gov.br/cidadesat/default.php>>. Acesso em: 13 out. 2020.

LEENDERS, Roger Th.A.J.. Modeling social influence through network autocorrelation: constructing the weight matrix. **Social Networks**, [S.L.], v. 24, n. 1, p. 21-47, jan. 2002. Elsevier BV. [http://dx.doi.org/10.1016/s0378-8733\(01\)00049-1](http://dx.doi.org/10.1016/s0378-8733(01)00049-1).

LONGLEY, Paul A. et al. **Sistemas e Ciência da Informação Geográfica**. 3. ed. Porto Alegre: Bookman, 2013. 540 p.

LUIZ, C. et al. **Observatório das Metrôpoles: Níveis de integração dos municípios brasileiros em RMS, RIDES e AUS à dinâmica da metropolização**. [s.l.: s.n.].

NASCIMENTO, M. M.; ARAÚJO, H. M. DE. **A urbanização extensiva de Aracaju e a formação de novos aglomerados habitacionais: avaliação a partir da desagregação de dados dos CENSOS-IBGE / The extensive urbanization of Aracaju and the formation of new housing agglomerates: evaluation (...)**. Caderno de Geografia, v. 28, n. 52, p. 166, 2018.

OLIVEIRA, L. S.; ANDRADE, A. C. S. **Dinâmica da paisagem da Zona de Expansão de Aracaju-Se**. Scientia Plena, v. 8, p. 1–6, 2012.

PNUD – PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO; IPEA – INSTITUTO DE PESQUISA ECONÔMICA E APLICADA; FUNDAÇÃO JOÃO PINHEIRO. **Atlas de desenvolvimento humano do Brasil de 2013**. 2013. Disponível em: <<http://www.atlasbrasil.org.br/2013/>>. Acesso em: 13 out. 2020.

TYSZLER, Marcelo. **Econometria Espacial: discutindo medidas para a matriz de ponderação espacial**. 2006. 155 f. Dissertação (Mestrado) - Curso de Administração Pública e Governo, Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, São Paulo, 2006.

YAMAMOTO, Jorge Kazuo. **Estatística, análise e interpolação de dados geoespaciais**. São Paulo: Grafica Paulos, 2020.

## APÊNDICE A

Quadro 1 – Variável dependente e variáveis explicativas testadas nos modelos.

Variável	Descrição
<b>Variável Dependente</b>	
preco	Valor total do imóvel a ser considerado na regressão, em reais
<b>Variáveis Estruturais</b>	
infra	Variável que representa a soma dos elementos de infraestrutura presentes na face do logradouro (água, esgoto, galeria pluvial, guias e/ou sarjetas, iluminação pública e pavimentação). 1 – soma menor ou igual a 3; 2 – soma igual a 4; 3 – soma igual a 5; 4 – soma igual a 6
sit_quad	Indica a situação modelada para regressão (0 – Meio de quadra, Encravado ou Interno; 1 – Mais de uma frente)
topog	Indica a topografia modelada para regressão (0 – Aclive, Declive ou Acidentado; 1 – Plano)
testada	Indica a testada principal do lote em metros
cota	Cota de terreno da unidade em metros quadrados (multiplicação da fração ideal pela área do terreno).
area_unid	Indica a área total construída da unidade em metros quadrados
num_pav	Indica o número de pavimentos da construção
pad_bom	Padrão da construção (0,50; 0,70; 1,00; 1,20; 1,35)
pad_ot_esp	Padrão da construção (0,50; 0,70; 1,00; 1,20; 1,35)
cons_conserv	Indica o estado de conservação do imóvel, se é conservado ou não
idade_atual	Idade da construção em anos
pisc	Indica se imóvel possui piscina (0 – Sem; 1 – Com)
<b>Variáveis de Localização</b>	
renda	Variável proxy representada pela renda média mensal do responsável do imóvel, em salários mínimos, ajustada a uma superfície de tendência, construída pelo processo de krigagem
zona_urb	Indica se o imóvel está localizado na zona de expansão urbana ( 0 ) ou na zona urbana ( 1 )
av_rod	Indica se imóvel localiza-se em avenida ou rodovia (0 – Não; 1 – Sim)
perc_com	Variável proxy que representa o percentual de unidades comerciais na face de quadra
<b>Variáveis de Distâncias</b>	
dist_agrn	Variável que representa a distância ao aglomerado subnormal mais próximo
dist_shopp	Variável que representa a distância ao shopping mais próximo
<b>Variáveis de Informação</b>	
a2019	Variável dummy que indica se dado do ano de 2019 (0 – Não; 1 – Sim)
a2020	Variável dummy que indica se dado do ano de 2020 (0 – Não; 1 – Sim)
or_trans	Variável dummy que indica tratar-se de transação (0 – Não; 1 – Sim)
or_oferta	Variável dummy que indica tratar-se de oferta (0 – Não; 1 – Sim)

## APÊNDICE B

Tabela 4 – Estimativas e erros padrão dos Modelo MQO e MQG.

Variável/Intercepto	Estimativa	MQO		MQG (HAC)	
		Erro padrão	significância	Erro padrão	significância
(Intercepto)	7,8558	0,0525	0,001%	0,0624	0,000%
infra	0,0207	0,0048	0,001%	0,0049	0,003%
sit_quad	0,0429	0,0114	0,016%	0,0120	0,034%
topog	0,2245	0,0317	0,000%	0,0289	0,000%
ln_testada	0,1086	0,0122	0,000%	0,0139	0,000%
ln_cota	0,3417	0,0106	0,000%	0,0137	0,000%
ln_area_unid	0,2962	0,0111	0,000%	0,0134	0,000%
num_pav	0,0752	0,0097	0,000%	0,0111	0,000%
pad_bom	0,0778	0,0092	0,000%	0,0095	0,000%
pad_ot_esp	0,1587	0,0237	0,000%	0,0247	0,000%
cons_conserv	0,0665	0,0082	0,000%	0,0083	0,000%
idade_5	0,0431	0,0198	2,946%	0,0209	3,894%
pisc	0,0400	0,0178	2,495%	0,0186	3,152%
ln_renda	0,2932	0,0089	0,000%	0,0098	0,000%
zona_urb	0,2597	0,0147	0,000%	0,0149	0,000%
av_rod	0,0518	0,0122	0,002%	0,0143	0,030%
dens_vert	0,0118	0,0023	0,000%	0,0027	0,001%
perc_com	0,0016	0,0003	0,000%	0,0003	0,000%
dist_shopp	-0,0291	0,0030	0,000%	0,0034	0,000%
a2019	-0,0178	0,0092	5,251%	0,0082	3,110%
a2020	-0,0265	0,0105	1,179%	0,0099	0,771%
or_trans	0,1223	0,0097	0,000%	0,0090	0,000%
or_oferta	0,3099	0,0091	0,000%	0,0095	0,000%
Estatística F	1228,364	Log likelihood		546,196	
R <sup>2</sup>	0,884	AIC		-1046,392	
R <sup>2</sup> ajustado	0,883	BIC		-904,180	

Notas: i) O prefixo "ln\_" antes da variável indica que foi aplicada uma transformação pelo logaritmo natural.

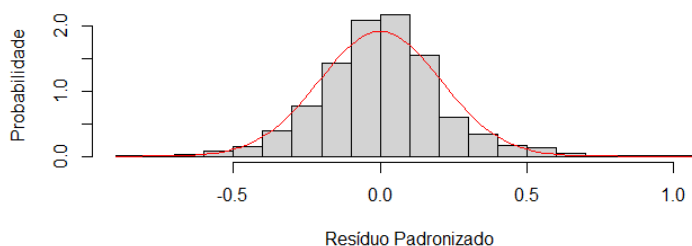


Figura 5 - Histograma dos resíduos padronizados.