MODELOS ECONOMÉTRICOS ESPACIAIS PARA AVALIAÇÃO EM MASSA DE TERRENOS

Spatial Econometrics Applied to Real Estate Mass Appraisal

Sandro Ricardo Vasconcelos Bandeira
Secretaria Municipal das Finanças de Fortaleza
sandro.bandeira@sefin.fortaleza.ce.gov.br

Antônio Augusto Ferreira de Oliveira
Secretaria Municipal das Finanças de Fortaleza
augusto.oliveira@sefin.fortaleza.ce.gov.br

Luan Victor Vasconcelos Noberto
Secretaria Municipal das Finanças de Fortaleza
luan.noberto@sefin.fortaleza.ce.gov.br

Resumo:
O valor venal de mercado é entendido como a quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, em uma data de referência, dentro das condições do mercado vigente. Com o objetivo de estimar o valor de mercado de um imóvel, as normas brasileiras elegem vários métodos igualmente válidos, dentre os quais, destaca-se o método comparativo direto de dados de mercado, com o uso de inferência estatística, por meio do método dos Mínimos Quadrados Ordinários (MQO). Percebe-se que a localização do imóvel, por si só, exerce um papel de suma importância na definição de seu valor, assumindo, portanto, uma notável contribuição na formação dos preços dos imobiliários. Neste trabalho, foi proposto dois macromodelos, o primeiro contemplou somente a análise clássica, por meio do MQO, o segundo tomando em conta os comportamentos espaciais da variável dependente, preço, e dos resíduos do modelo MQO, com uso das abordagens *Spatial Autoregressive Models* (SAR) e *Spatial Error Models* (SEM), respectivamente. Ao final, mediu-se a performance de cada uma das diferentes abordagens, comparando-se a precisão dos modelos, por meio das estatísticas de Akaike e Schwarz e de métricas definidas pela *International Association of Assessing Officers* (IAAO), para definir aquele que apresenta os resultados mais consistentes. Foi ainda realizado estudo de caso para estimar valores venais de terrenos no município de Fortaleza a fim de exemplificar os estudos teóricos apresentados e solidificar o entendimento. Concluiu-se que o uso do modelo espacial, com base na contribuição espacial dos erros (SEM), produziu os melhores resultados dentre todos os testados.

Palavras-chave: Avaliação em massa; modelos clássicos MQO; modelos econométricos espaciais.

Abstract:
In order to estimate the market value of a property, Brazilian standards elect several equally valid methods, among which the direct comparative market method stands out with statistical inference by adopting the Method of Ordinary Least Squares (OLS). When we insert the first law of geography by Waldo Tobler (1970) in statistical study it can be noticed that the location of a property by itself plays an extremely important role in the definition of this market value, assuming, therefore, a notable contribution to form property prices. In this work, two macro-models were proposed, the first contemplated only the classical analysis, by means of the MQO, then an analysis was elaborated using spatial econometric tools, also appreciating the spatial behavior of the price dependent variable and of the residues of the MQO model, using Spatial Autoregressive Models (SAR) and Spatial Error Models (SEM), respectively. Therefore, the performance of each one of the approaches will be measured through a sample test - which was not included in the initial modeling - while comparing the model precision by using Akaike and Schwarz statistics to define the one that presents the most accurate and consistent results. In this work,

it's been developed a study case to estimate venal values of land in Fortaleza (main city of Ceará State, Brazil) in order to exemplify the theoretical studies presented as well as to solidify a better understanding about this matter. In conclusion, we may say that the use of the spatial model, based on the Spatial Contribution of the model residuals produced the best results among all those tested.

Keywords: Real estate mass appraisal; ordinary least squares; spatial econometrics.

1. INTRODUÇÃO

Com fundamento na norma brasileira de avaliação de imóveis para fins urbanos, ABNT NBR 14.653, o valor de mercado é entendido como a estimativa de valores que o imóvel alcançaria em uma transação à vista, cuja negociação seja voluntária e consciente, em uma data de referência, dentro das condições do mercado vigente. Assim, o valor de mercado é o preço justo pago por um imóvel por um comprador desejoso de comprar para um vendedor desejoso de vender, ambos com plena liberdade de exercer suas escolhas.

Mister ressaltar que aferir, com exatidão, o valor do mercado de um bem, sem participar da transação comercial propriamente dita, é tarefa hercúlea, senão impossível. Dessa forma, deve-se fazer uso de técnicas estatísticas, cujo objetivo é estimar com a maior acurácia possível, ou ainda, reduzir ao máximo a diferença entre o valor efetivo observado e esta estimativa.

O interesse em se prever o valor de imóvel é bastante diverso, abrangendo várias áreas econômicas. Além daqueles operadores do mercado imobiliário, naturalmente os maiores interessados em avaliação de imóveis, podemos citar também as sociedades empresárias em geral, para estimar valores de seu ativo imobilizado, bancos e entes subnacionais, principalmente as prefeituras, responsáveis por impostos como o IPTU e ITBI.

São diversas as metodologias utilizadas para avaliar, ficando a critério do profissional de engenharia de avaliações a escolha de qual método irá suprir de forma mais eficiente as necessidades exigidas pela situação ora apresentada. Este trabalho, contará com a utilização de dois métodos de avaliação, sendo eles: método comparativo direto de dados de mercado por tratamento dos dados com regressão linear múltipla (RLM), por meio do método dos mínimos quadrados ordinários (MQO) e regressão espacial (RE), através da econometria espacial, utilizando-se das abordagens do erro espacial correlacionado (SEM) e do preço defasado espacialmente (SAR), conforme metodologia proposta por Anselin e Rey (2014).

O uso da abordagem espacial se justifica na ideia extremamente racional que atribui a correlação entre áreas nobres e imóveis mais caros, assim como o seu inverso, postulado pela primeira Lei da Geografia de Waldo Tobler (1970), ensinando que *"as coisas são relacionadas entre si, mas coisas próximas estão mais relacionadas do que coisas mais distantes"*.

Portanto, o trabalho visou apresentar uma sistemática de cálculo enquadrada nos trâmites da norma brasileira, ABNT NBR 14.653-2:2011, mormente com o estabelecido em seu Anexo C (fls. 42-43), com a introdução do viés espacial, a fim de que se possa apurar a influência de elementos vizinhos entre si. A título de estudo de caso, utilizou-se de uma amostra de dados de mercado (ofertas) na tipologia territorial. Entretanto, não há nenhum óbice para a expansão do estudo, com sua aplicação em imóveis prediais em geral.

2. MATERIAS E MÉTODOS

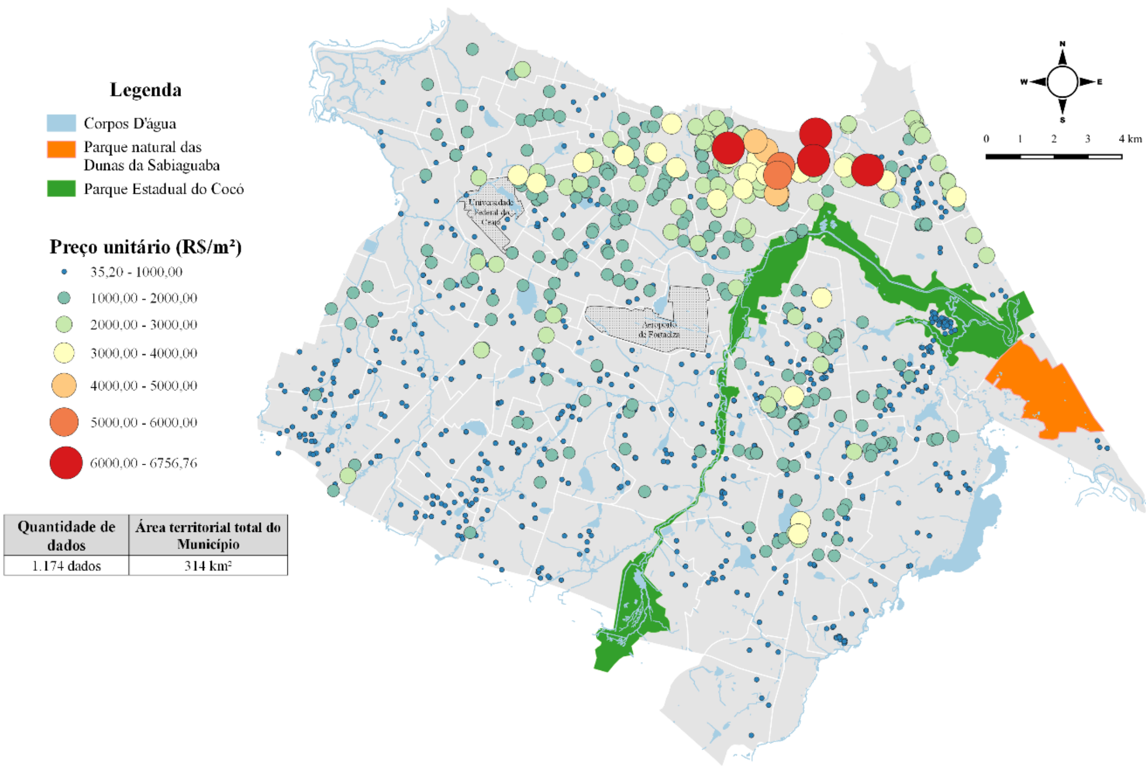
2.1. Área de estudo e descrição dos dados

O Município de Fortaleza é a capital do estado do Ceará, com população estimada em 2,6 milhões de pessoas, segundo projeção do IBGE para o ano 2018, sendo a quinta mais populosa cidade do Brasil. Possui área territorial de 314,93 km², o que torna a mais adensada entre todas as capitais com 7.786,44 hab./km², segundo o censo de 2010.

De acordo com o cadastro imobiliário municipal mantido na Secretaria Municipal das Finanças de Fortaleza (SEFIN), o número de inscrições municipais ou unidade autônomas imobiliárias sujeitas à tributação do Importo Predial e Territorial Urbano (IPTU) está na casa das 779 mil inscrições. Já o número total de lotes cadastrados é de aproximadamente 385 mil, dentre os quais, cerca de 74 mil não estão edificados, isto é, sendo terrenos vazios, representando menos que 10% do total das inscrições imobiliárias. Salienta-se que a tributação do IPTU em Fortaleza se dá pelo método evolutivo, onde a composição do valor venal do imóvel, que é a base de cálculo desse imposto, se dá pela soma do valor da parte territorial com a parte edificada depreciada.

Os 1.174 dados, todos representados na modalidade de "oferta", usado para composição da amostra trabalhada para esse artigo, foram disponibilizados pela SEFIN no intervalo de tempo compreendido entre os meses de julho de 2019 a julho de 2020. A distribuição espacial dessas amostras pode ser visualizada no Mapa 1:

Mapa 1 - Distribuição espacial da amostra.



Fonte: Elaborados pelos autores (2020).

A variável explicada foi o preço unitário do terreno em R$/m², representando a divisão do preço observado (R$) dividido pela sua área (m²). As variáveis explicativas e a explicada utilizadas para as modelagens encontram-se na Tabela 1:

Tabela 1 – Descrição das variáveis testadas/utilizadas no modelo.

Variáveis	Descrição	Média	Mínimo	Máximo
distvp	Distância em m à via principal mais próxima.	228,43	4,46	1.441,56
distsh	Distância ao shopping mais próximo.	1.725,16	9,79	7.663,84

Variáveis	Descrição	Média	Mínimo	Máximo
dscom	Densidade de comercialização de lotes que estão no mesmo trecho de logradouro do lote corrente.	0,09	0,01	1
dassprec	Distância ao assentamento precário mais próximo.	285,19	0	1.383,37
vbt	Valor de referência do terreno do IPTU.	39,90	3,02	1.532,69
renda	Renda do responsável *krigeada* IBGE 2010.	2,86	0,40	18,89
idhed	IDH, dimensão educação por bairro.	0,81	0,65	0,95
iaeq	Índice de aproveitamento equivalente.	1,69	0,01	3
lotcnd	0-fora de condomínio; 1-em condomínio.	0,01	0	1
numfr	número de frentes.	1,25	1	4
test	comprimento da testada principal em m.	18,69	3	90,97
area	área do terreno em m².	508,21	27	2.000,00
agua	0-sem água; 1-com água.	0,86	0	1
esg	0-sem esgoto; 1-com esgoto.	0,44	0	1
pav	0-sem pavimentação asfáltica (ou concreto) no trecho de logradouro; 1-com pavimentação asfáltica (ou concreto) no trecho de logradouro.	0,70	0	1
dv	Densidade de verticalização (kernel) (condomínios com elevador).	1,09	0	28,12
vunit	Valor unitário do terreno em R$/m².	735,48	35,20	6.756,76

Fonte: Elaborado pelos autores (2020).

2.2. Método de regressão linear multivariado, com base nos mínimos quadrados ordinários (MQO)

A equação de regressão linear múltipla, com n variáveis é dada pela Equação 1:

$$Y=\alpha+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n+\varepsilon \quad (1),$$

onde:

Y – Valor Venal (variável dependente).
α – Intercepto (coeficiente linear do modelo).
$\beta_{(1,2,3,\ldots,n)}$ – Coeficientes/parâmetros dos regressores (variáveis) usadas do modelo.
$X_{(1,2,3,\ldots,n)}$ – Regressores (variáveis independentes) usadas do modelo (vagas, área, dormitórios, natureza do evento etc).
ε – Variável aleatória que representa o erro (resíduo) experimental.

Os modelos de regressão linear possuem os seguintes pressupostos que devem ser obedecidos: 1) relação Linear entre variáveis independentes (X) e dependente (Y): o ajuste só é válido para relações lineares; 2) os valores das variáveis independentes (X) são fixos em repetidas amostras, não aleatórios – faz-se a pressuposição que dado um valor de X, Y irá variar segundo uma distribuição de probabilidade com valor esperado dado por $E(Y|X_i)$; 3) as esperança condicional dos erros é igual a zero; 4) a variabilidade dos erros é constante, ou seja, os erros são homocedásticos; 5) os erros são não autocorrelacionados, ou seja, não há correlação entre valores ordenados dos erros, no tempo e no espaço; 6) os erros apresentam distribuição normal[1].

2.3 Dependência espacial e matrizes de vizinhança

Na análise do mercado imobiliário, a variável localização exerce um papel de suma importância nos modelos. Ora, é extremamente factível a ideia de que os preços dos imóveis vizinhos *i* e *j* possam se correlacionar espacialmente entre si, bem como os resíduos observados entre os valores observados (reais) e estimados pelo modelo.

Portanto, essa correlação espacial entre os dados observados é regra e não exceção, e isso traz como consequência a inobservância da hipótese 5, dos pressupostos de regressão linear acima. Na forma matricial, chega-se à seguinte conclusão:

$$Y_i = f(Y_j, X), \quad i, j = 1,..., n \text{ e } i \neq j \qquad (2),$$

Os preços dos vizinhos *j* interferem no preço do imóvel *i*. Uma das maiores dificuldades encontradas na utilização de métodos comparativos de dados de mercado por regressão linear múltipla, na busca de modelos de avaliação, está em considerar a localização, vez que sua inobservância poderia causar sérios problemas de predição, pois dados localizados espacialmente (que é o caso dos imóveis), em geral, apresentam autocorrelação ou covariância espacial.

Para se inserir a contribuição da variável localização (efeitos espaciais) nas modelagens, o primeiro passo é aferir se há mesmo a presença desses efeitos espaciais. A dependência espacial tem sido diagnosticada na literatura de duas formas distintas: pela análise gráfica do variograma ou utilizando testes estatísticos específicos, que utilizam matrizes de vizinhança, como o teste global de Moran (I), caminho que vamos seguir nesse trabalho.

Desta feita, para se efetuar o cálculo do teste de Moran (I), de acordo com Anselin e Rey (2014), é necessário definir previamente uma matriz de vizinhança (por contiguidade ou distâncias), conhecida como W. No caso das matrizes de vizinhança por contiguidade, cada elemento é igual a um se *i* e *j* forem vizinhos e zero, em caso contrário. Por convenção, os elementos da diagonal principal da matriz W são iguais a zero, ou seja, $W_{ij(i=j)}=0$, vez que um elemento não pode ser vizinho de si mesmo. No final, produz-se uma matriz esparsa binaria (apenas com 0 e 1), cuja maior vantagem seria minimizar o esforço computacional nos cálculos (implicando uma maior rapidez nas estimações).

Uma outra forma, até mais eficiente, de se construir uma matriz de vizinhança, seria considerar a importância dos vizinhos por meio de uma ponderação correspondente ao inverso da distância (d_{ij}^{-1}) ou ao inverso do quadrado da distância entre eles (d_{ij}^{-2}) (chamada de matriz de vizinhança por distâncias). Em geral, a matriz de W_{ij}, é padronizado por linha, assumindo a nomenclatura W_s.

Neste caso, cada elemento de W_s, representado por W_{sij}, é obtido dividindo-se W_{ij} pela soma dos elementos da linha *i* a que pertence, configurando-se na fórmula abaixo:

$$W_{sij} = \frac{w_{ij}}{\sum_{ij=1}^{n} w_{ij}}. \qquad (3),$$

Nesta matriz, os elementos das linhas somam 1. Ao contrário das matrizes de

[1] Não é um pressuposto necessário para que os estimadores de MQO sejam MELNV (melhor estimador linear não viesado, conforme o Teorema Gauss–Markov, mas necessário para que as inferências sejam válidas.

contiguidade, aqui se produz uma matriz numérica, bem mais precisa em representar a contribuição espacial da localização na formação dos preços. Entretanto produz um esforço computacional muito maior, na aferição de seus resultados.

No presente trabalho, optamos por usar uma matriz de vizinhança montada pelo inverso da distância ponderada por linha e com raio de determinação dos vizinhos de 1.394m, que corresponde ao valor mínimo para que todos os dados tenham, pelo menos, um vizinho.

Por último, faz-se mister observar que a norma brasileira, ABNT NBR 14.653-2:2011, em seu anexo C, faculta o uso da abordagem espacial, não o tornando obrigatório ainda.

2.4 Teste de Moran (I)

Como já se comentou, o teste de Moran (I) tem como finalidade atestar se há dependência espacial dos resíduos da regressão linear. Este índice é uma medida global da autocorrelação espacial, indicando o grau de associação espacial do conjunto de dados.

O seu valor estatístico é calculado pela seguinte expressão:

$$I = \frac{n}{W_s} . \frac{\sum_i^n \sum_j^n w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i^n (y_i - \bar{y})^2} \qquad (4),$$

Onde:

I – índice de Moran’s I, que é uma métrica usada para testar a hipótese de autocorrelação espacial;

w_{ij} – elemento da matriz de distância para o par de vizinhos i e j;

W_s – matriz que representa a ponderação para cada uma das linhas da matriz ω_{ij};

y_i e $\bar{y}$ – representam o valor da variável na localização i e sua média;

n – número de dados.

De maneira geral, o índice de Moran’s I é um teste de hipóteses, onde a hipótese nula (H_0) é de independência espacial, obtendo-se o valor estatisticamente zero neste caso. Valores positivos (entre 0 e +1) indicam para correlação direta, indicando que a localização possui vizinhos com valores semelhantes (alto – alto e baixo – baixo), e negativos (entre 0 e – 1) correlação inversa, indicando que a localização possui vizinhos com valores deferentes (alto – baixo e baixo – alto).

2.5 Modelo de defasagem espacial (SAR)

A teoria acerca dos modelos espaciais foi desenvolvida recentemente (por volta dos anos 1960) e seu uso em avaliações de imóveis é mais recente ainda (anos 1980). Infelizmente, esses modelos ainda estão sendo pouco utilizados em relação aos outros modelos econométricos tradicionais de regressão, o que ocasiona um certo engessamento e dificuldade de vencer determinados paradigmas, tais como o elevado grau de discricionariedade da construção da matriz de distâncias W, que pode ser construída sob vários panoramas, conforme vimos no item 2.3.

A dependência espacial criada por meio da interação entre os preços dos imóveis vizinhos causa o efeito chamado de defasagem espacial, conhecido como efeito de vizinhança que influi na determinação dos preços. Ou seja, quando há intenção entre um vendedor e um comprador em transacionar um imóvel, é levado em consideração, não apenas características locacionais, de padrão construtivo ou estruturais deste próprio imóvel, mas também o preço dos imóveis vizinhos a ele. Portanto, a estimação do modelo apenas por MQO não é adequada, vez que seus parâmetros podem ser enviesados e inconsistentes.

No modelo de defasagem espacial, em inglês *Spatial Auto Regressive ou Spatial Lag Model* (SAR), a autocorrelação espacial presente é atribuída ao comportamento endógeno da variável dependente preço (Y), por meio da seguinte expressão (4):

$$Y = \rho WY + X\beta + \varepsilon \qquad (5),$$

Onde:

Y: variável preço dependente endógena, representada por um vetor coluna de respostas Y_i de n observações;

X: é o vetor contendo as variáveis independentes exógenas;

β: coeficientes de regressão dessas variáveis independentes;

ε: erros aleatórios com média zero e variância σ^2 constante, contendo os resíduos ε_i da equação;

W: matriz de vizinhança espacial, de contiguidade ou distâncias;

ρ: coeficiente espacial autorregressivo, corresponde ao efeito médio da variável dependente relativo à vizinhança espacial na região em questão (no caso de $\rho=0$, a autocorrelação espacial dos preços investigada não existe).

A ideia dos modelos SAR é utilizar a mesma ideia dos modelos AR (autorregressivos) em séries temporais, por meio da incorporação de um termo de defasagem (*lag*) entre os regressores da equação. A sugestão tradicional para a estimação do coeficiente é a utilização do estimador de mínimos quadrados ordinários (MQO). No entanto, quando o vetor de covariáveis (variáveis do lado direito da equação) é correlacionado com o resíduo da regressão (endogeneidade da variável dependente) sabe-se que o estimador MQO é inconsistente ou enviesado, resultando em uma estimativa inconsistente para o coeficiente (ponderador) "ρ". Conforme lição de Ywata e Albuquerque (2011), utiliza-se da estimação via máxima verossimilhança, "que não sofre do problema de inconsistência do estimador de mínimos quadrados ordinários, devido à endogeneidade do regressor WY, para estimar os parâmetros ρ e σ2, bem como da distribuição normal multivariada para o vetor de resíduos ε". Os mesmos autores também ensinam que podemos testar a significância do parâmetro ρ, utilizando-se o teste de Wald, o teste da razão de verossimilhança ou o teste dos multiplicadores de Lagrange (LM), que, a rigor, foi o escolhido nesse trabalho.

2.6. Modelo de defasagem espacial dos resíduos (SEM)

Diferentemente do modelo anterior, nesse caso a interação espacial ocorre nos resíduos devido a efeitos não modelados e que não estão distribuídos aleatoriamente no espaço (ALMEIDA, 2012).

Os modelos SEM também partem da especificação de modelos autoregressivos (AR) e de modelos de médias móveis (MA) para observações no tempo. Note-se que, ao contrário dos modelos SAR, os modelos SEM não apresentam a variável resposta como uma função direta dos seus *lags* espaciais. A autocorrelação espacial nos modelos SEM aparece nos termos de erro (diferenças entre os valores observados e estimados no tratamento estatístico, por meio do MQO).

De acordo com Dantas, Magalhães e Vergolino (2007), a autocorrelação espacial nos erros ocorre quando se utiliza variáveis *proxies* resultantes da divisão artificial das unidades geográficas na área em estudo. Por exemplo, a utilização da variável renda do responsável por setor censitário (dado do IBGE), ou a utilização dos índices de desenvolvimento humano por bairro (IDH), sem uma devida suavização destas em uma superfície, podem ocasionar erros de medida pelo efeito de transbordamento[2], que acaba fazendo com que a variável se propague para além da sua fronteira. O modelo do erro espacial pode ser expresso formalmente conforme a Equação 5:

$$Y= X\beta+\lambda W_{\varepsilon}+ \xi \qquad (6),$$

Y: variável preço dependente endógena, representada por um vetor coluna de respostas Y_i de n observações;

X: é o vetor contendo as variáveis independentes exógenas;

β: coeficientes de regressão dessas variáveis independentes;

W_{ε} = matriz de erros com efeito espacial;

ξ = erros aleatórios com média zero e variância σ^2;

λ = coeficiente autorregressivo.

O vetor de resíduos ξ possui distribuição normal multivariada, com média nula e matriz de covariância $\sigma^2 I$. O coeficiente escalar λ indica a intensidade da autocorrelação espacial entre os resíduos da equação observada. Mais especificamente, esse parâmetro mensura o efeito médio dos erros dos vizinhos em relação ao resíduo da região em questão.

Dantas, Magalhães e Vergolino (2007), ensinam que, nos casos em que há dependência espacial dos erros, a estimação do modelo apenas por MQO conduziria a estimativas não viesadas, mas ineficientes dos parâmetros (β), devido à estrutura não diagonal da matriz de variância dos resíduos, havendo outros estimadores lineares que produzem variâncias menores.

Ainda, Ywata e Albuquerque (2011) aduzem que, especificamente para o modelo SEM, o estimador linear com variância mínima, que traria melhores resultados, é o estimador de mínimos quadrados generalizados – MQG (*generalized least squares* – GLS).

2.9 Medidas de desempenho e acurácia

Como medidas de desempenho e acurácia para comparação entre os diversos métodos, as seguintes métricas definidas pelos padrões da *International Association of Assessing Officers* (IAAO*,* 2013). Assim as seguintes medidas foram utilizadas no trabalho:

i) **Nível de avaliação** (*sales ratio*) mediano (***SR$_m$***):

$$SR_m = mediana\ de \frac{valor\ estimado}{valor\ observado} \qquad (7),$$

ii) **Coeficiente de dispersão** (**COD**):

$$COD = \frac{100}{SR_m} \times \left(\frac{\sum_{i=1}^{n} |SR_i - SR_m|}{n} \right) \qquad (8),$$

onde: SRi é o nível de avaliação de cada terreno individualmente considerado e n é o número total de

[2] Cabe aqui uma observação importante. Nesse trabalho, utilizamos algumas variáveis espaciais. Entretanto, apenas a mais importante (renda) foi suavizada, por meio de técnicas de geoestatística (*Krigagem*).

dados da amostra

iii) Média percentual absoluta do erro (MAPE):

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \left| \frac{valor\ observado_i - valor\ estimado_i}{valor\ observado_i} \right| \quad (9)$$

iv) Raiz quadrada da média dos erros ao quadrado (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(valor\ observado_i - valor\ estimado_i)^2}{n}} \quad (10)$$

v) Coeficiente de determinação (r²):

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(\text{valor observado}_i - \text{valor estimado}_i)^2}{\sum_{i=1}^{n}(\text{valor observado}_i - \text{valor observado}_{médio})^2} \quad (11)$$

3. RESULTADOS E DISCUSSÃO

3.1. Resultados estatísticos do modelo MQO

As tabelas 2 e 3 abaixo trazem resultados estatísticos do modelo MQO, já a tabela 4 traz o teste de hipótese de especificação linear do modelo (teste de RESET).

Tabela 2 – Estimativa dos parâmetros do modelo MQO.

Variável	Parâmetro	Erro padrão	Estatística t	ρ-value
(Intercepto)	4,5300	0,471	9,62	$< 2x10^{-16}$
distvp	- 0,0005	0,000	-8,01	$2,74x10^{-15}$
LN (distsh)	- 0,0338	0,025	-1,34	0,18146
LN (dscom)	0,0318	0,010	3,17	0,001571
dassprec	0,0001	0,000	1,16	0,248162
LN (vbt)	0,3615	0,029	12,65	$< 2x10^{-16}$
LN (renda)	0,0830	0,029	2,86	0,004274
idhed	1,2540	0,441	2,85	0,004486
LN (iaeq)	0,0597	0,031	1,93	0,053867
lotcnd	0,8978	0,159	5,66	$1,89x10^{-8}$
LN (numfr)	0,0982	0,045	2,18	0,029781
test	0,0072	0,002	3,77	0,000172
LN (area)	- 0,0571	0,029	-1,96	0,050825
agua	0,1232	0,043	2,84	0,004657
esg	0,0626	0,037	1,69	0,091236
pav	0,1213	0,036	3,35	0,000837
dv	0,0079	0,005	1,48	0,138854

Fonte: Elaborado pelos autores (2020)

Cabe destacar que, algumas variáveis independentes, bem como a dependente preço, tiveram seus valores transformados pelo logaritmo natural (neperiano – LN). Com respeito aos valores de ρ-value calculados, todas as variáveis tiveram seus valores calculados abaixo de 10%, exceto as variáveis distsh(18,1%), dassprec(24,8%) e dv(13,9%), enquadrando o trabalho no grau I de fundamentação, segundo a NBR 14.653.

É importante observar que, na busca da necessária linearização da dispersão dos dados para utilização do método dos mínimos quadrados (teste de White), adotamos transformações de escalas na variável dependente preço. Além disso, esta transformação apresenta também outras características desejáveis a serem consideradas, vez que certamente produziremos um modelo não tendencioso, mais eficiente e consistente (melhor coeficiente de determinação).

Vale lembrar que as variáveis independentes inicialmente são testadas sem transformação, depois usamos o logaritmo natural, para saber se há vantagens, tais como o maior controle naquelas que apresentam grande amplitude e dispersão de valores. A quantidade dessas variáveis também e objeto de estudo e controle, pois partimos de dezenas delas, mantidas no cadastro territorial municipal (CTM) e observatório urbano de valores (OUV), para finalmente modelar as poucas que se tornam estatisticamente significantes e relevantes para explicar a variabilidade dos preços.

Essas variáveis apresentaram crescimento positivo com o preço, exceto distvp (distância a via principal), distsh (distância a shopping centers) e área do terreno. As duas primeiras representam distâncias a equipamentos valorizantes da cidade. Portanto, apresentam crescimentos inversamente proporcionais aos preços. Já o crescimento negativo da área do terreno em relação à variável dependente é explicado pela teoria da utilidade marginal decrescente.[3]

Tabela 3 - Estatísticas do modelo MQO

Parâmetro	Regressão	Estimativa
Coeficiente de Correlação	0,8048	0,7894
Coeficiente de Determinação	0,6477	0,6232
Coeficiente de Determinação Ajustado	0,6429	0,6233
Desvio Padrão	0,4838	536,024
Grau de liberdade	1.157	
(F) Fischer/Snedecor	133	
Significância (F)	0,01	
Número de dados amostra	1.174	
Número de variáveis independentes	17	

Fonte: Elaborado pelos autores (2020)

Cabe observar que, após diversas interações, encontramos como melhor coeficiente de determinação ajustado ($r_a^2 = 64,3\%$), valor que ainda não é o ideal. Entretanto, analisando os testes de RESET (tabela 4) e de Fischer/Snedecor (F), este modelo foi o mais adequado.

Tabela 4 - Teste de hipótese de RESET – Especificação

Parâmetro	VALOR
H_0	Não existe erro de especificação.
H_1	Os parâmetros do modelo não são lineares
RESET	1,3928
df1	2
df2	1.155
ρ-value	0,2488 (H_0)

Fonte: Elaborado pelos autores (2020)

[3] A teoria define que há uma relação econômica em que o valor de um determinado bem ou serviço diminui à medida que o seu consumo é feito em larga escala. Isso significa dizer que, de acordo com a teoria, um produto passa a custar menos a partir do momento em que ele se torna muito abundante, fazendo com que a sua utilidade marginal diminua.

Concluímos pela tabela 4, que os parâmetros do modelo MQO possuem comportamento linear, quando os comparamos com a variável dependente (preço), ou seja, não há erro de especificação, acatando o pressuposto 1 do item 2.2 deste trabalho.

3.2. Equação do modelo MQO

Abaixo, a equação do modelo MQO:

$$\text{valor} = e^{(4{,}529803371 - 0{,}0005355931139 \cdot \text{distvp} - 0{,}03381640254 \cdot \ln(\text{distsh}) + 0{,}03181996221 \cdot \ln(\text{dscom}) + 7{,}02188987\text{E-}005 \cdot \text{dassprec} + 0{,}3614526341 \cdot \ln(\text{vbt}) + 0{,}08296634787 \cdot \ln(\text{renda}) + 1{,}254471275 \cdot \text{idhed} + 0{,}05972745559 \cdot \ln(\text{iaeq}) + 0{,}8978081199 \cdot \text{lotcnd} + 0{,}09822196524 \cdot \ln(\text{numfr}) + 0{,}0071825371 \cdot \text{test} - 0{,}05709807195 \cdot \ln(\text{área}) + 0{,}1232308769 \cdot \text{agua} + 0{,}06263818136 \cdot \text{esg} + 0{,}1213329076 \cdot \text{pav} + 0{,}007927164532 \cdot \text{dv})} \quad (12)$$

3.3. Matriz de vizinhança e Índice Global de Moran I

A matriz de vizinhança foi montada pelo inverso da distância, ponderada por linha e com raio de determinação dos vizinhos de 1.394m que é o valor mínimo para que todos os dados tenham pelo menos um vizinho. O índice global de Moran I indicou dependência espacial dos resíduos do modelo MQO indicando a necessidade de se utilização de um modelo econométrico espacial. Sua estatística foi de 0,3228, indicando autocorrelação positiva global com *ρ-value* de $2{,}2 \times 10^{-16}$ (rejeitando a hipótese nula de não autocorrelação).

3.4. Testes dos multiplicadores de Lagrange para autocorrelação espacial

Tabela 5 - Estatística dos multiplicadores de Lagrange

Modelo	Estatística do teste	AIC
LMErro	195,93	1.431,289
LMLag	188,42	1.445,416

Fonte: Elaborado pelos autores (2020)

Os multiplicadores de Lagrange Robusto do Erro ($LM\lambda$) e da Defasagem - Lag (LM_ρ) são usados para definir qual modelo de dependência espacial terá melhor comportamento. Segundo a Tabela 5, ambos foram significantes (apresentaram-se significativos ao nível de 1%), ou seja, os dois modelos espaciais (defasagem e erro) poderiam servir. Entretanto, verifica-se que a maior estatística dos testes (195,93) e menor dispersão de Akaike (1.431,29) se referem ao modelo de erro espacial (SEM), sendo esse, então, o mais indicado a ser utilizado com o coeficiente $\lambda = 0{,}50295$ (ρ-value $< 2{,}22 \times 10^{-16}$ – significante).

3.5. Comparativo de desempenho e acurácia

Tabela 6 – Comparativo de desempenho e acurácia entre os modelos.

Medida/Modelo	MQO	SAR	SEM[4]
Nível de avaliação	1	0,99	0,99
COD (%)	41,36%	35,34%	34,75%

[4] Em todos os índices testados, o modelo SEM (modelo espacial dos erros/resíduos) teve melhor comportamento que os demais, corroborando a conclusão que tivemos, ao calcular os multiplicadores de *Lagrange* acima.

RMSE (R$/m²)	532,13	484,52	474,84
MAE (R$/m²)	329,97	286,35	279,29
MAPE (%)	41,25%	35%	34,40%
Coef. Determinação (r²)	0,62	0,69	0,70

Fonte: Elaborado pelos autores (2020)

Analisando os diversos resultados extraídos das respectivas modelagens, verificou-se que o modelo SEM possui um poder de explicação melhor(r²) que os demais. Pelos demais testes, percebemos que esse modelo espacial possui menor variância (dispersão) dos desvios e deve apresentar os melhores resultados, sendo, assim, a melhor escolha. O coeficiente de dispersão da mediana (COD) ainda é alto, vez que são recomendados valores menores que 20% para terrenos e encontramos valores próximos a 35%, cabendo averiguar se há áreas com poucos dados e/ou analisar fronteiras geográficas, por exemplo. Todos os modelos tiveram nível de avaliação iguais a 1 (um), o que indica um desejável ajuste equilibrado nas projeções em relação aos preços observados.

4. CONCLUSÃO

O trabalho envolveu a realização de um estudo de caso que comparou os modelos clássicos de regressão linear múltipla pelo MQO e de regressão espacial sobre amostra de terrenos ofertados no Município de Fortaleza/CE, num intervalo de uma ano, cujos resultados levaram a conclusão de melhor performance do modelo econométrico do erro espacial (SEM).

Nesse diapasão, foi realizada a escolha das variáveis mais apropriadas, muitas delas relacionadas à localização de cada imóvel, com base em testes estatísticos, e se estimou o modelo de regressão por meio de regressão linear multivariada MQO.

Como visto, verificou-se a existência de autocorrelação espacial nos resíduos MQO, apesar deste não apresentar erro de especificação, pois as formas de relacionamento da variável dependente e independentes foram lineares. Essa autocorrelação espacial é comum nos modelos de avaliação em massa, o que acarreta restrições graves nos seus pressupostos essenciais.

Assim, como inovação, esse trabalho procurou sanar esses problemas, utilizando um modelo econométrico puramente espacial, dado que se evidenciou a correlação espacial dos resíduos e da variável dependente, quando se aplicou o teste de Moran (I), após a construção da matriz de vizinhança W. Testou-se várias alternativas de construção de matrizes de vizinhança, sendo a com base no inverso da distância entre os pontos (d^{-1}) a que apresentou a estatística de Moran's I mais significante. Verificou-se também, que a utilização de todos os dados para a construção da matriz de vizinhança acarretou um grande esforço computacional, mesmo usando bibliotecas em R e Python. Adotou-se uma distância mínima de 1.394m entre os pontos, que correspondeu a garantia de que cada ponto tivesse pelo menos um vizinho.

Por fim, observando o teste de LM robusto da defasagem espacial e dos erros, bem como os padrões da IAAO, concluiu-se que a metodologia espacial do erro (SEM), com enfoque na correlação espacial dos resíduos, tornou-se a melhor escolha.

Em decorrência do trabalho apresentado, vislumbram-se algumas recomendações voltadas ao desenvolvimento de novas pesquisas, bem como algumas aplicações práticas.

Primeiramente, sugere-se uma maior aplicação de técnicas de econometria espacial na engenharia de avaliação de imóveis, vez que há uma escassez de produções científicas que utilizem essa metodologia, promovendo um maior destaque da variável localização, por ser, indiscutivelmente, aquela que apresenta maior correlação com os valores de imóveis.

Além disso, o estudo em questão pode ser aperfeiçoado com o uso de técnicas combinadas de interpolação geoestatística (*krigagem*), trazendo como um dos grandes benefícios o cálculo da área de contágio espacial, que por sua vez poderia ser utilizado na matriz de distâncias.

REFERÊNCIAS

ABNT. **NBR 14653-1**: avaliação de bens. Rio de Janeiro: [s.n.], 2019.

ABNT. **NBR 14653-2**: avaliação de bens parte 2: imóveis urbanos. Rio de Janeiro: [s.n.], 2011.

ALMEIDA, Eduardo. **Econometria Espacial Aplicada**. Campinas, SP: Editora Alínea, 2012.

ANSELIN, Luc; REY, Sergio J. **Modern Spatial Econometrics in Practice**: A Guide to Geoda, Geodaspace and Pysal. Chicago: GeoDa Press LLC, 2014. 394p.

BANDEIRA, Sandro R. V. **Regressão espacial e avaliação de Terrenos**: Um estudo de caso para a cidade de Fortaleza/CE, Dissertação (Mestrado em Economia) – FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E CONTABILIDADE – FEAAC – Universidade Federal do Ceará, Fortaleza, 2019.

DANTAS, Rubens Alves; MAGALHÃES, André Matos; VERGOLINO, José Raimundo de Oliveira. **Avaliação de imóveis**: a importância dos vizinhos no caso de Recife. Econ. Apl., Ribeirão Preto, v. 11, n. 2, 2007.

INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS - IAAO. **Standard on Ratio Studies**, 2013. Standard on Mass Appraisal of Real Property. 2017.

SILVA, Everton da. **Cadastro técnico multifinalitário**: base fundamental para avaliação em massa de imóveis. 2006. 219 f. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis-SC, 2006.

VALDUGA, Leonel de Moura Brizola; ZANCAN, Evelise Chemale. **Utilização de modelo de regressão espacial para avaliação de terrenos na Cidade de Criciúma, SC.** Artigo submetido ao Curso de Engenharia Civil da Universidade do Extremo Sul Catarinense - UNESC, Criciúma-SC, 2018

YWATA, Alexandre Xavier de Carvalho. ALBUQUERQUE, Pedro Henrique de Melo. Métodos e Modelos em Econometria Espacial. **Revista Brasileira de Biomassa**, São Paulo, v. 29, n. 2, p. 273-306, 2011.