



UNIVERSIDADE FEDERAL DO CEARÁ – UFC
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E
CONTABILIDADE – FEAAC
PROGRAMA DE ECONOMIA PROFISSIONAL – PEP

SANDRO RICARDO VASCONCELOS BANDEIRA

REGRESSÃO ESPACIAL E AVALIAÇÃO DE TERRENOS: UM ESTUDO DE CASO
PARA A CIDADE DE FORTALEZA-CE

FORTALEZA

2019

SANDRO RICARDO VASCONCELOS BANDEIRA

**REGRESSÃO ESPACIAL E AVALIAÇÃO DE TERRENOS: UM ESTUDO DE CASO
PARA A CIDADE DE FORTALEZA-CE**

Dissertação submetida à Coordenação do Programa de Economia Profissional – PEP, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do grau de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Orientador: Prof. Dr. Andrei Gomes Simonassi

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- B164r Bandeira, Sandro Ricardo Vasconcelos.
 Regressão espacial e avaliação de terrenos: um estudo de caso para a Cidade de Fortaleza-Ce /
 Sandro Ricardo Vasconcelos Bandeira. – 2019.
 67 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia,
 Administração, Atuária e Contabilidade, Mestrado Profissional em Economia do Setor Público,
 Fortaleza, 2019.
 Orientação: Prof. Dr. Andrei Gomes Simonassi.
1. Regressão espacial. 2. Avaliação de imóveis. 3. Tributação Imobiliária Municipal. I. Título.

CDD 330

SANDRO RICARDO VASCONCELOS BANDEIRA

**REGRESSÃO ESPACIAL E AVALIAÇÃO DE TERRENOS: UM ESTUDO DE CASO
PARA A CIDADE DE FORTALEZA-CE**

Dissertação submetida à Coordenação do Programa de Economia Profissional – PEP, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do grau de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Aprovada em: **20 de dezembro de 2019.**

BANCA EXAMINADORA

Prof. Dr. Andrei Gomes Simonassi (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Henrique Felix Silva
Universidade Federal do Ceará (UFC)

Prof. Dr. Fabrício Carneiro Linhares
Universidade Federal do Ceará (UFC)

À Deus.

Aos meus pais, Maria do Carmo e Silo Renan.

À minha esposa, Liana, e aos meus filhos,
Rafael e Eduardo.

Ao meu irmão Cássio Bandeira e ao grande
amigo Antônio Augusto F. Oliveira.

AGRADECIMENTOS

Agradeço à Deus, grato pelo precioso dom da vida.

Aos meus pais, pelo imenso amor e dedicação.

À minha esposa e filhos, pela vivência de uma família feliz.

Ao meu irmão Cássio, pelos incentivos nas horas mais difíceis.

Aos colegas da Secretaria de Finanças de Fortaleza e em modo especial, ao amigo-irmão Augusto Oliveira, cuja imensurável e inesgotável contribuição tornou esse trabalho possível.

RESUMO

Os tributos imobiliários possuem em comum o fato de possuírem como base de cálculo o valor venal de mercado, 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 mercado, com o uso de inferência estatística, por meio do método dos mínimos quadrados ordinários (MQO). Ocorre que, ao se inserir no estudo estatístico a primeira lei da geografia de Waldo Tobler (1970), percebe-se que a localização de imóvel, por si só, exerce papel de suma importância na definição deste valor, assumindo, portanto, uma notável contribuição na formação dos preços dos imóveis. Foi proposto dois macromodelos, o primeiro contemplou somente a análise clássica, por meio dos mínimos quadrados ordinários (MQO), com ajustes dos intervalos de confiança dos parâmetros, por meio dos procedimentos de Newey-West (MQG/HAC) e, por fim, uma análise usando ferramentas de econometria espacial, apreciando também os comportamentos espaciais da variável dependente preço (SAR) e os resíduos do modelo MQO (SER). Ao final, medir-se-á a performance de cada uma das abordagens sobre uma amostra de teste não participante da modelagem inicial, comparando-se a precisão dos modelos, por meio das estatísticas de Akaike e Schwarz para definir aquele que apresenta os resultados mais consistentes. Nesse trabalho, foi 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, na qual se chegou à conclusão que o uso do modelo espacial, com base na contribuição espacial do preço (SAR) produziu os melhores resultados dentre todos os testados.

Palavras-chave: Regressão espacial. Avaliação de imóveis. Tributação Imobiliária Municipal.

ABSTRACT

The real estate taxes have in common the fact that they have market value as their basis of calculation, that can be described as the most probable amount for which a good would be voluntarily and consciously negotiated, on a reference date, within the conditions of the current market. 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 Least Squares (MLS). 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 value, assuming, therefore, a notable contribution to form property prices. Two macromodels were proposed where the first included only classical analysis, using the Method of Least Squares (MLS), considering safe parameters adjustments by using Newey-West procedures and finally by proposing an analysis through spatial econometric tools, also appreciating the spatial behaviors of price as a dependent variable (SAR) and the MLS model residuals (SER). 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 (capitol 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 Price produced the best results among all those tested.

Keywords: Spatial regression. Property valuation. Municipal Real Estate Taxation.

LISTA DE FIGURAS

Figura 1 - Fluxograma das etapas utilizadas na pesquisa de regressão espacial.....	20
Figura 2 - Tipos de matrizes de contiguidade (Queen, Rook e Bishop).....	27
Figura 3 - Forma da matriz de distâncias W (914 x 914).....	46

LISTA DE GRÁFICOS

Gráfico 1 - Multicolinearidade, normalidade e independência serial dos resíduos e de dispersão dos valores estimados e observados do modelo MQO.....	61
Gráfico 2 - Índice de MORAN (I) para os resíduos padronizados do modelo MQO.....	62
Gráfico 3 - Resíduos da Regressão espacial (SAR) – Amostra de teste.....	64
Gráfico 4 - Disposição espacial das amostras de treinamento e validação (teste).....	65

LISTA DE QUADROS

Quadro 1 - Equações dos modelos utilizados no trabalho.....	58
Quadro 2 - Descrição das variáveis utilizadas nos modelos.....	66

LISTA DE TABELAS

Tabela 1 - Níveis de multicolinearidade.....	39
Tabela 2 - Cálculos estatísticos do modelo MQO.....	42
Tabela 3 - Fator de inflação da variância das variáveis do modelo MQO.....	44
Tabela 4 - Resultados estatísticos do modelo de defasagem espacial e erro espacial.....	47
Tabela 5 - Medidores de acurácia da IAAO.....	49
Tabela 6 - Cálculos estatísticos dos modelos MQO, MQG (HAC) e SAR.....	57
Tabela 7 - Matriz de correlação entre variáveis do modelo MQO.....	59
Tabela 8 - Elementos da matriz W padronizada para os 13 vizinhos do ponto 852, calculado pelo GeoDa.....	63

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO DA LITERATURA	14
3	ASPECTOS METODOLÓGICOS	19
3.1	Base de dados	19
3.2	Fluxograma da pesquisa	19
3.3	Método de regressão linear multivariado, com base nos mínimos quadrados ordinários (MQO)	21
3.4	Análise da dependência espacial dos dados	25
3.4.1	<i>Matriz W de distância (ou contiguidade)</i>	26
3.4.2	<i>Testes para detectar autocorrelação espacial</i>	28
3.4.2.1	<i>Teste de Moran (I)</i>	29
3.4.2.2	<i>Testes de LM robusto (erro e defasagem)</i>	30
3.5	Método de regressão espacial	31
3.5.1	<i>Modelo de Defasagem Espacial (SAR)</i>	33
3.5.2	<i>Modelo de Defasagem Espacial dos Resíduos (SER)</i>	35
3.6	Críticas aos modelos de dependência espacial	36
3.7	Medidas de desempenho recomendadas pela IAAO	37
3.8	Estatísticas de Akaike (AIC) e Schwarz (BIC)	39
4	RESULTADOS	41
4.1	Análise dos dados por meio do modelo MQO – preços hedônicos	42
4.2	Análise dos dados por meio de modelos econométricos espaciais	45
4.2.1	<i>Análise da dependência espacial – índice de Moran (I) e multiplicadores de Lagrange (testes LM)</i>	45
4.2.2	<i>Dimensionamento da matriz de distâncias (W)</i>	45
4.2.3	<i>Escolha do modelo da dependência espacial, por meio dos multiplicadores de Lagrange (testes LM)</i>	47
4.2.4	<i>Análise de desempenho dos modelos clássicos (MQO) e espacial por defasagem, pelos critérios da IAAO (International Association of Assessing Officers)</i>	49
4.2.5	<i>Distribuição dos resíduos da regressão espacial - SAR</i>	50
5	CONCLUSÃO	51
	REFERÊNCIAS	53
	APÊNDICES	57
	ANEXOS	66

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. 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 previsão.

O interesse em se estimar 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 comerciais em geral, para estimar valores de seu ativo imobilizado, bancos e os entes governamentais.

Do ponto de vista governamental, por exemplo, o conhecimento do preço marginal que o consumidor está disposto a pagar por um determinado imóvel, também chamado de preço implícito ou hedônico, pode ser de grande interesse na elaboração e implantação de políticas habitacionais e urbanas, uma vez que, no longo prazo, são as preferências do consumidor que determinam a configuração das cidades. É extremamente racional a ideia 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, a ser detalhada no corpo desse trabalho.

Nesse contexto, propõe-se utilizar o modelo de regressão espacial como solução inovadora, complementando o tradicional modelo de regressão multivariada (clássico – MQO). Assim, o trabalho visou apresentar uma sistemática de cálculo enquadrada nos trâmites da norma brasileira – NBR 14.653 – com a introdução do viés espacial, a fim de que se possa apurar a influência de elementos vizinhos entre si, bem como mostrar como se comportam os imóveis nas proximidades de polos valorizantes ou desvalorizantes. A título de estudo de caso, utilizou-se de uma amostra composta apenas de dados territoriais. Entretanto, não há nenhum óbice para a expansão do estudo, com sua aplicação em imóveis prediais.

Outro ineditismo do trabalho pode ser observado quando usamos o modelo MQG (método dos mínimos quadrados generalizados) para ajustar o modelo tradicional MQO e promover estimativas mais consistentes dos parâmetros β das variáveis, diante da esperada e indesejada heterocedasticidade dos resíduos.

Considerando o exposto, o trabalho envolveu a realização de um estudo de caso que comparou os modelos de regressão clássico e espacial nos terrenos do município de Fortaleza/CE, cujos resultados levaram a conclusão pela opção do modelo de regressão espacial.

Além desta introdução, o trabalho está estruturado em mais quatro seções. Na segunda seção, que trata sobre a revisão da literatura, evidenciam-se alguns trabalhos relevantes publicados sobre a aplicação da econometria espacial na avaliação dos imóveis urbanos.

Na terceira seção, apresentam-se os aspectos metodológicos, com algumas evidências empíricas que motivaram a análise mais robusta feita no exercício empírico.

A quarta seção expõe o estudo de caso, mostrando os resultados da comparação entre os modelos de regressão linear multivariada e espacial aplicados para a avaliação dos valores dos terrenos do município de Fortaleza/CE.

Por fim, no quinto capítulo, apresentam-se as considerações finais do trabalho com a sugestão de novos horizontes a construir.

2 REVISÃO DA LITERATURA

Em função da extensa literatura a respeito do tema em questão, esse capítulo se restringirá a apresentar alguns trabalhos que utilizaram modelos econométricos espaciais para analisar a dependência espacial dos valores de imóveis, seja por meio dos resíduos (Modelo SER) ou pela própria defasagem espacial dos próprios preços (Modelo SAR).

Preliminarmente, Olmo (1995) investigou uma hipótese amplamente aceita de que o preço da habitação é determinado por dois tipos de características: estrutural e localização. Sob essa hipótese, uma metodologia foi apresentada para estimar a parcela espacial dos preços dos imóveis e também seus aluguéis: o método iterativo e residual por *Krigagem* (IRK). Os aspectos metodológicos são complementados pela apresentação de um estudo de caso para a cidade de Granada, Espanha.

Posteriormente, Trivelloni e Hochheim (2006) apresentaram um método para a determinação do valor da localização dos imóveis usando técnicas de econometria espacial e geoestatística. Relataram que, dentre todas as variáveis que influenciam no valor dos imóveis, as de localização são as mais complexas de analisar e modelar. Ademais, a multiplicidade de fatores ambientais, sociais e econômicos que influenciam no valor dos imóveis é difícil de ser modelada adequadamente pelos métodos inferenciais tradicionais, provocando problemas de especificação nos modelos e autocorrelação espacial nos resíduos, comprometendo a confiabilidade da avaliação. Os modelos de estatística espacial, em especial os de regressão espacial e os métodos geoestatísticos (*Krigagem*) foram usados em forma combinada para estimar o valor da localização dos imóveis.

Na oportunidade, uma aplicação do método foi realizada para uma amostra de mercado de dois bairros do Município de São José, Santa Catarina, Brasil. Assim, o método proposto permitiu encontrar um índice do valor da localização fortemente significativo no modelo de regressão para todos os tipos de imóveis da região e que modelava de forma consistente o efeito do conjunto de polos de valorização na região. O método eliminou toda autocorrelação espacial nos resíduos do modelo de regressão melhorando o poder explicativo e a confiabilidade da avaliação.

No mesmo ano, Silva (2006) pesquisou o cadastro técnico multifinalitário como base fundamental para avaliação em massa de imóveis, com o objetivo de tratar de assuntos relacionados ao valor da propriedade, a fim de contribuir para maior qualidade e transparência em questões como a tributação imobiliária, já que existia carência de procedimentos para a modernização do aparelhamento fiscal dos municípios. Nesse contexto, a criação de um banco

de dados do mercado imobiliário integrado ao cadastro técnico multifinalitário, como um dos objetivos daquela pesquisa, foi uma condição necessária ao estabelecimento de um processo contínuo de avaliação coletiva dos imóveis e a solidificação de uma cultura técnica no que diz respeito ao valor da propriedade, que atenderiam com maior eficácia as decisões políticas do Poder Público Municipal.

No referido trabalho, observou-se que o cadastro técnico possibilitou a elaboração de análises espaciais e estatísticas no sentido de se produzir uma avaliação em massa de imóveis fundamentada nos dados que o compõe. Assim, um modelo de regressão múltipla foi ajustado a uma amostra de terrenos pesquisados em uma porção da cidade de Blumenau/SC, levando a resultados que possibilitaram a sua aplicação no desenvolvimento de uma planta de valores genéricos de terrenos.

No ano seguinte, Dantas, Magalhaes e Vergolino (2007) apresentaram a importância dos imóveis vizinhos na avaliação em massa na cidade de Recife/PE, onde observaram que os dados associados à posição que ocupam no espaço eram caracterizados pela dependência espacial. Seguindo a literatura internacional à época, este trabalho reconheceu a importância da questão espacial e mostrou como se poderia diagnosticar e incorporar os efeitos espaciais na especificação de modelos hedônicos, ao mesmo tempo em que analisou o mercado imobiliário da cidade do Recife. Os resultados indicaram que a consideração da questão espacial, com base em distâncias a polos de influência, não foi capaz de explicar completamente as variações dos preços em relação à localização da habitação, uma vez que existia uma verdadeira interação entre os dados, de forma que cada edifício funcionava como um micropolo de influência sobre os seus vizinhos. Isso significava dizer que as negociações de compra dos apartamentos não ocorreriam de forma independente, como considerado nos modelos tradicionalmente estimados, mas que existia uma verdadeira interação entre estes preços, de forma que uma negociação de um apartamento por um preço elevado tenderia a elevar os preços dos imóveis vizinhos. Assim sendo, a solução para estimação segura da equação de preços hedônicos, na presença de efeitos espaciais, deveria ser feita com base na econometria espacial, levando em consideração todas as interações espaciais possíveis entre os dados e servindo como uma *proxy* para variáveis locais não consideradas explicitamente no modelo.

Ywata e Albuquerque (2011) discutiram diversos modelos econométricos para estimação de modelos paramétricos na presença de dependência espacial, com dados *cross-section*¹. Inicialmente, foram trabalhados modelos de dependência espacial com lags espaciais

¹ Dados de corte transversal ou *cross-section* são uma amostra de dados usados em estatística ou econometria, sejam indivíduos, consumidores, empresas, cidades, etc., em um dado ponto no tempo.

da variável resposta ou *lags* espaciais do resíduo, com estimação dos parâmetros feita via estimativa por máxima verossimilhança. Em seguida, uma análise crítica destes modelos foi apresentada, com discussão de testes para detectar presença de dependência espacial.

Florêncio (2012) investigou modelos espaciais para avaliações de imóveis em massa, de relevante interesse para a administração pública, especialmente para fins tributários, com base em modelos aditivos generalizados para posição, escala e forma (GAMLSS). Contudo, relatou que as características singulares dos imóveis fazem com que a análise dos valores dos bens seja uma tarefa complexa e o emprego de técnicas não apropriadas podem resultar em modelos irrealistas e de conclusões equivocadas. Citou que, em parte, as deficiências dos modelos de avaliação em massa estão relacionadas aos seguintes aspectos: (i) a teoria não determina a forma funcional nem as variáveis relevantes para a estimação da função de preços hedônicos e (ii) os modelos estatísticos tradicionalmente utilizados podem apresentar estimadores viesados, ineficientes ou inconsistentes quando defrontados com não-normalidade, heterocedasticidade e autocorrelação espacial, bastante comuns em dados imobiliários.

Assim, visando lidar com estas dificuldades, o mencionado trabalho propôs uma metodologia para avaliações em massa, baseada no uso de modelos espaciais a partir do emprego da classe de modelos de regressão proposta por Rigby e Stasinopoulos (2005 *apud* FLORÊNCIO, 2012), a qual permitia o ajuste de uma ampla família de distribuições para a variável resposta e possibilitava a modelagem direta, utilizando funções paramétricas e/ou não-paramétricas, da estrutura de regressão da variável de interesse. Dessa forma, incorporados os efeitos da correlação espacial nos modelos GAMLSS, mostrou, mediante análise empírica com dados de terrenos urbanos situados em Aracaju/SE, que a estrutura de regressão GAMLSS aparentava ser mais apropriada para a estimação dos parâmetros da função de preços hedônicos do que os modelos usualmente empregados, como o modelo normal de regressão linear clássico.

Fermo e Hochheim (2013) analisaram a influência de erros de medida de posicionamento em modelos de regressão espacial, no qual o diagnóstico da presença de efeitos espaciais foi obtido através de testes estatísticos que utilizaram matrizes de vizinhança, ou matriz de pesos espaciais, cuja determinação necessitou do conhecimento do posicionamento dos imóveis. Este pode ser determinado por meio de georreferenciamento dos dados da amostra. Assim, o trabalho teve como objetivo principal analisar a influência de erros de medidas de posicionamento dos imóveis da amostra no desenvolvimento de modelos de regressão espacial, com foco na avaliação da qualidade de três ferramentas de georreferenciamento: Mapa Cadastral Georreferenciado, GPS de Navegação e o aplicativo Google Earth. No caso estudado, foi selecionada uma amostra de 328 dados de mercado (apartamentos) na cidade de Criciúma

(SC). Após a identificação da ferramenta que fornecia o melhor modelo, foi feita uma simulação para avaliar a sensibilidade do modelo a erros. Foram simulados erros de diversas grandezas nas coordenadas originais. Por fim, os resultados mostraram que a ferramenta mais adequada para fazer o georreferenciamento da amostra é o mapa cadastral georreferenciado, porém as demais ferramentas também poderiam ser usadas sem prejuízos para a avaliação dos imóveis. O estudo também mostrou que o modelo desenvolvido não foi muito sensível a erros na determinação das coordenadas dos imóveis.

Silva, Hochheim e Silva (2016) estudaram a importância da variável localização em modelos de regressão linear e espacial, além do uso de geoestatística em avaliações de imóveis. Assim, foram propostas alternativas de variáveis que poderiam representar e caracterizar a localização do imóvel, como distância ou inserção ao polo valorizante e o uso da renda familiar. Para tal, foi escolhida como região de estudo o bairro Centro em Florianópolis/SC. Para a área de estudo, uma variável que se tornou eficaz foi a renda média do chefe de família, produzido no censo 2010 do IBGE, apresentando ótimos resultados atestando que, para a amostra em questão, os locais com as maiores rendas familiares tendem a apresentar os maiores valores de imóveis, atestando a correlação espacial desta variável.

Valduga e Zancan (2018) apresentaram um estudo da utilização de modelagem por regressão espacial aplicada ao mercado de terrenos da cidade de Criciúma – SC, onde foi obtido uma amostra de 460 dados, aferido e interpretado por 5 variáveis independentes formadoras de valor: Área total, distância à via principal mais próxima, número máximo de pavimentos, renda e avenida centenário. Posteriormente, criou-se uma matriz de vizinhança correspondente ao inverso do quadrado da distância entre os pontos e se realizou um modelo de regressão linear múltipla. Os testes dos multiplicadores de *Lagrange* LM do erro e da defasagem espacial, apresentaram-se significantes ao nível de 1%, confirmando assim a autocorrelação espacial no modelo. Visto isso, realizou-se o modelo do erro espacial e o modelo da defasagem espacial, todos se enquadraram nos critérios da NBR 14.653-2 (2011). Por fim, esses modelos foram comparados por meio dos critérios de Akaike e Schwarz, que representam a qualidade do ajuste, tendo o erro espacial apresentado os menores valores nos critérios em relação aos outros modelos, sendo então o mais indicado a ser utilizado para a avaliação de terrenos na cidade de Criciúma – SC.

Recentemente, Dall’Agnol, Betzek e outros (2018) avaliaram a aplicação do software R para análise geoestatística, interpolação de dados por *krigagem* ordinária e geração de mapas temáticos de produtividade agrícola, onde mostraram a análise geoestatística, interpolação de dados por *krigagem* ordinária e a geração de mapas temáticos de produtividade

de uma área agrícola, visando representar a variabilidade espacial nessa área. Eles utilizaram procedimentos computacionais para análise geoestatística e interpolação de dados, por meio de técnicas e bibliotecas geoestatísticas já existentes no software R. Como característica específica dessa pesquisa foi realizada a implementação para a automatização da escolha do melhor modelo e parâmetros do semivariograma, fundamentais na interpolação de dados por krigagem. Os procedimentos executados foram eficientes e capazes de identificar os melhores ajustes para o semivariograma, conseqüentemente foram criados mapas temáticos com dados interpolados, que apresentam de maneira consistente, a distribuição espacial da produtividade na área pesquisada.

Por fim, Bandeira *et al.* (2019) investigaram a acurácia das predições de avaliação em massa de terrenos urbanos por modelos espaciais no município de Fortaleza/CE com o comparativo de 2 (duas) abordagens para mensurar a dependência espacial dos preços de imóveis observados: a regressão espacial, através de modelos de defasagem espacial e erro espacial, e a geoestatística com a krigagem ordinária dos resíduos da regressão clássica pelos mínimos quadrados ordinários (MQO) e posterior reincorporação no modelo inicial. Ademais, foi apresentada uma alternativa de utilização da regressão robusta dos erros MQG dada a presença de heterocedasticidade e autocorrelação serial (HAC) pelo procedimento de Newey-West. Ao final, mediram a performance de cada uma das abordagens sobre uma amostra de teste não participante da modelagem inicial escolhida aleatoriamente dos dados iniciais.

As pesquisas apresentadas destacam importantes resultados referentes a modelagens espaciais estatísticas, as quais poderiam servir como arcabouço científico para justificar os valores venais dos imóveis para fins tributários, tanto para base de cálculo do IPTU quanto do ITBI, solucionando os frequentes questionamentos técnicos sobre a acurácia dos valores venais emitidos pelas prefeituras municipais.

Seguindo a mesma linha dos trabalhos pesquisados, o presente trabalho apresenta uma metodologia de cálculo que poderia ser usada em avaliações em massa de imóveis, com base na econometria espacial, para por exemplo, unificar as bases de cálculos dos tributos imobiliários, ITBI e IPTU possibilitando uma melhor eficácia técnica e jurídica de cobrança. Para tanto, sugerimos o aproveitamento dos modelos espaciais baseados na análise da (esperada) correlação espacial dos erros e da própria variável dependente, juntamente com a já tradicional abordagem efetuada nos moldes da regressão linear multivariada, por meio do método dos mínimos quadrados ordinários (MQO).

3 ASPECTOS METODOLÓGICOS

Nessa seção, descreve-se a base de dados e especifica-se os modelos econométricos utilizado para obtenção dos resultados apresentados na Seção 4.

3.1 Base de dados

A base de dados da amostra é composta por 914 (novecentos e quatorze) terrenos dispostos geograficamente por todo o município de Fortaleza, conforme o Apêndice H, extraída do banco de dados de ofertas imobiliárias da Secretaria de Finanças da Prefeitura Municipal de Fortaleza/CE. Estes dados contemplam valores de transações, ofertas e avaliações de ITBI ocorridas nos anos de 2018 e 2019. Captou-se também, nos mesmos moldes, uma amostra de validação, composta de 423 (quatrocentos e vinte e três) terrenos, utilizados para aferir o comportamento dos modelos.

3.2 Fluxograma da pesquisa

As coordenadas geográficas de um imóvel, via de regra, são incorporadas a um modelo de regressão clássica de forma indireta, normalmente por meio de TSA² (superfícies de tendências).

Este trabalho propõe como inovação a incorporação direta das variáveis de localização ao já tradicional modelo clássico de regressão – MQO, aplicando-se técnicas de regressão espacial (SAR e SER), usando preceitos de econometria espacial (matriz de distâncias, defasagem espacial de preços e dos erros).

A pesquisa em questão será conduzida com fundamento no fluxograma da Figura 1, adaptado de Almeida (2012), com o desenvolvimento de três modelos econométricos, um clássico (MQO) e dois espaciais (SAR e SER), cuja sequência de cálculo obedecerá às fases a seguir:

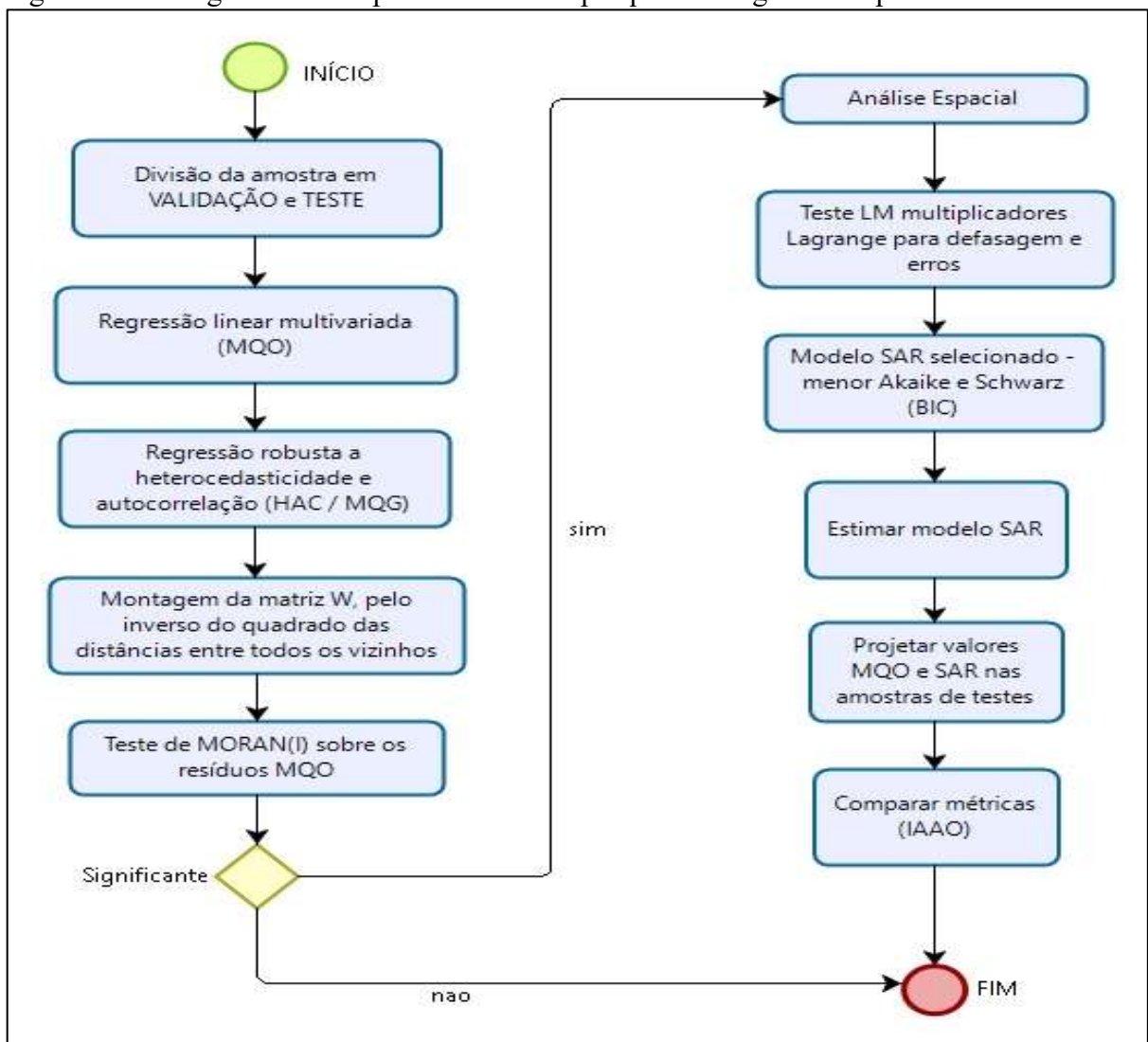
- a) Estimativa do modelo de regressão linear multivariada (MQO);
- b) Calcular a matriz de distâncias (W);
- c) Calcular o índice de Moran (I) do resíduo do modelo MQO para testar se há dependência espacial nos dados;

² Trend Surface Analysis.

- d) Em havendo dependência espacial, calcular o teste LM robusto para escolher entre o modelo de defasagem do preço (SAR – *Spatial autoregressive models*) e o modelo de defasagem dos resíduos (SER – *Spatial autoregressive errors*);
- e) Utilizar parâmetros de acurácia da norma americana (IAAO – *International Association of Assessing Officers*) e as estatísticas de Akaike e Schwarz dos três modelos (MQO, SAR e SER) para justificar a melhor escolha.

A seguir, apresenta-se o fluxograma dos trabalhos, conforme Bandeira *et al.* (2019) adaptado de Anselin e Rey (2014):

Figura 1 – Fluxograma das etapas utilizadas na pesquisa de regressão espacial



Fonte: Adaptado de Anselin e Rey (2014) e Almeida (2012).

Nos itens seguintes, serão apresentados o detalhamento do processo de cálculo a ser aplicado no estudo de caso, de forma tornar mais compreensível todo o trabalho a ser desenvolvido.

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

Considera-se que o principal desafio para qualquer sistema de tributação sobre a propriedade imobiliária é a correta (precisa) avaliação dos imóveis. Por sua vez, a ABNT/NBR 14.653.p1 (2011), em seu item 3.44, preconiza que “*valor de mercado é a quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente.*”.

A norma americana *International Association of Assessing Officers* – IAAO, indo na mesma direção, já inovara ao apresentar também a possibilidade de avaliar imóveis em massa, ao ensinar que o valor de mercado para fins de tributação é geralmente determinado por meio da aplicação de técnicas de avaliação em massa de imóveis, que corresponde ao processo de avaliação de um grupo de propriedades em uma data e com a utilização de dados comuns, métodos padronizados e testes estatísticos.

De modo similar e complementar, Silva e Verdinelli (2000 *apud* SILVA, 2006) discorrem sobre o tema e definem que “*a avaliação em massa consiste na determinação de valores para todos os imóveis situados dentro de um determinado perímetro, pelo emprego de procedimentos avaliatórios, que devem ser respaldados legalmente*”. Essa determinação de valores deve ser fundamentada por uma metodologia que evite ao máximo o emprego de subjetivismo, tanto dos procedimentos quanto das informações ou dados a serem processados, e que procure adequar os mesmos à realidade do mercado imobiliário.

Em geral, os imóveis são bens heterogêneos, ou seja, neles vários elementos mercadológicos interagem entre si para a formação do preço, tais como: área, frente, número de dormitórios, suítes, localização etc. Esses preços são revelados para os agentes econômicos por meio dos preços observados dos produtos diferenciados e as quantidades específicas das características associadas a eles, constituindo nosso banco de dados, com finalidade comparativa.

No caso de imóveis, são considerados como atributos (variáveis comparativas), em geral, características decorrentes dos seus aspectos físicos (área, padrão construtivo, número de vagas na garagem etc.) e de localização (bairro onde se situa o imóvel, distância a polos de

influência etc.), bem como de aspectos econômicos (condições de pagamento do imóvel, natureza do evento: em oferta ou efetivamente vendido etc.). Portanto, há uma infinidade de elementos comparativos viáveis (variáveis) a serem testados, para se inferir com a maior eficiência possível o valor de um imóvel.

Em primeiro lugar, analisa-se a metodologia clássica de regressão linear, por meio do método dos mínimos quadrados ordinários (MQO), que, em linhas gerais, é apresentado da seguinte forma (1):

$$V_V(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (1)$$

onde:

$V_V(Y)$ – Valor Venal (variável dependente);

β_0 – Intercepto (coeficiente linear do modelo);

$\beta_{(1,2,3,\dots,n)}$ – Coeficientes/Parâmetros dos Regressores (variáveis) usadas do modelo;

$X_{(1,2,3,\dots,n)}$ – 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 as seguintes prerrogativas que devem ser obedecidas:

- a) Os resíduos (ε) devem seguir a distribuição normal, possuir média zero, variância constante (homocedasticidade) e serem independentes entre si (não autocorrelacionados);
- b) Não há aleatoriedade das variáveis exógenas independentes (são fixas; somente a endógena dependente é aleatória);
- c) O número de observações deve exceder o número de coeficientes a ser estimado (micronumerosidade);
- d) Não pode existir nenhuma relação linear (correlação) entre as variáveis independentes.

Cardim (2000 *apud* SILVA; HOCHHEIM; SILVA, 2016), reforça a ideia de se observar esses pressupostos, quando ensina que *“na abordagem tradicional lida-se com técnicas bem estabelecidas pela estatística, abordagens como média, variância, desvio padrão, conceitos de normalidade, partindo do princípio de hipóteses e modelos lineares no qual supõe independência entre os dados”*.

Assim, para que as estimativas com base na modelagem estatística, por meio dos modelos MQO possam funcionar todos os pressupostos acima precisam ser testados e

comprovados, senão se corre o risco de trabalhar com modelos viesados (tendenciosos e inconsistentes) nas suas predições, além da possibilidade de cometer *overfitting* ou *underfitting*, quando inadvertidamente, deixa-se apenas para a informática a importante tarefa de revisar os métodos e resultados, principalmente quando se desprezam amostras de validação (teste).

Dessa forma, para a aceitação de uma equação, é necessária a realização de uma análise da normalidade dos resíduos bem como o diagnóstico da existência ou não de heterocedasticidade. Para tanto utilizam-se, entre outros, testes como o de Jarque-Bera, White e de Breusch-Pagan, além da possibilidade de usar a regressão robusta aos erros (MQG) quando tivermos que resolver o problema da heterocedasticidade e/ou autocorrelação serial dos desvios.

A estimação dos parâmetros (β) é realizada pela inferência estatística. Tradicionalmente, esta estimação tem sido realizada pelo Método dos Mínimos Quadrados Ordinário (MQO), que consiste em fornecer uma equação de regressão que minimize a soma de desvios quadráticos entre os valores observados da variável dependente Y , e os valores estimados desta variável dependente $Y\hat{1}$ ³.

Cabe informar que, em virtude da heterogeneidade do mercado imobiliário, a norma brasileira ABNT NBR 14653p2 é mais tolerável que a técnica econométrica tradicional, quando se analisa a significância do teste de *t student*, cuja finalidade é testar a validade dos parâmetros (β) de regressões. Em geral, esse teste é usado para confirmar se a variável que está sendo usada na regressão está realmente contribuindo para a estimativa ou se a mesma é estatisticamente nula.

A econometria tradicional normalmente não considera válidas as variáveis (X) cujos parâmetros (β) apresentam significância superior a 10% ao teste de *t student*, considerando-se um intervalo de confiança de 90% em torno do estimador central. Entretanto, a norma NBR 14653-2 aceita valores de significância (somatório do valor das caudas) de cada variável até 30% e ainda tolera uma maior abertura do intervalo de confiança em torno deste estimador central (média, moda ou mediana), permitindo valores de 80%. Este documento técnico atribui uma graduação para cada valor: até 10% Grau III, até 20% Grau II e até 30% Grau I. (ABNT, 2011).

Já foi apresentado neste trabalho a primeira Lei da Geografia de Waldo Tobler, onde ficou pacificado a importância da variável localização na avaliação de imóveis.

González e Formoso (2000), consideram esta variável como sendo a mais importante a ser testada, pois ela é relacionada com a fixação espacial do produto (imobilidade).

³Para maior aprofundamento, recomenda-se a leitura de “Análise de Modelos de Regressão Linear – com Aplicações” por Charnet e outros (1999).

O valor de localização está relacionado com a acessibilidade (oferta e qualidade de vias e meios de transporte) e com as características da vizinhança, ou seja, do uso do solo no entorno próximo do imóvel. Quando a análise estatística se resume ao modelo MQO, a medição destes efeitos é difícil, pois não são quantificáveis diretamente, sendo medidos indiretamente através de variáveis *proxy*, tais como a renda média da população ou à distância ao centro comercial-histórico da área urbana. Os modelos mais comuns de análise das áreas urbanas consideram apenas um pólo de atração (chamado de *Central Business District*) visando simplificar a análise. Assim, o uso esperado da localização como variável em modelos econométricos deverá ser conduzida na presença de estatística espacial, gerada pelo tratamento deste atributo.

Assim, quando analisamos o comportamento espacial dos imóveis, certamente encontraremos violação aos pressupostos dos modelos lineares MQO, vez que a localização certamente apresentará autocorrelação ou covariância espacial, podendo, também, contaminar os resíduos deste modelo. A brilhante lição de Dantas (2003 *apud* VALDUGA; ZANCAN, 2018), complementa:

A autocorrelação ou dependência espacial pode afetar o termo de erro, a variável dependente ou ambos. A não consideração deste efeito, pode gerar sérios problemas de estimação, pois, na presença de autocorrelação espacial nos resíduos, os parâmetros “ β ” do modelo de regressão linear múltipla são ineficientes e, neste caso, testes de significância como t de *student* e F de *Snedecor*, bem como os intervalos de confiança inferidos, não são mais válidos e as decisões tomadas com base neles são enganosas. Ocorrido isso, o modelo de regressão linear múltipla deverá ser substituído pelos modelos espaciais, estimados por Econometria espacial. (grifo nosso)

Na mesma linha, Trivelloni (2005 *apud* SILVA; HOCHHEIM; SILVA, 2016):

A estatística tradicional falha nesses conceitos, apresentando uma certa continuidade nos dados de acordo com a localidade, a amostra passa a apresentar dependência espacial demonstrando autocorrelação, sendo melhor representada através de modelos espaciais. (grifo nosso)

Portanto, este método tradicional de regressão baseada em estimativas com base no método dos mínimos quadrados ordinários (MQO), aos poucos, foi cedendo o lugar para a prática de avaliações, envolvendo o uso de inteligência artificial e fórmulas matemáticas multivariadas para estabelecer o relacionamento entre as características das propriedades e os preços de venda, permitindo, deste modo, a estimação de valores de mercado para outras propriedades não objetos de vendas recentes, com a consequente inclusão da localização.

Na engenharia de avaliações de imóveis, a inserção da componente espacial nos modelos MQO, no primeiro momento, foi implantada com o uso das TSA (*Trend Surface Analysis*), ou análise por superfície de tendências, mantendo-se os pressupostos de regressão linear e para, por fim, usar as técnicas combinadas de geoestatística e econometria espacial.

Faz-se necessário comentar que procedimentos de geoestatística (*krigagem*) com uso de interpolações não será objeto deste estudo.

Segundo Caixa (2018), cabe pontuar que, em decorrência da heterogeneidade dos imóveis, a avaliação de imóveis recorre usualmente ao modelo de preços hedônicos (DANTAS; MAGALHÃES; VIRGOLINO, 2007 *apud* CAIXA, 2018), que nada mais é do que um subtipo dos modelos MQO tradicionais.

Assim, o modelo hedônico associa o preço dos imóveis com suas características, com o intento de achar a contribuição marginal de cada atributo. A forma funcional mais comum para o modelo hedônico é a log-linear (2) (SHEPPARD, 1997 *apud* CAIXA, 2018):

$$\text{Ln}(V_v) = \beta_0 + \beta_i X_i + \varepsilon \quad i=(1,2,3,\dots,n) \quad (2)$$

3.4 Análise da dependência espacial dos dados

Uma variável é correlacionada com outra quando uma delas está, de alguma forma, relacionada com a outra, ou seja, quando qualquer alteração de gradiente dela repercute nesta (s) outra (s) positivamente ou negativamente. Assim, por exemplo, dizemos que há correlação entre o preço do terreno e sua área territorial, vez que quanto maior for a área, normalmente, mais caro será este terreno. Essa relação pode ser efetuada de diversas formas.

Existem diferentes coeficientes para medir o grau desta correlação, sendo o coeficiente de correlação de Pearson (r), obtido pela divisão da covariância de duas variáveis pelo produto dos seus desvios padrão, um dos mais utilizados, quando estudados relações lineares, como é o caso nesse estudo.

No nosso caso, estudamos as relações entre variáveis, tendo como suas coordenadas geográficas como elemento de comparação. Assim, ao coletar o preço dos imóveis em uma área nobre de uma cidade, percebemos que seus valores tendem a ser muito próximos entre si, ou ainda, os preços dos seus vizinhos influenciam no valor de um imóvel. Dizemos, por fim, que a variável preço possui correlação com ela mesma (autocorrelacionada) em razão da localização geográfica dos imóveis.

Em econometria espacial, investigamos se há autocorrelação espacial entre os dados da amostra, a qual pode ser definida como a coincidência entre valores similares e similaridades locais. Portanto, quando valores altos ou baixos para uma variável aleatória tendem a agrupar-se no espaço, temos o processo de autocorrelação espacial positiva.

No entanto, pode acontecer também de as unidades espaciais serem circundadas por unidades com valores significativamente distintos, ou seja, pode ocorrer que altos valores sejam acompanhados por vizinhos com valores baixos, ou vice-versa, processo que se denomina autocorrelação espacial negativa.

Embora os dois processos sejam igualmente importantes e dignos de consideração, a autocorrelação espacial positiva é, sobretudo, a mais intuitiva, e é encontrada, com maior frequência nos fenômenos econômicos. Na maior parte das vezes, um processo que apresenta autocorrelação espacial negativa é de difícil interpretação.

Uma amostra de dados espacialmente autocorrelacionada contém menos informação do que outra não autocorrelacionada e essa perda de informação precisa ser levada em conta nos testes de estimação e de diagnóstico (essência do problema de autocorrelação espacial em econometria aplicada).

3.4.1 Matriz W de distância (ou contiguidade)

De acordo com Anselin e Rey (1999), para diagnosticar a presença de efeitos de dependência espacial, é necessário definir previamente uma matriz de pesos espaciais (matriz de vizinhança), conhecido como W . Esta pode ser definida de diversas formas, o que traz críticas aos modelos econométricos espaciais (ver item 3.6). Nesse trabalho, optamos por construir uma matriz de distâncias (W) considerando o inverso quadrático da distância entre os pontos vizinhos (d^{-2}), embora alguns autores tradicionalmente optem pelo uso de matrizes de contiguidade espacial.

Assim, cada elemento é igual a distância entre i e j , se i e j forem vizinhos e igual a zero caso i e j não sejam vizinhos. Por convenção, os elementos da diagonal principal da matriz W são iguais a zero, ou seja, $W_{(i=j)}=0$, vez que um elemento não pode ser vizinho de si mesmo.

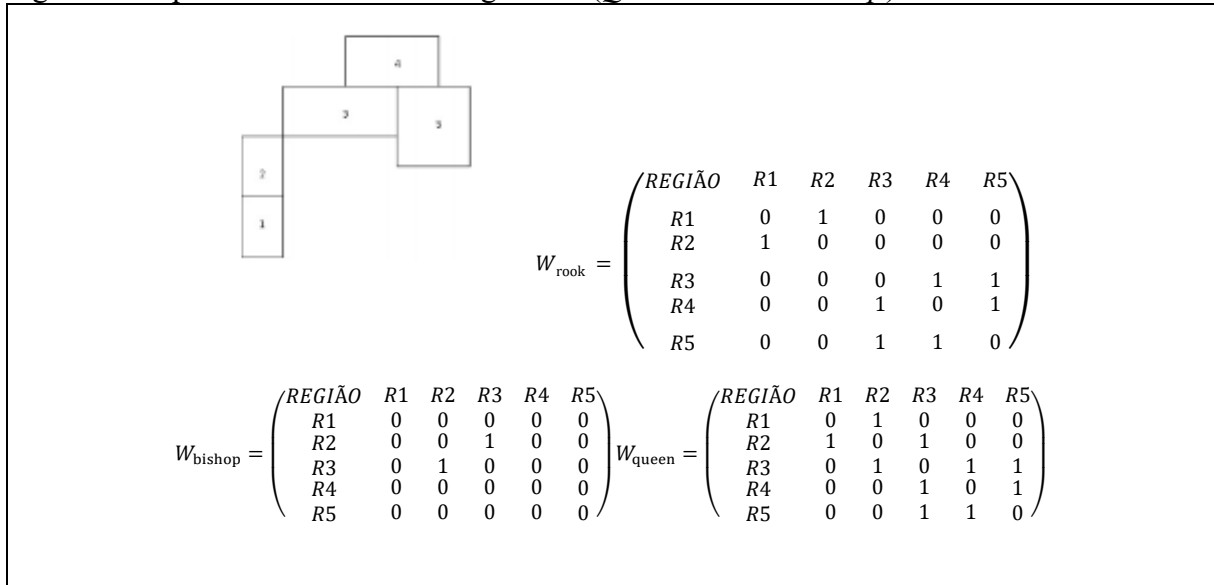
Segundo Anselin (1999a *apud* TRIVELLONI; HOCHHEIM, 2006):

Os elementos que compõem a matriz W são não estocásticos (aleatórios) e são externos ao modelo. Eles estão geralmente baseados nos conceitos de contiguidade dos dados ou na distância entre observações, baseados numa função decrescente com a distância entre observações. Esta estrutura de variação é definida pelo pesquisador diferentemente de outros métodos como os de estimação direta da covariância que calculam a covariância a partir dos próprios dados. Esta falta de obrigatoriedade de uma justificativa teórica na seleção dos pesos espaciais e o fato de serem exógenos ao modelo e sem ambiguidades explica a popularidade dos pesos derivados geograficamente.

Como já adiantado, pode-se também elaborar vários tipos de contiguidade para a matriz de distâncias, afastando-se de técnicas métricas, produzindo uma matriz apenas com

zeros (ausência de vizinhos) e uns (vizinhos). É usual adotar os movimentos das peças do xadrez, para padronizar essa contiguidade, conforme Figura 2, extraída de Chagas (2016):

Figura 2 – Tipos de matrizes de contiguidade (*Queen, Rook e Bishop*)



Fonte: Chagas (2016)

Assim, tem-se:

- Contiguidade do tipo *Rook* (torre) ocorre quando apenas as fronteiras com extensão diferente de zero são consideradas, não se levando em conta os vértices.
- Contiguidade do tipo *Bishop* (bispo) considera vizinhos todas aquelas localidades que possuem apenas um vértice em comum;
- Contiguidade do tipo *Queen* (rainha) quando além das fronteiras com extensão diferente de zero, puderem ser considerados os vértices como contíguos.

A matriz de contiguidade espacial W , com elementos 0 ou 1, é conhecida como matriz não padronizada, em contraposição à matriz W_s normalizada. Para diminuir o trabalho computacional dos cálculos, a matriz de vizinhança (W) deve ser padronizada 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, ou seja, $W_{sij} = \frac{W_{ij}}{\sum w_{ij}}$.

Portanto, a matriz W_s possui todas as linhas com soma igual a 1. A defasagem espacial da variável “y” no ponto “i” pode ser interpretada como sua média ponderada na vizinhança de “i”.

Felizmente, pela própria definição da matriz de contiguidade W , pode-se tratá-la como matriz esparsa; ou seja, a grande maioria dos elementos de W são nulos. Para matrizes

esparsas, existe uma literatura bem desenvolvida sobre algoritmos que tornam o processo computacional mais eficiente. Assim, apesar de a codificação do estimador de máxima verossimilhança não ser trivial (é preciso programar algumas rotinas para matrizes esparsas), o esforço computacional pode ser bastante reduzido.

Cabe repisar, que se pode elaborar diversos tipos de matrizes de vizinhança, como aquelas em que há uma ponderação correspondente ao inverso da distância ou ao inverso do quadrado da distância entre eles, sendo essas as mais comuns, no ambiente de avaliação imobiliária. A escolha da melhor situação é realizada comparando-se os valores com a área de influência de um semivariogramas (amplitude) ou pela otimização na estimação do modelo.

3.4.2 Testes para detectar autocorrelação espacial

Vimos acima, que a autocorrelação ocorre quando analisamos correlações da própria variável estudada, como na chamada autocorrelação espacial do preço, cujo objetivo é calcular a contribuição dos preços dos seus vizinhos na formação do valor de um imóvel avaliando qualquer. Dizemos aqui que, a variável preço é defasada espacialmente.

A autocorrelação apresenta como problema principal, violar as hipóteses de independência dos erros, um dos pressupostos básicos de modelo MQO, vez que embora o estimador MQO gere estimativas não-viesadas, haverá problemas no cálculo dos desvios-padrão dos coeficientes e nos valores de seus testes de hipóteses.

Por exemplo, para atestar essa autocorrelação, a partir da comparação entre os valores de uma amostra e de seus vizinhos, utilizamos indicadores de uma estatística de produtos cruzados do tipo (3) – índice local de Moran:

$$\Pi(d) = \sum_{i=1}^n \sum_{j=1}^n w_{ij}(d) \varepsilon_{ij} \quad (3)$$

Onde, a matriz W_{ij} fornece uma medida de correlação entre as variáveis aleatórias que pode ser o produto destas variáveis, como no caso do índice de Moran para áreas.

Os principais testes para detectar a autocorrelação espacial são Moran I, LM Robusto (erro) e LM Robusto (defasagem).

O índice global de Moran I é a expressão de autocorrelação considerando apenas o primeiro vizinho e consiste em um teste cuja a hipótese nula é de independência espacial, entretanto este teste não identifica o tipo de modelo a usar (erro ou defasagem espacial), e por esse motivo, são realizados testes específicos como o LM robusto (erro) e o LM robusto (defasagem) para detectar os efeitos de autocorrelação serial dos erros e defasagem espacial na

variável dependente respectivamente e ao se comparar esses valores com outros testes estatísticos (Akaike e Schwarz/BIC, por exemplo) chega-se à conclusão sobre qual modelo escolher.

3.4.2.1 Teste de Moran (I)

O teste de Moran I permite determinar se os resíduos do modelo de regressão por mínimos quadrados apresentam autocorrelação espacial. Este índice é uma medida global da autocorrelação espacial, indicando o grau de associação espacial do conjunto de dados.

O valor estatístico é calculado pela seguinte expressão (4):

$$I = \frac{n}{ws} \cdot \frac{\sum_i^n \sum_j^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i^n (y_i - \bar{y})^2} \quad (4)$$

onde:

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

W_{ij} – Elemento da matriz de distância para o par i e j ;

W_s – matriz que representa a ponderação para cada uma das linhas da matriz W ;

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, consiste em um teste cuja hipótese nula é de independência espacial, obtendo-se o valor zero neste caso. Valores positivos (entre 0 e +1) indicam para correlação direta, e negativos (entre 0 e -1) correlação inversa.

De acordo com Des Rosiers *et al.* (2001 *apud* SILVA, 2006):

Assim, a análise de Moran's I se apresenta como um teste paramétrico para testar se, e em que grau, a estrutura espacial observada é aleatória. Como a análise de regressão múltipla requer que as observações sejam independentes uma das outras para que os testes de hipóteses sejam confiáveis, testar a presença de autocorrelação espacial é um pré-requisito para uma boa modelagem hedônica.

Cabe alertar que se deve ter atenção no modelo, pois esse teste é afetado pela ausência de normalidade e pela presença de heterocedasticidade, o que pode invalidar as conclusões inferenciais resultantes das estimações. Entretanto, a utilização do método de Monte Carlo (via *bootstrap*) e a utilização da transformação de Box-Cox na variável dependente podem auxiliar na correção desses problemas, além da tradicional correção por regressão robusta, por meio dos procedimentos de Newey-West (HAC/MQG), utilizados nesse trabalho, por proporcionarem estimativas mais consistentes (não viesadas) que o modelo tradicional MQO.

3.4.2.2 Testes de LM robusto (erro e defasagem)

Segundo Anselin (2005), os testes baseados no Multiplicador de *Lagrange* (testes LM) também são calculados a partir dos resíduos do modelo MQO, mas ao contrário do que ocorre no teste de Moran, aqui se pode saber qual dos modelos espaciais se deve utilizar. Dessa maneira, serão apresentados os testes LM robusto (erro) e o LM robusto (defasagem) para detectar os efeitos de autocorrelação e defasagem espacial nos resíduos MQO e na variável dependente, respectivamente.

No caso do modelo de erros espaciais (SER), os resíduos são modelados na forma $\varepsilon = \lambda W\varepsilon + \xi$ (Equação 9) e, para se testar a hipótese de ausência de autocorrelação espacial, o interesse reside em se testar se a hipótese nula ($H_0=0$), ou ainda se $\lambda = 0$, está presente.

Analogamente, para o modelo de defasagem espacial, também testamos a inexistência de autocorrelação espacial da variável dependente. Assim, em caso de aceitação da hipótese nula ($H_0=0$), tem-se $\rho=0$ (7).

De acordo com Anselin e Rey (1999), o teste LM Robusto (erro) é realizado a partir da estatística representada pela Equação (5) que tem distribuição qui-quadrado (χ^2) com um grau de liberdade, sob a hipótese nula de não existência de autocorrelação espacial no termo do erro. Portanto, a rejeição da hipótese nula (H_0) confirma a autocorrelação espacial do erro, ou seja, deveremos nesse teste encontrar um p -value próximo de zero, para atestar a correlação espacial dos erros (5), encontrando um $\lambda \neq 0$ (9).

$$LM(erro) = \frac{\left[\frac{e'we}{\frac{s^2}{n}} \right]^2}{[tr(w^2+w'w)]} \approx \chi^2 \quad (5)$$

onde:

e: Vetor de resíduos de mínimos quadrados (MQO);

W: Matriz de distâncias (W' é a sua matriz transposta);

$s^2 = \frac{e'e}{n}$: Estimativa da máxima verossimilhança da variância do modelo $Y = X\beta + \varepsilon$;

n: Número de dados da amostra;

tr: Operador denominado traço da matriz.

Do mesmo modo, o teste LM Robusto (defasagem) também é um teste assintótico, que tem distribuição qui-quadrado (χ^2) com um grau de liberdade, sob a hipótese nula de não existência de defasagem espacial na variável dependente, segundo a expressão a seguir (6):

A rejeição da hipótese nula (H_0) confirma a autocorrelação espacial da variável, ou seja, deveremos nesse teste encontrar um ρ -value também próximo de zero, encontrando um $\rho \neq 0$ (7).

$$LM(defasagem) = \frac{\left[\frac{e'wY}{s^2} \right]^2}{\left[(WX\beta)' \frac{(MWX\beta)}{s^2} + tr(w^2 + w'w) \right]} \approx X^2 \quad (6)$$

onde:

e: Vetor de resíduos de mínimos quadrados (MQO);

W: Matriz de distâncias (W' é a sua matriz transposta);

Y: Vetor de observação da variável dependente;

$s^2 = \frac{e'e}{n}$: Estimativa da máxima verossimilhança da variância do modelo $Y = X\beta + \varepsilon$;

X: Vetor de observação das variáveis independentes;

β : Vetor dos parâmetros das variáveis independentes (modelo MQO);

$M = I - X(X'X)^{-1}X'$, onde I é a matriz identidade (matriz identidade é uma matriz diagonal, cujos elementos da diagonal principal são todos iguais a 1 e os demais valores são nulos);

tr: Operador denominado traço da matriz.

E ainda, segundo Trivelloni (2005 *apud* VALDUGA; ZANCAN, 2018):

A dependência espacial pode ser incorporada nos modelos clássicos de regressão de duas formas: em modelos de defasagem espacial, como um regressor adicional na forma de uma variável dependente espacialmente defasada (WY), ou em modelos de erro espacial como uma estrutura espacialmente defasada no erro da regressão ($W\varepsilon$).

3.5 Método de regressão espacial

Segundo Trivelloni e Hochheim (2006), pode-se afirmar que:

A dinâmica urbana produz uma contínua mudança nos efeitos destes fatores sobre o valor dos imóveis: investimentos públicos, empreendimentos econômicos e sociais, atividades comerciais que abrem ou fecham, empreendimentos imobiliários, loteamentos, todos estes fatores produzem mudanças nas características de uma vizinhança e criam novos polos de valorização imobiliária. Quase toda mudança na fisionomia urbana afeta direta ou indiretamente o valor dos imóveis localizados na sua área de influência, podendo transformar o valor de localização no fator mais dinâmico do valor dos imóveis.

Desse modo, explicar tal dinâmica, identificando os fatores que formam os preços dos imóveis, é um dos principais focos da avaliação de bens imóveis. As áreas com características semelhantes tendem a apresentar preços de imóveis também semelhantes. A razão disso é facilmente vista quando se analisa a chamada de primeira Lei da Geografia de

Waldo Tobler (1970), segundo a qual “*todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes*”.

Esse postulado exerce grande influência nas estimativas do valor dos imóveis, pois é de se esperar que os erros nas modelagens não apresentem a aleatoriedade esperada, dada a uniformidade de características (variáveis) entre os imóveis.

Além disso, vimos que a independência e a inexistência de auto correlação residual são pré-requisitos para que um modelo linear MQO funcione adequadamente. O problema é que, na prática, esses pressupostos não serão observados, o que irá inviabilizar a utilização de tais modelos.

Portanto, a influência de vizinhança pode ser explicada na simples escolha de um comprador que tem duas opções de imóveis de mesmo preço e características, intrínsecas e extrínsecas, muito semelhantes entre si, que opta pelo imóvel que tem a vizinhança próxima considerada de maior valor.

Ainda nas palavras de Trivelloni e Hochheim (2006), extrai-se que:

Quando as variáveis de localização não são corretamente especificadas podem gerar outro tipo de problemas nos modelos de regressão. Além da perda de poder de explicação do modelo, o tratamento incompleto ou inadequado dos fatores de localização pode provocar a autocorrelação espacial dos resíduos do modelo de regressão. A presença de autocorrelação espacial nos resíduos invalida uma das hipóteses básicas da inferência estatística e questiona a validade dos testes de hipótese da regressão tornando ineficientes as estimativas e os resultados do modelo.

Segundo Anselin (1998 *apud* CAIXA, 2018),

O reconhecimento de que as técnicas econométricas tradicionais podem falhar diante da presença de autocorrelação espacial, situação comum em dados de corte transversal (*cross section*) possivelmente defasados geograficamente, é uma motivação para o crescente interesse na econometria espacial.

Na avaliação de imóveis, na qual normalmente os dados são coletados segundo um plano de amostragem com coordenadas definidas, torna-se muito importante, quando de sua análise, que a configuração geométrica espacial seja considerada. Com frequência, resultados devem ser apresentados sob a forma de mapas, o que exige uma metodologia específica.

Desse modo, quando se testa a influência da proximidade dos elementos entre si, deve-se medir sua autocorrelação espacial, que por sua vez, é a expressão que mede essa dependência espacial. Assim, Des Rosiers e Thériault (1999 *apud* SILVA, 2006) explicam que a autocorrelação espacial mede o grau de semelhança ou dissimilaridade entre lugares ou pontos como uma função da distância que os separa.

Na literatura de avaliação imobiliária, existem duas grandes escolas para o tratamento espacial dos dados de mercado. Alguns autores têm proposto o uso das técnicas de

econometria espacial e seus modelos de regressão espacial (SAR e SER, por exemplo), enquanto outros têm preferido o uso de técnicas geoestatísticas, como a aplicação de *krigagem ordinária* e elaboração de semivariogramas, para encontrar uma área de contágio espacial – região em que há presença de autocorrelação entre vizinhos, usada, por exemplo, na elaboração da matriz de distâncias da econometria espacial, por exemplo.

Distância entre pontos além do limite dessa região de contágio espacial não implicaria autocorrelação da variável estudada, ou seja, haveria a completa aleatoriedade espacial dessa variável nos pontos, que a rigor, não seriam considerados vizinhos ($W_{ij}=0$), proporcionando um esforço computacional muito menor no cálculo da matriz de distâncias.

Trivelloni e Hochheim (2006) ensinam que a discussão sobre qual das duas escolas tem fundamentos teóricos e metodológicos mais adequados para sua aplicação na área da engenharia de avaliações é estéril, levando a conclusão que se deve combinar os dois estudos. Portanto, a econometria espacial e a geoestatística têm sido aplicadas conjuntamente na análise de dados imobiliários apresentando mais vantagens e potencialidades que desvantagens e limitações. Neste trabalho, adotaremos somente as técnicas de econometria espacial para a análise de nossos dados. Os estudos acerca do comportamento espacial usando *krigagem* dos resíduos MQO e a sua reinserção no modelo ficarão reservados para novos estudos.

3.5.1 Modelo de Defasagem Espacial (SAR)

A dependência espacial criada por meio da interação entre os preços dos imóveis 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 locais, de padrão construtivo ou estruturais, mas também o preço dos imóveis vizinhos. A estimação do modelo apenas por mínimos quadrados ordinários não é adequada, sendo os parâmetros não viesados, apenas se o coeficiente autoregressivo for nulo ($\rho = 0$).

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). Ocorre que esses modelos ainda estão sendo pouco utilizados em relação aos outros modelos econométricos de regressão (a NBR 14.653, de 2011, ainda não o tornou obrigatório) 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 , além de outras críticas técnicas, as quais foram condensadas no item 3.6 deste trabalho.

Segundo Trivelloni e Hochheim (2006), “os primeiros estudos sobre a existência de autocorrelação espacial nos dados do mercado imobiliário foram propostos por Dubin (1988) e o primeiro trabalho de aplicação de modelos de regressão espacial foi proposto por Can (1990 e 1992)”.

No modelo de defasagem espacial, em inglês SAR (*Spatial Auto Regressive* ou *Spatial Lag Model*), a autocorrelação espacial ignorada é atribuída ao comportamento endógeno da variável dependente preço (Y), por meio da seguinte expressão (7):

$$Y = \rho WY + X\beta + \varepsilon \quad (7)$$

Onde Y (variável dependente endógena) é um vetor coluna, contendo “n” observações na amostra para a variável resposta y_i , o coeficiente escalar ρ corresponde ao coeficiente espacial autoregressivo (representando o efeito médio da variável dependente relativo à vizinhança espacial na região em questão). O termo ε corresponde aos erros aleatórios com média zero e variância σ^2 constantes, contendo os resíduos ε_i da equação, X seria o vetor contendo as variáveis independentes exógenas, β os coeficientes de regressão dessas variáveis independentes e W a matriz de vizinhança espacial.

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 lag entre os regressores da equação.

Uma das primeiras sugestões para a estimação do coeficiente é a utilização do estimador de mínimos quadrados ordinários. No entanto, quando o vetor de covariáveis (variáveis do lado direito da equação) é correlacionado com o resíduo da regressão, sabe-se que o estimador de mínimos quadrados ordinários é inconsistente. Esta correlação entre os resíduos e o regressor é observada no modelo em (7). Portanto, estimação via mínimos quadrados ordinários resultaria em uma estimativa inconsistente para o coeficiente (ponderador) “ ρ ”.

Assim, pode-se utilizar estimação via máxima verossimilhança, que não sofre do problema de inconsistência do estimador de mínimos quadrados ordinários, devido à endogeneidade do regressor WY. Em linhas gerais, a estimação via máxima verossimilhança dos parâmetros ρ e σ^2 parte da distribuição normal multivariada para o vetor de resíduos ε .

Uma vez dentro do arcabouço de estimação via máxima verossimilhança, pode-se recorrer a vários dos resultados para este tipo de estimador. Pode-se, então, 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 foi o escolhido nesse trabalho. Testando-se a

significância do parâmetro ρ , se está implicitamente testando a presença de dependência espacial das observações para a variável resposta y_i .

3.5.2 Modelo de Defasagem Espacial dos Resíduos (SER)

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 SER também partem da especificação de modelos AR (autoregressivos) e de modelos MA (médias móveis) para observações no tempo. Note-se que, ao contrário dos modelos SAR, os modelos SER não apresentam a variável resposta como uma função direta dos seus lags espaciais. A autocorrelação espacial nos modelos SER aparece nos termos de erro.

Ainda, de acordo com Dantas, Magalhães e Vergolino (2007), a autocorrelação espacial nos erros está no fato de utilização de 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 do índice de desenvolvimento humano por bairro, sem uma devida suavização destas em uma superfície, podem ocasionar erros de medida pelo efeito de transbordamento (a variável se propaga para além da sua fronteira).

O modelo do erro espacial pode ser expressado formalmente conforme a Equação 8:

$$Y = X\beta + \varepsilon \quad (8)$$

Onde:

$$\varepsilon = \lambda W\varepsilon + \xi \quad (9)$$

$W\varepsilon$ = matriz de erros com efeito espacial;

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

λ = coeficiente autoregressivo.

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.

Segundo Anselin (1988 *apud* DANTAS; MAGALHÃES; VERGOLINO, 2007), 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.

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

3.6 Críticas aos modelos de dependência espacial

Ywata e Albuquerque (2011) mostraram que, apesar do seu uso bastante disseminado, os modelos paramétricos para tratamento de dependência espacial (SAR e SER) vêm recebendo algumas críticas na literatura. Alguns dos pontos levantados pelos críticos e citados pelos autores são importantes para:

- 1) Antecipar aos usuários alguns cuidados e limitações acerca dos quais eles devem estar cientes;
- 2) Fornecer um certo balizamento para pesquisas futuras para os modelos espaciais, de maneira a corrigir/amenizar algumas dessas limitações.

O maior questionamento levantado acerca dos modelos espaciais econométricos se aplica mais fortemente à existência (ou não) de especificação paramétrica para capturar corretamente a dependência espacial.

Vimos que, no caso de testes de hipótese para a presença ou não de dependência espacial, os testes atualmente disponíveis (conforme seção 3.4.2) se comportam de forma bastante satisfatória em situações regulares. Nos casos mais irregulares, como os encontrados na engenharia de avaliações, utiliza-se frequentemente os testes LM Robustos.

De maneira geral, o embasamento teórico para a modelagem em econometria espacial ainda se encontra em um estágio inicial, visto que os primeiros estudos acerca do assunto possuem menos de 60 anos. Dessa forma, uma das dificuldades é encontrar um modelo que responda bem a todos os tipos de situação.

Entre as limitações para os modelos espaciais, podem-se citar:

- 1) Hipótese improvável e desnecessária de normalidade dos resíduos;
- 2) O fato de Y depender dos seus próprios lags espaciais pode implicar que Y também dependa dos lags espaciais do vetor de covariáveis X, incorrendo no

problema de reflexão (*reflexion problem*), apontado por Manski (1993). A consequência prática é que a inclusão de lags espaciais de X pode ocasionar uma matriz com altíssimo grau de multicolinearidade;

- 3) Os modelos SAR, assim como todos os demais modelos ARMA, possuem a prerrogativa de possuir linearidade nos parâmetros (são lineares). Entretanto, na análise espacial pode haver a necessidade de especificações não lineares da relação entre o vetor resposta Y e o vetor dos regressores X ;
- 4) Os modelos SAR não levam em consideração a presença de correlação entre o vetor de regressores X e os resíduos ε , causada pela presença de regressores endógenos em X e/ou pela presença de heteroscedasticidade condicionada a esses regressores. No nosso modelo, corrigimos esse problema utilizando uma regressão robusta aos erros (MQG), por meio do estimador HAC (Newey West), sobre os resíduos do modelo MQO, vez que, por meio dos testes de White e Breusch-Pagan-Godfrey, atestamos a existência desta heterocedasticidade;
- 5) Há fortes críticas à representação excessivamente simplista de toda a dependência espacial em um único coeficiente ρ ou λ ;
- 6) A matriz de contiguidade W implica um alto grau de arbitrariedade na sua especificação, principalmente levando-se em consideração a irregularidade dos mapas de municípios e de setores censitários, ocasionando o fenômeno do transbordamento.

3.7 Medidas de desempenho recomendadas pela IAAO

Como medidas de desempenho e acurácia para comparação entre os diversos métodos, as seguintes métricas foram utilizadas no trabalho:

- **Nível de avaliação** (*sales ratio*) mediano (SR_m):

$$SR_m = \text{mediana de } \frac{\text{valor projetado}}{\text{valor observado}} \quad (10)$$

- **Coefficiente de dispersão (COD):**

$$COD = \frac{100}{SR_m} \times \left(\frac{\sum_{i=1}^n |SR_i - SR_m|}{n} \right) \quad (11)$$

Onde: SR_i é o nível de avaliação de cada terreno individualmente considerado e n é o número total de dados da amostra.

- **Média percentual absoluta do erro (MAPE):**

$$MAPE = \frac{100}{n} \times \sum_{i=1}^n \left| \frac{\text{valor observado}_i - \text{valor projetado}_i}{\text{valor observado}_i} \right| \quad (12)$$

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{valor observado}_i - \text{valor projetado}_i)^2}{n}} \quad (13)$$

- **Coefficiente de determinação (r^2):**

$$r^2 = 1 - \frac{\sum_{i=1}^n (\text{valor observado}_i - \text{valor projetado}_i)^2}{\sum_{i=1}^n (\text{valor observado}_i - \text{valor observado}_{\text{média}})^2} \quad (14)$$

Convém ressaltar que o nível de avaliação e o coeficiente de dispersão são medidas consagradas internacionalmente para aferir performance de avaliações em massa, segundo IAAO – *International Association of Assessing Officers* (2013).

A velocidade com a qual as variâncias e covariâncias aumentam pode ser vista com o Fator de Inflação da Variância (FIV), definido por Gujarati e Porter (2011) como:

$$FIV = \frac{1}{(1-r_j^2)} \quad (15)$$

Onde r_j^2 representa o coeficiente de determinação parcial de X_j em relação as demais variáveis X_j (com $j=1, 2, 3, \dots, k$) e mostra como a variação de um estimador é inflada pela presença da multicolinearidade.

Para os casos em que a multicolinearidade é elevada, ou seja, quando X_j se encontra altamente correlacionado com as demais variáveis X , r_j^2 assume valores próximos a unidade, fazendo com que os resultados de FIV também sejam elevados.

Gujarati e Porter (2011) ainda esclarecem que “quando r^2 aproxima-se de 1, o FIV aproxima-se do infinito. Ou seja, quando a colinearidade aumenta, a variância de um estimador aumenta e no limite, pode se tornar infinita. Caso não haja colinearidade entre as variáveis, o FIV será 1”.

De acordo com Doane e Seward (2008), quando o regressor X_j se encontrar correlacionado com os demais, seu r_j^2 será alto, diminuindo o denominador da Equação (15) e elevando o seu Fator de Inflação da Variância. Este FIV a partir de 2 (dois) é considerado moderado e a partir de 10, forte.

Além disso, Doane e Seward (2008) ensinam que altos valores de FIV podem alargar os intervalos de confiança dos coeficientes da regressão, tornando a estatística t menos

confiável. Convém por oportuno e relevante ressaltar, que um FIV de 10 de um regressor não recomenda sua imediata retirada do modelo, sob pena de má especificação do mesmo (retirar a variável área do terreno em uma modelagem de preços de terrenos, *a priori*, não faz sentido).

A seguir a Tabela 1, adaptada de Doane e Seward (2008), indica a relação entre o aumento do grau de correlação entre as variáveis e o aumento do FIV, ou seja, quanto maior a correlação entre as variáveis dependentes maior será o FIV e o nível dessa correlação:

Tabela 1 – Níveis de multicolinearidade

CORRELAÇÃO	FIV	NÍVEIS
0,40	1,19	FRACA
0,60	1,56	MÉDIA
0,75	2,29	FORTE
0,85	3,60	MUITO FORTE
> 0,85	3,60	FORTÍSSIMA

Fonte: Adaptado de Doane e Seward (2008).

3.8 Estatísticas de Akaike (AIC) e Schwarz (BIC)

A escolha do modelo apropriado, do ponto de vista estatístico, é um tópico extremamente importante na análise de dados. Busca-se o modelo mais parcimonioso, isto é, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Nesta linha, diversos critérios para seleção de modelos são apresentados na literatura (ver: BOZDOGAN, 1987; WOLFINGER, 1993, LITTEL *et al.*, 2002).

Dentre os critérios para seleção de modelos, os critérios baseados no máximo da função de verossimilhança (MFV) são os mais utilizados, com maior ênfase o Teste da Razão de Verossimilhança (TRV), o Critério de Informação de Akaike (AIC) e o Critério Bayesiano de Schwarz (BIC).

O teste da razão de verossimilhança é apropriado para testar dois modelos, desde que um dos modelos seja um caso especial do outro (modelos aninhados). O TRV usa a estatística LR dada por: $LR = 2 \times [L_2 - L_1]$, sendo L_2 máximo do logaritmo natural da função de verossimilhança (MLFV) para o modelo mais parametrizado (M_2) e L_1 para modelo mais simples (M_1).

Se M_1 for o modelo de melhor ajuste ao conjunto de dados LR tem distribuição assintótica $\chi^2_{\delta, v}$, com δ o parâmetro de não centralidade e v graus de liberdade, onde v é a

diferença entre o número de parâmetros dos modelos. A hipótese que o modelo M_1 apresenta melhor ajuste é rejeitada caso $LR > \chi^2_{\delta, v}$.

O Critério de Informação de Akaike (AIC) admite a existência de um modelo “real” que descreve os dados que é desconhecido, e tenta escolher dentre um grupo de modelos avaliados, o que minimiza a divergência de Kullback-Leibler (K-L). O valor de K-L para um modelo f com parâmetros θ , em relação ao modelo “real” representado por g é $L(g, f(\theta)) = \int g(y) \ln \left[\frac{g(y)}{f(y|\theta)} \right] dy$. Esta divergência está relacionada à informação perdida por se usar um modelo aproximado e não o “real”.

A estimativa do AIC para um determinado modelo é dada por: $AIC = -2L + 2K$, em que, L é o MLFV do modelo com os parâmetros θ e K o número de parâmetros. *O modelo com menor valor de AIC é considerado o modelo de melhor ajuste.*

O Critério Bayesiano de Schwarz (BIC) tem como pressuposto a existência de um “modelo verdadeiro” que descreve a relação entre a variável dependente e as diversas variáveis explanatórias entre os diversos modelos sob seleção.

Assim o critério é definido como a estatística que maximiza a probabilidade de se identificar o verdadeiro modelo dentre os avaliados. O valor do critério BIC para um determinado modelo é dado por: $BIC = -2L + 2K \ln(N)$, com N o número de observações. *O modelo com menor BIC é considerado o de melhor ajuste.*

Os três critérios apresentados apesar de conceitualmente diferentes acerca dos modelos em avaliação, utilizam o mesmo estimador estatístico, ou seja, o máximo da função de verossimilhança como medida do ajustamento. Entretanto, definem valores críticos diferentes. Esta é a diferença fundamental entre eles.

Com o teste da razão de verossimilhança, considera-se por hipótese que o modelo mais simples é o de melhor ajuste, até que se observem, dado um nível α de significância, diferenças estatísticas para um modelo mais completo.

Utilizando-se o AIC admite-se que dentre os modelos avaliados nenhum é considerado o que realmente descreve a relação entre a variável dependente e as variáveis explanatórias, ou o “modelo verdadeiro” e então, tenta-se escolher o modelo que minimize a divergência (K-L).

Com o Critério Bayesiano de Schwarz (BIC), está implícito que existe o modelo que descreve a relação entre as variáveis envolvidas e o critério tenta maximizar a probabilidade de escolha do verdadeiro modelo.

4 RESULTADOS

O estudo de caso buscou encontrar o melhor modelo de regressão, usando a tradicional regressão linear multivariada clássica (MQO) e apresentou como inovação, a análise da dependência espacial dos dados, por meio da regressão espacial (SAR e SER), através de ferramentas de econometria espacial.

Para isso, elaborou-se uma amostra de 914 dados de terrenos para elaboração dos modelos e mais 423 dados para validação, distribuídos espacialmente em todo o município de Fortaleza/CE, conforme o Apêndice H.

Seguindo os ensinamentos de Anselin (2005), primeiramente foi feita a implementação do modelo tradicional (regressão linear clássica) para, em seguida, avaliar a presença de autocorrelação espacial e, conseqüentemente, a necessidade de um modelo espacial. Assim, se não houver relação espacial, pode-se retornar ao modelo clássico de regressão com maior segurança do seu ajuste.

Entretanto, em havendo essa dependência (relação) espacial, a literatura a trata de duas formas distintas: pela geoestatística, por meio da análise gráfica do semivariograma (com uso de *krigagem*) ou utilizando econometria espacial, por meio de testes estatísticos específicos, utilizando matrizes de vizinhança. Seguiremos apenas o caminho da econometria espacial.

Utilizando a análise espacial por meio da econometria, de acordo com Anselin e Rey (1999), para diagnosticar a presença de efeitos de dependência espacial, é necessário definir previamente uma matriz de pesos espaciais (matriz de vizinhança), conhecido como W . Essa é uma matriz sintética em que cada elemento é igual a um se i e j forem vizinhos e igual a zero caso i e j não sejam vizinhos. Por convenção, os elementos diagonais da matriz W são iguais a zero, ou seja, $W_{ij}=0$, pois um elemento não será considerado vizinho dele mesmo.

Por fim, em caso de estimativas por modelos espaciais, tem-se que escolher aquele que melhor se enquadra para os dados pesquisados, justificando a escolha, com base em testes estatísticos e parâmetros de acurácia estabelecidos pela norma americana – IAAO, usando a amostra de testes. Para os modelos estudados utilizou-se a estimação da variável endógena preço unitário do terreno (dependente), transformada com base no logaritmo natural e mais 15 (quinze) variáveis independentes – exógenas, dispostas no Anexo A deste trabalho.

Para facilitar a organização dos dados, foi elaborada uma planilha eletrônica, contendo informações referentes ao endereço de cada imóvel (rua e bairro); características físicas que foram consideradas relevantes para a formação do valor do terreno (situação cartográfica, topografia, área total, testada, renda, distância a via principal mais próxima, valor

unitário) e elementos de localização, tais como: coordenadas planas em *Universal Transversa de Mercator* (UTM), IDH de bairro, loteamento em condomínio, vista da beira mar, distância a via principal etc.

Optou-se por trabalhar com as coordenadas em UTM, pois os *softwares* Quantum Gis Versão 2.18.26 (QGIS), GeoDa e Sistema de Avaliação de Bens Versão 1.0 beta (SAB – Espacial) utilizados no presente trabalho, possuem uma *interface* mais amigável para esse tipo de coordenada.

4.1 Análise dos dados por meio do modelo MQO – preços hedônicos

Para elaboração do modelo MQO, foram utilizados os seguintes softwares:

- 1) Action Stat Pro da Empresa EstatCamp;
- 2) SisDEA desenvolvido pela Empresa Pelli Sistemas;
- 3) MsExcel da Microsoft;
- 4) Sistema de Avaliação de Bens Espacial – SAB da Dantas Engenharia; e
- 5) EViews da IHS Inc. (para elaboração do modelo robusto aos erros – MQG).

O modelo estatístico MQO apresentou os seguintes resultados gerais, dispostos na Tabela 2 a seguir:

Tabela 2 – Cálculos estatísticos do modelo MQO

Parâmetros	Regressão MQO
Coeficiente de Correlação (r)	0,836
Coeficiente Determinação (r ²)	0,700
Fisher – Snedecor (F)	139,70
Significância modelo	0,01
Akaike	1.194,38
Schwarz	1.271,76
Log likelihood	-581,189

Fonte: Elaboração do autor

Os cálculos das variáveis independentes, bem como seus respectivos coeficientes e estatística do modelo por regressão linear (MQO) estão dispostos no Apêndice A. Todos os *p-value* foram inferiores a 10% ao teste *t* de *student*, o que indica o aproveitamento de todas as variáveis (todas se tornaram significantes), com grau de fundamentação III, conforme a NBR 14.653.

Analisando os pressupostos básicos dos modelos MQO, percebeu-se que os testes de Breusch-Pagan, White e Koenker-Basset rejeitaram a hipótese nula de homocedasticidade a um nível de significância de 1%, e atestaram também a autocorrelação serial dos resíduos.

Esses resultados já eram esperados, conforme amplos comentários nesse trabalho acerca da heterogeneidade dos preços dos imóveis, o que viola o pressuposto a) do item 3.3.

Portanto, os desvios do modelo MQO são heterocedásticos e autocorrelacionados. Entretanto, é possível ajustar erros-padrão, estatísticas t, F e LM de forma a torná-las válidas mesmo na presença destes problemas, por meio do procedimento de Newey-West, através de uma regressão linear robusta aos erros (MQG).

Os desenvolvimentos relativos aos estimadores de mínimos quadrados generalizados (MQG) são semelhantes aqueles apresentados para a obtenção dos estimadores de MQO. O modelo MQG fornece estimadores (não viesados) mais consistentes que o MQO.

Isso significa que é possível descrever novas estatísticas que funcionam independentemente do tipo de heterocedasticidade presente na população. Esses métodos são os procedimentos robustos em relação à heterocedasticidade, onde a variância dos erros não é constante. Portanto, estimam-se variâncias consistentes, por meio da aplicação de métodos robustos em relação à heterocedasticidade, providência que se tornou trivial, pois muitos programas estatísticos e econométricos calculam essas estatísticas como uma opção (utilizou-se o *Eviews*).

Convém observar que encontramos um *p-value* de 14,12% para a variável LN da área do terreno, que a deixaria não significativa (acima de 10%). Entretanto, a NBR 14.653 p2 (norma brasileira de avaliação de imóveis) permite valores até 30%, além do mais, sua retirada certamente implicaria em problemas de especificação do modelo, dada a inquestionável relevância desta variável. Entretanto, cabe observar que, em razão da sua manutenção, o grau de fundamentação do modelo reduziu para II (no modelo MQO inicial, o grau de fundamentação era III – máximo).

O apêndice A mostra os valores do modelo MQO corrigidos pelo procedimento Newey – West (HAC). Já a equação do modelo de regressão linear multivariado (MQO) está disposta no Apêndice B. A matriz de correlação das variáveis independentes se encontra no Apêndice C, na qual não se observa nenhum valor superior a 80%, valores que podem ser corroborados pelo FIV (fator de inflação da variância), segundo a Tabela 3. Assim, pode-se afirmar que as variáveis são independentes e não apresentam correlação entre si.

Tabela 3 – Fator de inflação da variância das variáveis do modelo MQO

Variável	Fator de Inflação da Variância (FIV)	Interpretação
Loteamento condomínio	1,19	Fraca
Avenida	1,42	Fraca
Esquina	1,16	Fraca
LN renda	1,89	Média
LN área terreno	1,78	Média
LN percentual área preservação	1,10	Fraca
Infraestrutura ajustada	1,35	Média
Índice aproveitamento básico	1,61	Média
Dens. Comerc. Trecho logradouro	1,24	Média
LN Valor m ² face quadra IPTU 2014	3,04	Forte
Transação	1,22	Média
Oferta	1,46	Média
LN Distância Via Principal	1,46	Média
LN Interação incorporação 200	1,61	Média
LN IDH Educação	2,43	Forte

Fonte: Elaboração do autor

Um dos mais importantes pressupostos dos modelos de regressão linear, versa sobre a exigência de normalidade nos resíduos (consultar item 3.3). Entretanto, pode-se garantir a normalidade desses resíduos, com base no Teorema do Limite Central (TLC), que, em vias gerais, indica que nas amostras aleatórias simples com qualquer distribuição, com média μ e variância σ^2 , a distribuição amostral da média aproxima-se de uma distribuição normal, com média μ e variância $\frac{\sigma^2}{n}$ à medida que o tamanho da amostra aumenta.

O teorema do limite central é um dos principais resultados da estatística. Este teorema diz que se uma amostra é grande o suficiente (acima de 30 dados. No nosso caso, temos 914 dados), qualquer que seja a distribuição da média amostral, a distribuição será aproximadamente normal.

Portanto, garante-se que os resíduos da regressão MQO seguem uma distribuição normal, por meio do teorema do limite central.

Os gráficos de análise dos resíduos do modelo de regressão, bem como o de dispersão do preço, na transformação logarítmica (modelo de preços hedônicos) e natural estão no Apêndice D.

4.2 Análise dos dados por meio de modelos econométricos espaciais

4.2.1 Análise da dependência espacial – índice de Moran (I) e multiplicadores de Lagrange (testes LM)

O índice de Moran (I) é comumente utilizado na análise de dados como um dos mais importantes testes de dependência espacial. Aplicou-se esse teste sobre os resíduos MQO, que, conforme supracitado nos pressupostos, deveria apresentar inexistência de autocorrelação espacial, para que este modelo MQO pudesse ser usado sem receios de interpretações equivocadas.

A forma gráfica do índice apresenta quatro situações, dispostas em cada um dos quatro quadrantes. Os quadrantes AA (alto-alto) e BB (baixo-baixo) indicam pontos com valores semelhantes, de associação positiva (AA) ou negativa (BB), representando uma dependência espacial. Já os quadrantes BA (baixo-alto) e AB (alto-baixo) indicam associação negativa (ou contrária), no sentido de que uma localização possui vizinhos com valores distintos, portanto, uma dispersão sem dependência espacial.

Analisando apenas o gráfico de Moran dos resíduos MQO do modelo observou-se que há dependência espacial positiva (coeficiente angular da reta é positivo), segundo Apêndice E. Cabe observar, porém, que o índice de Moran I só indica a presença de autocorrelação espacial, mas não indica qual dos modelos espaciais deve ser utilizado.

4.2.2 Dimensionamento da matriz de distâncias (W)

Utilizaremos uma matriz de distância W, com base no inverso do quadrado da distância entre pontos vizinhos. Por convenção, os elementos diagonais da matriz W são iguais a zero, ou seja, $W_{(i=j)}=0$, vez que um elemento não pode ser vizinho de si mesmo.

O primeiro desafio é saber se os pontos A e B são vizinhos ou não. Para isso, usaremos o valor de *minimo_threshold_distance* fornecido pelo GeoDa. Este valor indica que a distância mínima em que podemos garantir que existe pelo menos um par de pontos vizinhos. Menores valores que esse valor mínimo garantiria a existência de mais pontos vizinhos; enquanto que, maiores distâncias não produziram pontos vizinhos.

Como a amostra possui 914 dados, a matriz W foi gerada com 914 linhas e 914 colunas, ou seja, do tipo 914 x 914 (835.396 elementos), segundo a Figura 3, onde W_{12} seria o

inverso do quadrado da distância entre os pontos 1 e 2, W_{13} seria o inverso do quadrado da distância entre os pontos 1 e 3 e assim sucessivamente.

Figura 3 – Forma da matriz de distâncias W (914 x 914)

$$= \begin{pmatrix} 0 & w_{1\ 2} & w_{1\ 3} & & w_{1\ 913} & w_{1\ 914} \\ w_{2\ 1} & 0 & w_{2\ 3} & & w_{2\ 913} & w_{2\ 914} \\ w_{3\ 1} & w_{3\ 2} & 0 & & w_{3\ 913} & w_{3\ 914} \\ w_{4\ 1} & w_{4\ 2} & w_{4\ 3} & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & 0 & w_{913\ 914} \\ w_{914\ 1} & w_{914\ 2} & w_{914\ 3} & & w_{914\ 913} & 0 \end{pmatrix}$$

Fonte: Elaboração do autor

A seguir, foi feita a padronização de cada linha, na qual a matriz assumiu a nomenclatura W_s . Neste caso, cada elemento de W_s , representado por W_{sij} foi obtido dividindo-se W_{ij} pela soma dos elementos da linha “i” a qual pertence, ou seja $W_{sij} = \frac{W_{ij}}{\sum w_{ij}}$.

Nessa matriz W_s , os elementos das linhas somam 1. Assim, a defasagem espacial da variável “y” no ponto “i” pode ser interpretada como sua média ponderada na vizinhança de “i”.

Segundo cálculo do GeoDa, a distância mínima para que cada dado tivesse pelo menos 1 (um) vizinho foi de aproximadamente 1.550m (*min_threshold_distance*). Portanto, para distâncias inferiores a 1.550m certamente mais de 1 par de vizinhos serão encontrados. Analogamente, distâncias maiores indicaria inexistência de vizinhança, alocando o valor zero para a interação entre esses pontos, na construção da matriz de distancias W.

É necessário esclarecer que, foram testadas outras matrizes de distâncias e realizados os correspondentes testes de correlação espacial. O objetivo foi encontrar a maior distância a ser usada na matriz, para que se possa garantir a existência pelo menos dois pontos (1 par) vizinhos.

A título de exemplo, o Apêndice F mostra a área de contágio do ponto 852 e seus 13 vizinhos calculados pelo GeoDa, bem com as os elementos da matriz W padronizada de distâncias para cada par de vizinhos.

4.2.3 Escolha do modelo da dependência espacial, por meio dos multiplicadores de Lagrange (testes LM)

Após o cálculo da matriz de distância (W) e com a confirmação da dependência espacial, por meio do teste de Moran (I), utilizamos os multiplicadores de *Lagrange* Robusto do Erro (LM_λ) e Multiplicador de *Lagrange* Robusto da Defasagem (LM_ρ), para definir qual modelo de dependência espacial terá melhor comportamento no caso em estudo, cujos resultados se encontram na Tabela 7.

Observa-se que ambos foram significantes (apresentaram-se significativos ao nível de 1%), ou seja, os dois modelos espaciais (defasagem e erro) poderiam servir como escolha. Entretanto, o modelo de defasagem espacial do preço apresentou menor Akaike, Schwarz (BIC) e maior coeficiente de determinação (r^2) e Ln likelihood (logaritmo natural da função maxiverossimilhança), sendo assim, este foi escolhido para a projeção final sobre os dados da amostra de teste, conforme Tabela 4.

Cabe ressaltar que não calculamos o modelo com base na correlação espacial dos resíduos (SER), em razão deste ter sido preterido pelo modelo de defasagem espacial.

Tabela 4 – Resultados estatísticos do modelo de defasagem espacial e erro espacial

MODELO	LM_ρ Robusto (Defasagem espacial)	LM_λ Robusto (Erro espacial)
Valor	22,896	10,80
<i>p-value</i>	0,0000	0,00102
Akaike	1,161,44	1.176,52
Coef. Determinação (r^2)	0,712459	0,707753
Coef. Correlação (r)	0,844	0,821
Schwarz	1.243,34	1.253,61
Ln likelihood	-563,72	-572,26

Fonte: Elaboração do autor

No modelo de defasagem espacial, a autocorrelação espacial ignorada é atribuída à variável endógena dependente, conforme expressa pela Equação 16 a seguir.

$$Y = \rho WY + X\beta + \varepsilon \quad (16)$$

Diante dessa possibilidade de autocorrelação espacial, ou efeito vizinhança, em seguida estimou-se o modelo de defasagem espacial de preços hedônicos. Os resultados da estimação estão explicitados no Apêndice A.

O Coeficiente de Correlação de 0,844 indicou uma correlação fortíssima da variável dependente com as variáveis independentes utilizadas no modelo. Já o Coeficiente de

Determinação de 0,7124 indicou que 71,24% da variação da variável dependente (preço) foi explicada pelas variáveis independentes exógenas utilizadas.

Observou-se ainda que todas as variáveis independentes se apresentaram significativas ao nível de 10% (o máximo calculado foi de 9,67% para IDHEduc), enquadrando-se no grau III de fundamentação conforme NBR 14.653-2:2011.

Além disso, destaca-se que o ρ (Coeficiente Autoregressivo) foi positivo e com um valor da estatística z de 5,809 (significante a menos de 1%), implicando a existência de um forte efeito de defasagem espacial positiva na amostra. Isto é, neste caso, a formação dos preços de mercado não foi explicada apenas pelas suas características estruturais e locacionais tradicionalmente consideradas, mas também dependeu dos preços dos terrenos vizinhos, isto é, foram influenciados positivamente pelos vizinhos.

Ademais, verificou-se que o coeficiente da variável área do terreno no modelo MQO, que era acima de 10% (14,12%), o que enquadrava aquele modelo no grau II de fundamentação da norma, reduziu para 1,15%, garantindo uma melhor eficiência na estimativa, enquadrando o modelo espacial (SAR) no grau máximo (III) da norma.

Observou-se também que as alterações mais significativas nas magnitudes dos coeficientes ocorreram no intercepto e nas variáveis relacionadas à localização, que tiveram seus valores reduzidos.

Outro aspecto importante é que os valores obtidos nas estatísticas de Akaike e Schwarz no modelo espacial (SAR) foram bem inferiores aos encontrados no modelo tradicional (MQO), o que mostrou a melhoria no modelo ajustado pela análise espacial, segundo tabelas 2 e 7.

Pelos resultados obtidos, verificou-se que a consideração da questão espacial, com base em distâncias a polos de influência, ou dividindo o espaço em regiões ou zonas, não é capaz de explicar completamente as variações dos preços em relação à localização do terreno. Isto porque não são apenas os polos que afetam os preços dos terrenos, pois existe uma verdadeira interação entre os dados.

Assim, a solução para estimação segura da equação de preços hedônicos, na presença de efeitos espaciais, deve ser feita com base na econometria espacial, em que a variável endógena de defasagem espacial $W.LnY$ (valor unitário) deve contemplar todas as interações espaciais possíveis entre os dados, servindo como *proxy* para variáveis locacionais não consideradas no modelo MQO.

4.2.4 Análise de desempenho dos modelos clássicos (MQO) e espacial por defasagem, pelos critérios da IAAO (*International Association of Assessing Officers*)

Nesse tópico, analisou-se a acurácia (grau de eficiência) dos modelos, com base na norma americana (IAAO), sobre a amostra de testes, por meio dos seguintes balizadores:

- 1) COD – Coeficiente de dispersão da mediana;
- 2) MAPE (média percentual absoluta do erro);
- 3) RMSE (*Root Mean Squared Error*);
- 4) MAD/MAE (Desvio Padrão Absoluto da Média – *Mean Absolute Deviation*);
- 5) Desvio Padrão Quadrático da Média (MSD/MSE – *Mean Squared Deviation*);
- 6) Coeficiente de determinação r^2 (variável direta);
- 7) Coeficiente de determinação ajustado r_a^2 (variável direta);
- 8) Nível de avaliação (SR – *Sales Ratio*).

Analisando os diversos resultados extraídos das respectivas modelagens, verificou-se que o modelo SAR possui um poder de explicação melhor que o modelo clássico (MQO).

Em relação aos indicadores acima, o modelo que possui menor variância (dispersão) dos desvios deve apresentar os melhores resultados (exceto os três últimos indicadores), sendo, assim, a melhor escolha.

Cabe ressaltar que, o coeficiente de dispersão da mediana (COD) ainda precisaria de melhores ajustes, vez que são recomendados valores menores que 20% para terrenos e nos nossos cálculos encontramos valores próximos a 38%.

As exceções são observadas nos coeficientes de determinação (simples e ajustado), pois medem o grau em que modelo responde à variabilidade do preço, tendo em vista os dados coletados. Assim, quanto maior forem esses valores, melhor será esse ajuste.

Como se atestou por meio da Tabela 5, o modelo espacial SAR respondeu melhor a todos os medidores de acurácia.

Tabela 5 – Medidores de acurácia da IAAO

ESTIMADOR	MODELO CLÁSSICO (MQO)	MODELO ESPACIAL (SAR)
MAPE (média percentual absoluta do erro)	37,47	37,37%
RMSE (<i>Root Mean Squared Error</i>)	501,08	497,67
MAD/MAE (Desvio Padrão Absoluto da Média – <i>Mean Absolute Deviation</i>)	288,92	284,63
COD (Coeficiente de dispersão da mediana)	37,49%	37,43%

Continua

Tabela 5 – Medidores de acurácia da IAAO

ESTIMADOR	MODELO CLÁSSICO (MQO)	MODELO ESPACIAL (SAR)
Desvio Padrão Quadrático da Média (MSD/MSE – <i>Mean Squared Deviation</i>)	251.081,69	247.679,13
Coefficiente de determinação r^2 (variável direta)	0,6334	0,6394
Coefficiente de determinação ajustado r_a^2 (variável direta)	0,6330	0,639
Nível de avaliação (SR – <i>Sales Ratio</i>)	0,9995	0,9983
Log Likelihood	-581,189	-563,72

Fonte: Elaboração do autor

Como escolha final, percebeu-se que o modelo espacial é ligeiramente mais eficiente que o modelo clássico de regressão – MQO para os dados de teste, segundo os critérios da IAAO.

4.2.5 Distribuição dos resíduos da regressão espacial - SAR

Após a escolha do modelo espacial, elaborou-se um mapa com a distribuição espacial no município, dos resíduos desta regressão, disposto no Apêndice G.

A finalidade do mapa reside na percepção que a inclusão da modelagem espacial pulverizou a distribuição dos resíduos, que antes era composto com arranjos de *clusters* tendenciosos, que apontava regiões com desvios positivos e outras com desvios negativos.

5 CONCLUSÃO

O trabalho envolveu a realização de um estudo de caso que comparou os modelos clássico e de regressão espacial nos terrenos do município de Fortaleza/CE, cujos resultados levaram a conclusão pela opção do modelo de regressão espacial (SAR).

No presente caso, verificou-se a existência de autocorrelação espacial nos resíduos e preços do modelo de regressão por mínimos quadrados para as variáveis formadoras de valor. Além disso, o uso de modelo de regressão linear clássico (MQO), apesar de apresentar resultados semelhantes para os valores unitários estimados dos imóveis da amostra, resultou numa modelagem equivocada do mercado imobiliário local. Isto se explica porque este modelo não levou em conta a influência da vizinhança, o que ocasionou erros de especificação.

Para se atestar essas informações, primeiro obteve-se uma amostra que contemplou toda a extensão territorial. Posteriormente, foi realizada a escolha das variáveis mais apropriadas, com base em testes estatísticos e se estimou o modelo de regressão com variáveis múltiplas – MQO.

Verificou-se então que os modelos MQO, quando voltados à análise de preços de imóveis, sofrem restrições graves nos seus pressupostos essenciais, vez que é esperada a autocorrelação espacial dos erros e da própria variável endógena dependente – preço (endogeneidade), além da variância não constante (heterocedasticidade) desses resíduos.

Assim, como inovação, esse trabalho procurou sanar esses problemas, primeiro usando estimadores de NEWEY-WEST (HAC/MQG), depois utilizando um modelo puramente espacial, para atestar a correlação espacial dos resíduos e da variável dependente, fato que se evidenciou quando se aplicou o teste de Moran (I), após a construção da matriz de distâncias W .

Essa matriz foi calculada com base no inverso do quadrado da distância entre os pontos (d^{-2}), após se testar outras hipóteses de distâncias. Verificou-se também que a utilização de todos os dados para a construção da matriz de vizinhança acarretou em um desnecessário tempo de cálculo e grande esforço computacional, mesmo utilizando-se de bibliotecas já disponíveis nas linguagens R e Python.

Nesse sentido, adotou-se uma distância mínima de 1.550m, que correspondeu a garantia de existência de pelos menos 1 vizinho (1 par), ao se utilizar o *software* GeoDa. Portanto, as distâncias acima de 1.550 m não apresentaram vizinhos, resultando um valor nulo na matriz W .

Por fim, observando o teste de LM robusto da defasagem espacial e dos erros, bem como os balizadores da norma americana (IAAO), concluiu-se que a metodologia espacial (SAR), com enfoque na variável dependente preço, tornou-se a melhor escolha. Cabe ressaltar que o modelo baseado na correlação espacial dos resíduos (SER) não foi calculado.

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.

Ademais, o método proposto também poderia contribuir para a justiça fiscal por meio de sua aplicação na elaboração de planta de valores genéricos em municípios de médio e grande porte, além de possibilitar a uniformidade das bases de cálculo do IPTU e ITBI, resolvendo definitivamente um dos maiores gargalos tributários que os municípios brasileiros enfrentam.

REFERÊNCIAS

- ABNT. **NBR 14653-1**: avaliação de bens. Rio de Janeiro: [s.n.], 2019.
- _____. **NBR 14653-2**: avaliação de bens parte 2: imóveis urbanos. Rio de Janeiro: [s.n.], 2011.
- ACTION Stat Pro. Empresa EstatCamp. Versão 3.6.331.450, 2019.
- 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.
- ANSELIN, Luc. **Exploring Spatial Data with GeoDa™: A Workbook**. Department of Geography, University of Illinois, 2005. 244p.
- BANDEIRA, Sandro R. V.; OLIVEIRA, Antônio Augusto F. de; PRADO, João Freire; SOUZA, Heverton Alves de. Acurácia das Predições de Avaliação em Massa de Terrenos Urbanos por Modelos Espaciais. *In*: CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS - COBREAP, 20., 2019, Salvador. **Anais...** Salvador: IBAPE, 2019.
- BOZDONGAN, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v. 52, n. 3, p. 345-370, Sep. 1987.
- CAMARGO, E. **Desenvolvimento, Implementação e Teste de Procedimentos Geoestatísticos (Krigagem) no Sistema de Processamento de Informações Georreferenciadas (Spring)**. 1997. 148 f. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.
- CAMARGO, E. C. G.; FUCKS, S. D.; CÂMARA, G. Análise Espacial de Superfícies. *In*: INPE. **Análise Espacial de Dados Geográficos**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais – INPE, 2002. Cap. 3.
- CHAGAS, André Luís Squarize. **Econometria Espacial**. Aula 2: Matrizes de pesos espaciais Disponível em:
<https://edisciplinas.usp.br/pluginfile.php/990610/mod_resource/content/0/Aula%202_2016.pdf>. Acesso em: 28 out. 2019.
- DALL´AGNOL, Rodrigo Werner; BETZEK, Nelson Miguel; BAZZI, Cláudio Leones. Aplicação do Software R para Análise Geoestatística, Interpolação de Dados por Krigagem Ordinária e Geração de Mapas Temáticos de Produtividade Agrícola. **Revista de Computação Aplicada ao Agronegócio**, v. 1, n. 1, 2018.
- DANTAS, R. A. **Modelos espaciais aplicados ao mercado habitacional um estudo de caso para a cidade do Recife**. 2003. 114 f. Tese (Doutorado em Economia) – Centro de Ciências Sociais e Aplicadas, Universidade Federal de Pernambuco, Recife, 2003.

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.

DANTAS, Rubens Alves. **Relatório final das atividades PGVI**. Fortaleza, 2014.

DOANE, David P.; SEWARD, Lori E. **Estatística Aplicada à Administração e à Economia**. McGraw Hill Brasil, 2008. 864p.

EVIIEWS 10. IHS Markit Inc. 2017.

FERMO, Graziela Olivo; HOCHHEIM, Norberto. Influência de Erros de Medida de Posicionamento em Modelos de Regressão Espacial. *In*: CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS - COBREAP, 17., 2013, Florianópolis. **Anais...** Florianópolis: IBAPE, 2013.

FLORENCIO, Lutemberg de Araújo. Modelos Espaciais para Avaliações de Imóveis em Massa com Base em Modelos de Regressão GAMLSS. *In*: CONFERÊNCIA INTERNACIONAL DA LARES - LATIN AMERICAN REAL ESTATE SOCIETY, 12., 2012, São Paulo. **Anais...** São Paulo: LARES, 2012.

GEOUDA. Universidade de Chicago – IL. Versão 1.14.0. 2019.

GONZÁLEZ, Marco Aurélio Stumpf; FORMOSO, Carlos Torres. Análise conceitual das dificuldades na determinação de modelos de formação de preços através de análise de regressão. **Revista Engenharia Civil**, n. 8, 2000.

GUERRA, P. A. G. **Geoestatística operacional**. Rio de Janeiro: Departamento Nacional da Produção Mineral, 1988.

GUIMARÃES, Ednaldo Carvalho. **Geoestatística Básica e Aplicada**. Universidade Federal de Uberlândia – MG, v. 1, 2004, p. 78 (Apostila).

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria Básica**. Mc Graw Hill, 2011.

INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS - IAAO. Standard on Ratio Studies, 2013. www.iaao.org.

INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS - IAAO. Standard on Mass Appraisal of Real Property. 2017.

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS System for Mixed Models**. Cary: Statistical Analysis System Institute, 2002. 633p.

MANSKI, C. Identification of endogenous social effects: the reflection problem. **Rev. Econ. Stud.**, Oxford, v. 60, n. 3, p. 531-542, 1993.

MATHERON, G. **Les variables régionalisées et leur estimation**: une application de la théorie des fonctions aléatoires aux sciences de la nature. Paris: Masson et Cie, 1965.

MICROSOFT Excel. Microsoft Corporation. 2016.

OLMO, Jorge Chica. Spatial Estimation of Housing Prices and Locational Rents. **Urban Studies**, v. 32, n. 8, 1995.

PELLI NETO, Antônio *et al.* **Coletânea de Artigos de Avaliação de Imóveis CAIXA**. Brasília, 2018. 248p.

QGIS. Quantum Gis, Versão 2.18.26.

R (Software), Version 3.6.1.

SAB – ESPACIAL. Sistema de Avaliação de Bens, Dantas Engenharia, Versão 1.0.

SALVIANO, A. A. C. **Variabilidade de atributos de solo e de Crotalaria juncea L. em solo degradado do município de Piracicaba-SP**. 1996. 91 f. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1996.

SILVA, A. P. **Variabilidade espacial de atributos físicos do solo**. 1988. 104 f. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1988.

SILVA, Caroline Bernardo; HOCHHEIM, Norberto; SILVA, Everton. Estudo da Variável Localização em Modelos de Regressão Linear, Espacial e Geoestatística em Avaliação de Imóveis. *In*: CONGRESSO DE CADASTRO MULTIFINALITÁRIO E GESTÃO TERRITORIAL – COBRAC, 14., 2016, Florianópolis-SC. **Anais...** Florianópolis: UFSC, 2016.

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.

SISDEA. Pelli Sistemas Engenharia. Versão Home 1.50.

TRIVELLONI, C. A. P. **Método para determinação do valor da localização com uso de técnicas inferências e geoestatísticas na avaliação em massa de imóveis**. 2005. 171 f. Tese (Doutorado em Engenharia Civil) - Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2005.

TRIVELLONI, Carlos A. Peruzzo; HOCHHEIM, Norberto B. O Valor de Localização dos Imóveis: Determinação por Métodos de Análise Espacial. *In*: CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS - COBREAP, 12., 2006, Fortaleza. **Anais...** Fortaleza: IBAPE, 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

WOLFINGER, R. D. Covariance structure selection in general mixed models. **Communications in Statistics**, v. 22, p. 1079-1106, 1993.

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

APÊNDICES

APÊNDICE A – TABELA 6

Tabela 6 – Cálculos estatísticos dos modelos MQO, MQG (HAC) e SAR

MODELOS: Explicativas	LN Preço Unitário	LN Preço Unitário	LN Preço Unitário
Intercepto	4,669[19,93]	4,669[14,97]	2,972[8,05]
Loteamento condomínio	0,607[6,29]	0,607[10,97]	0,526[5,54]
Avenida	0,219[4,88]	0,219[4,48]	0,219[5,02]
Esquina	0,096[2,58]	0,096[2,47]	0,096[2,67]
LN renda	0,166[5,39]	0,166[4,71]	0,1222[3,94]
LN área terreno	-0,025[1,95]	-0,025[-1,47]	-0,036[-2,53]
LN percentual área preservação	-0,101[-4,55]	-0,101[-3,08]	-0,101[-4,70]
Infraestrutura ajustada	0,068[3,73]	0,068[3,24]	0,066[3,76]
Índice aproveitamento básico	0,089[3,17]	0,089[2,81]	0,075[2,72]
Dens. Comerc. Trecho logradouro	0,118[1,99]	0,118[1,92]	0,148[2,58]
LN Valor m ² face quadra IPTU 2014	0,370[12,99]	0,370[10,13]	0,293[9,52]
Transação	0,253[3,29]	0,253[3,32]	0,220[2,95]
Oferta	0,298[7,10]	0,298[8,21]	0,288[7,04]
LN Distância Via Principal	-0,04[-4,84]	-0,04[-4,32]	-0,046[-4,70]
LN Interação incorporação 200	0,009[2,34]	0,009[2,04]	0,009[3,32]
LN IDH Educação	1,820[4,72]	1,820[3,82]	0,674[1,66]
W_LN Valor Unitário (ρ)			0,281[5,81]
MÉTODO	MQO	MQG(HAC)	SAR
N	914	914	914
r²	0,708	0,708	0,712
Schwarz	1.271,76	1.271,76	1.234,34
Akaike	1.194,38	1.194,38	1.161,44

Fonte: Elaboração do autor.

Notas: a) Estatística T entre colchetes; b) Prefixo “LN” antes da variável indica a transformação com logaritmo natural; c) Variável dependente foi transformada para logaritmo natural; d) Todos os estimadores foram estatisticamente significantes a 1%.

APÊNDICE B – QUADRO 1

Quadro 1 – Equações dos modelos utilizados no trabalho

MODELO	EQUAÇÃO
MQO – Mínimos quadrados ordinários e HAC (MQG) – Mínimos quadrados generalizados.	$\text{Valor unitário} = e^{(4,669021816 + 0,607019033 * \text{loteamento condomínio} + 0,2197088959 * \text{avenida} + 0,096106115 * \text{esquina} + 0,1667232193 * \ln(\text{renda}) - 0,02897035501 * \ln(\text{área terreno}) - 0,1011664287 * \ln(\text{percentual área preservação}) + 0,06810982912 * \text{infraestrutura ajustada} + 0,08933436548 * \text{índice aprov básico} + 0,1180454434 * \text{densidade comercialização trecho} + 0,3706592342 * \ln(\text{valor m2 terreno face quadra i}) + 0,009913395181 * \ln(\text{interacao incorporacao 200}) + 0,2535008533 * \text{transação} + 0,2987057145 * \text{oferta} - 0,0492071645 * \ln(\text{Dist.Via Principal}) + 1,820558767 * \ln(\text{IDH Educação}))}$
SAR – Modelo de espacial de defasagem do preço unitário.	$\text{Valor unitário} = e^{(2,97247 + 0,526788 * \text{loteamento condomínio} + 0,219243 * \text{avenida} + 0,122292 * \ln(\text{renda}) - 0,0362822 * \ln(\text{área terreno}) - 0,101282 * \ln(\text{percentual área preservação}) + 0,0666228 * \text{infraestrutura ajustada} + 0,0751341 * \text{índice aprov básico} + 0,14846 * \text{densidade comercialização trecho} + 0,293853 * \ln(\text{valor m2 terreno face quadra IPTU}) + 0,00950948 * \ln(\text{interação incorporação 200}) + 0,220936 * \text{transação} + 0,28823 * \text{oferta} - 0,0466099 * \ln(\text{Dist.Via Principal}) + 0,674587 * \ln(\text{IDHEduc}) + 0,0964153 * \text{Esquina} + 0,281392 * W * \ln(\text{valor unitário}))}$

Fonte: Elaboração do autor.

APÊNDICE C – TABELA 7

Tabela 7 – Matriz de correlação entre variáveis do modelo MQO

Variáveis	Loteamento condomínio	Avenida	Esquina	LN renda	LN área terreno	LN percentual área preservação	Infraestrutura ajustada	Índice aproveitamento básico	Dens. Comerc. Trecho logradouro	LN Valor m2 face quadra IPTU 2014	Transação	Oferta	LN Distância Via Principal	LN Interação incorporação 200	LN IDH Educação	Valor Unitário
Loteamento condomínio	100,00%															
Avenida	-9,05%	100,00%														
Esquina	-7,79%	7,50%	100,00%													
LN renda	-0,20%	10,50%	7,64%	100,00%												
LN área terreno	-16,38%	28,61%	25,17%	8,57%	100,00%											
LN percentual área preservação	-3,71%	4,80%	-0,45%	-8,20%	17,10%	100,00%										
Infraestrutura ajustada	7,40%	16,18%	19,25%	22,16%	6,91%	-1,61%	100,00%									
Índice aproveitamento básico	-1,96%	3,96%	3,93%	17,95%	-3,75%	-17,34%	26,54%	100,00%								
Dens. Comerc. Trecho logradouro	-9,06%	27,71%	16,17%	9,54%	25,96%	1,83%	20,59%	14,65%	100,00%							

Continua

Conclusão

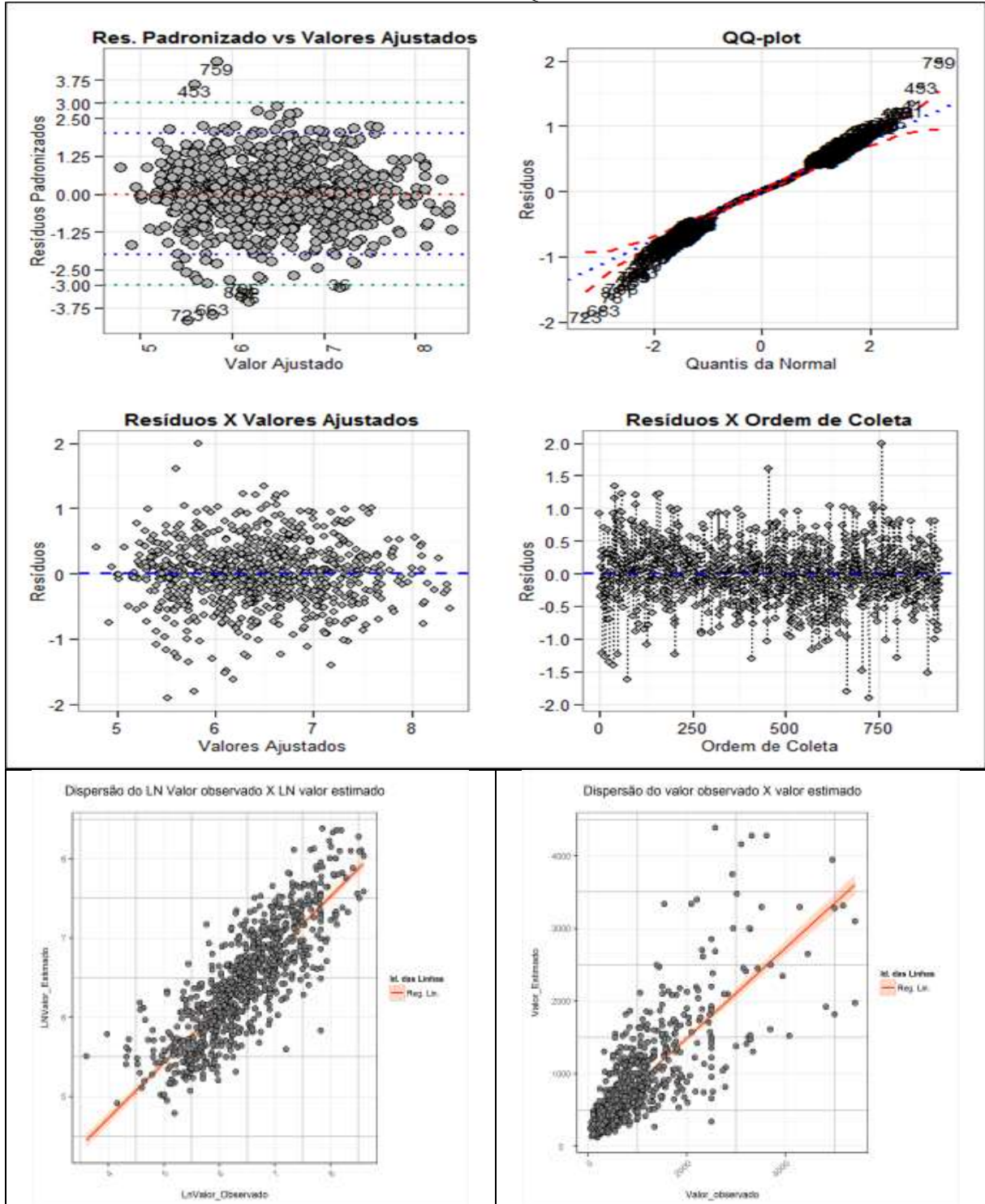
Tabela 7 – Matriz de correlação entre variáveis do modelo MQO

Variáveis	Loteamento condomínio	Avenida	Esquina	LN renda	LN área terreno	LN percentual área preservação	Infraestrutura ajustada	Índice aproveitamento básico	Dens. Comerc. Trecho logradouro	LN Valor m2 face quadra IPTU 2014	Transação	Oferta	LN Distância Via Principal	LN Interação incorporação 200	LN IDH Educação	Valor Unitário
LN Valor m2 face quadra IPTU 2014	-18,52%	26,03%	14,81%	61,88%	9,69%	-6,89%	37,30%	45,78%	25,84%	100,00%						
Transação	10,43%	6,86%	-4,74%	6,54%	-13,53%	-4,81%	-3,01%	-5,35%	-7,84%	-0,80%	100,00%					
Oferta	-24,25%	9,11%	15,70%	12,52%	34,34%	12,36%	11,20%	15,00%	15,81%	21,29%	-38,91%	100,00%				
LN Distância Via Principal	0,27%	-40,67%	-5,17%	-10,71%	-20,88%	-11,44%	-28,96%	-2,88%	-27,61%	-30,89%	2,77%	-14,71%	100,00%			
LN Interação incorporação 200	-12,64%	12,18%	28,49%	19,39%	53,62%	-0,60%	14,80%	15,08%	10,96%	24,57%	-12,42%	31,64%	-13,48%	100,00%		
LN IDH Educação	1,55%	1,95%	11,19%	51,97%	1,62%	-8,05%	41,49%	51,06%	13,99%	66,79%	-1,97%	16,40%	-23,79%	18,77%	100%	
Valor Unitário	1,04%	30,17%	19,14%	56,04%	12,35%	-13,10%	43,68%	43,55%	27,40%	75,45%	1,77%	27,77%	-37,19%	27,73%	63,4%	100%

Fonte: Elaboração do autor.

APÊNDICE D – GRÁFICO 1

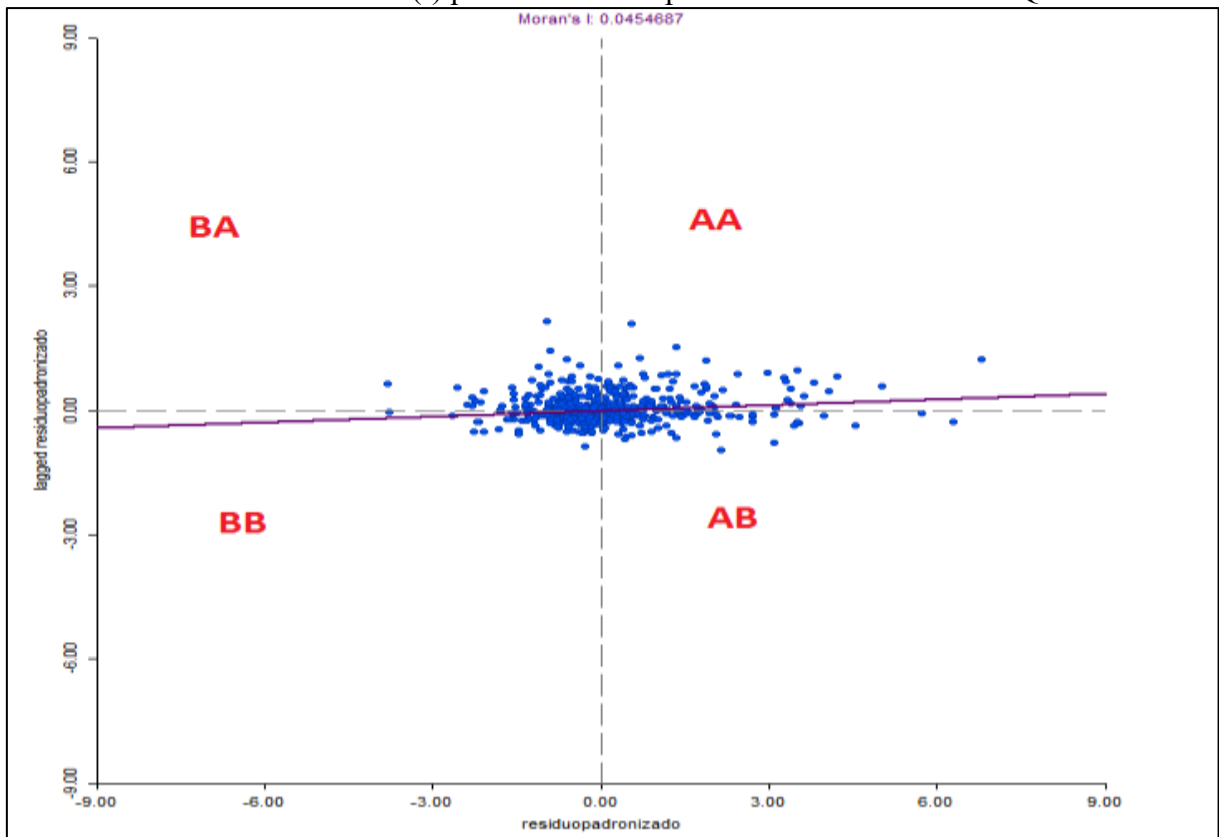
Gráfico 1 – Multicolinearidade, normalidade e independência serial dos resíduos e de dispersão dos valores estimados e observados do modelo MQO



Fonte: Elaboração do autor.

APÊNDICE E – GRÁFICO 2

Gráfico 2 – Índice de MORAN (I) para os resíduos padronizados do modelo MQO



Fonte: Elaboração do autor.

APÊNDICE F – TABELA 8

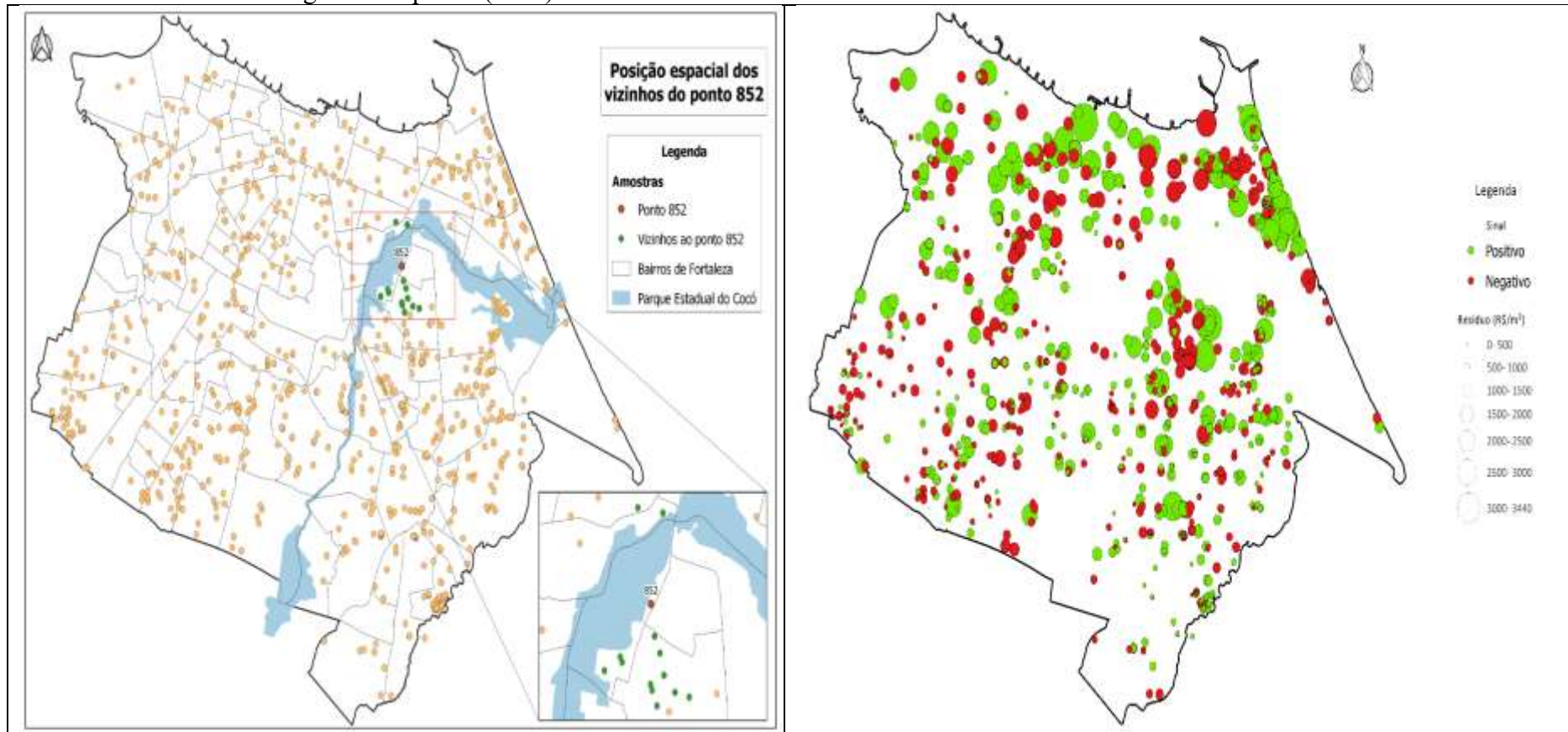
Tabela 8 – Elementos da matriz W padronizada para os 13 vizinhos do ponto 852, calculado pelo GeoDa

Referência	Vizinho	Distância	(1/d)²
852	569	1.831,14	2,98234E-07
852	899	2.407,115	1,72586E-07
852	812	2.331,896	1,839E-07
852	781	4.567,576	4,79323E-08
852	788	2.728,093	1,34364E-07
852	820	3.353,241	8,89345E-08
852	527	4.053,575	6,08588E-08
852	497	5.466,916	3,34592E-08
852	838	2.238,233	1,99613E-07
852	530	9.155,021	1,19311E-08
852	642	4.006,229	6,23058E-08
852	649	3.966,273	6,35675E-08
852	242	1.317,146	5,36545E-08

Fonte: Elaboração do autor.

APÊNDICE G – GRÁFICO 3

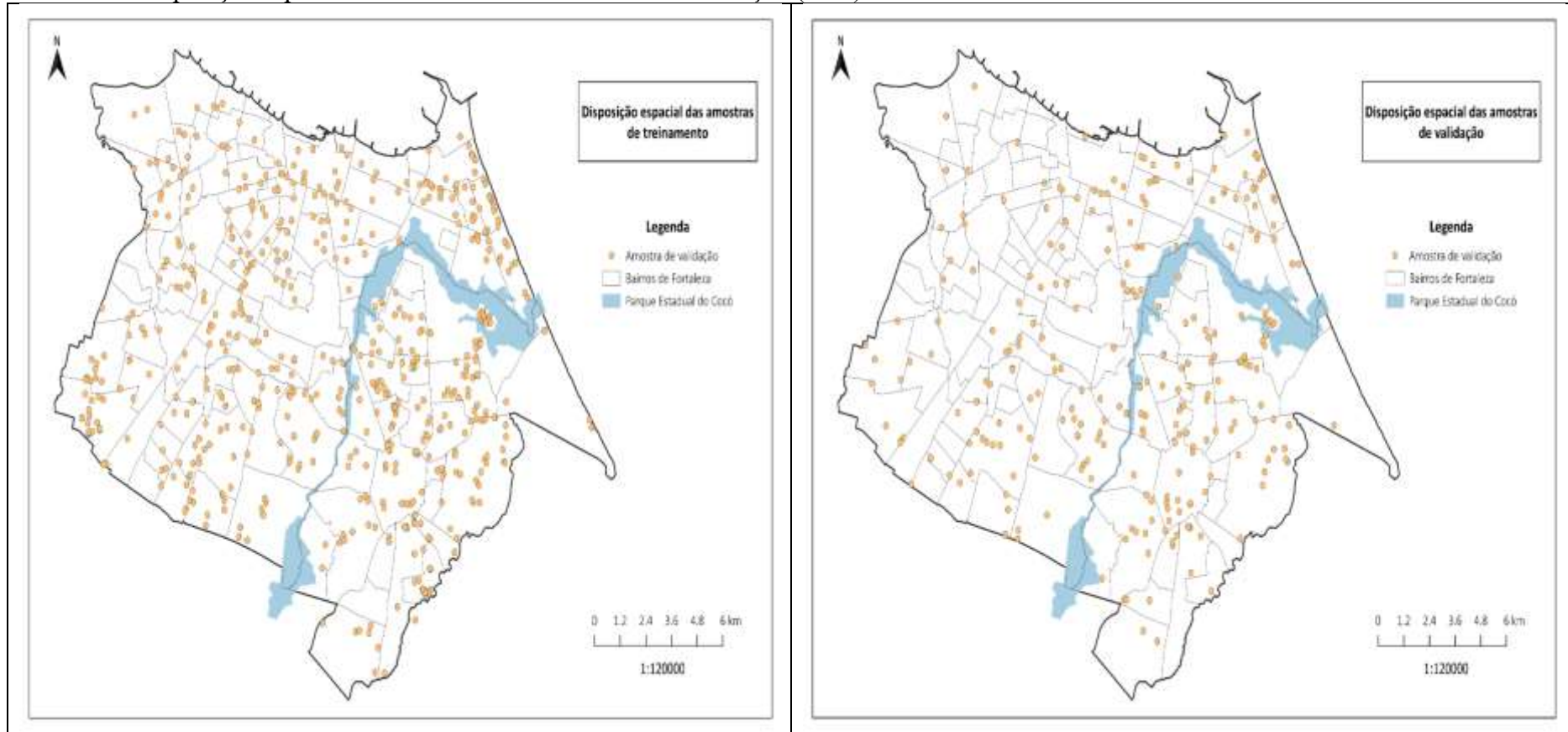
Gráfico 3 – Resíduos da Regressão espacial (SAR) – Amostra de teste



Fonte: Elaboração do autor

APÊNDICE H – GRÁFICO 4

Gráfico 4 – Disposição espacial das amostras de treinamento e validação (teste)



Fonte: Elaboração do autor.

ANEXOS

ANEXO A – QUADRO 2

Quadro 2 – Descrição das variáveis utilizadas nos modelos

Nome da Variável	Descrição
Avenida	Variável <i>dummy</i> positiva que assume valor 1(um) se o dado está situado em avenida ou rodovia e zero em caso contrário, segundo Dantas (2014).
Área do Terreno	Variável quantitativa continua negativa que indica sua área territorial (em m ²), segundo Dantas (2014).
Esquina	Variável <i>dummy</i> positiva que indica se o mesmo se encontra ou não em esquina, segundo Dantas (2014).
Loteamento em Condomínio	Variável <i>dummy</i> positiva que indica se o terreno se encontra em condomínio fechado. Essa variável foi colhida junto ao cadastro imobiliário do Município, segundo Dantas (2014).
Renda	Variável <i>proxy</i> positiva de macrolocalização, representada pela renda média do chefe da família, em salários mínimos, ajustada a uma superfície de tendência construída pelo processo de <i>Krigagem</i> , tomando-se como base os dados de renda média do responsável no setor censitário, divulgada pelo censo do IBGE (2010), segundo Dantas (2014).
Oferta	Variável <i>dummy</i> agrupada representando os dados coletados em OFERTA. Agrupada com TRANSAÇÃO. Quando OFERTA e TRANSAÇÃO são iguais a 0 (ZERO) simultaneamente, indica que o dado é uma avaliação de ITBI, segundo Dantas (2014).
Transação	Variável <i>dummy</i> agrupada representando as TRANSAÇÕES do mercado imobiliário ou valor declarado pelo contribuinte nas declarações de ITBI quando essa declaração esteja dentro do limite de mais ou menos 5% do avaliado pelo ITBI. Agrupada com OFERTA. Quando OFERTA e TRANSAÇÃO são iguais a 0 (ZERO) simultaneamente, indica que o dado é uma avaliação pura de ITBI, segundo Dantas (2014).
Valor Básico de Terrenos IPTU 2014	Variável <i>proxy</i> positiva indicando o valor unitário (R\$/m ²) base do terreno para o lançamento do IPTU, referente ao ano 2014, segundo Dantas (2014).
Densidade Comercialização no Trecho de logradouro	Variável <i>proxy</i> positiva de densidade de comercialização no trecho de logradouro onde está situado o imóvel. Representa o percentual de imóveis comerciais em relação ao total de imóveis no trecho de logradouro. Em casos de zero absoluto, para não inviabilizar sua transformação logarítmica, considerou-se 0,01, segundo Dantas (2014).
Percentual de área de Preservação (%)	Variável <i>proxy</i> positiva representando a área de preservação (ZPA1) que atinge o imóvel, segundo o plano diretor do Município de Fortaleza (PDPFor) (em m ²). Em casos de zero absoluto, para não inviabilizar sua transformação logarítmica, considerou-se 0,01, segundo Dantas (2014).

Continua

Quadro 2 – Descrição das variáveis utilizadas nos modelos

Nome da Variável	Descrição
Índice de Aproveitamento	Variável <i>proxy</i> positiva que representa o índice de aproveitamento onde está situado o imóvel segundo plano diretor. Em casos de zero absoluto, para não inviabilizar sua transformação logarítmica, considerou-se 0,01, segundo Dantas (2014).
Interação Incorporação 200	Variável <i>proxy</i> positiva de interação representado pelo produto das seguintes variáveis: área do terreno, índice de aproveitamento máximo equivalente, densidade de verticalização <i>kernel</i> 200 e lote de incorporação, segundo Dantas (2014).
Índice de Desenvolvimento Humano Educacional CENSO 2010	Variável <i>proxy</i> positiva do índice de desenvolvimento humano do ano de 2010 referente ao indicador <u>educação</u> no setor censitário onde se situa o imóvel, segundo Dantas (2014).
Infraestrutura ajustada	Variável quantitativa discreta que representa a soma dos elementos de infraestrutura, presentes em qualquer um dos trechos de logradouro para qual o imóvel tem frente, a saber: água, esgoto, galeria pluvial, sarjeta, iluminação pública e pavimentação. Essa variável foi ajustada da seguinte forma: a) valor 1 (um) se esta soma é menor igual a 3; b) 2 (dois) se esta soma é 4; c) 3 (três) se esta soma é 5 e d) 4 (quatro) se esta soma é 6, segundo Dantas (2014).
Valor Unitário (R\$/m ²)	Variável explicada ou dependente, que pode ser fornecida como o valor estimado em razão da sua área territorial, segundo Dantas (2014).

Fonte: Dantas (2014).