

MERCADO IMOBILIÁRIO E INFRAESTRUTURA DE TRANSPORTE: UMA AVALIAÇÃO DO PRÊMIO DE MERCADO PARA A PROXIMIDADE COM EQUIPAMENTOS DE TRANSPORTE COLETIVO

Área temática: 10- Economia Regional e Urbana

Arthur Bazolli Alvarenga

Mestrando em Economia Aplicada pela Universidade Federal de Juiz de Fora (PPGE/UFJF).

E-mail: arthur.bazolli@estudante.ufjf.br

Ari Francisco de Araujo Junior

Professor do Departamento de Economia do Ibmec Minas Gerais (Ibmec-MG).

E-mail: ari.junior@professores.ibmec.edu.br

Admir Antonio Betarelli Junior

Professor do Departamento de Economia da Universidade Federal de Juiz de Fora (UFJF).

E-mail: admir.betarelli@ufjf.br

RESUMO

Ainda que equipamentos de transporte coletivo como linhas de metrô possam trazer benefícios à população de um centro urbano, a estimação do valor atribuído a essas estruturas não pode ser diretamente estimada, pois não podem ser adquiridas no mercado pelos consumidores. Por isso, este trabalho utiliza a metodologia do modelo hedônico de preços para estimar indiretamente a avaliação atribuída à proximidade das estações de metrô (CMSP) e de trem (CPTM) no mercado imobiliário residencial de São Paulo – SP. Foram estimados oito modelos, sob diferentes hipóteses de estrutura urbana: monocêntrica, duocêntrica, ou, ainda, relaxando a premissa de centralidade, com a inclusão de um índice de acessibilidade cumulativa. Demais aspectos do mercado imobiliário também foram estimados, como características dos imóveis e criminalidade do bairro. Os resultados apontam que a proximidade às estações de metrô resulta em um prêmio de mercado positivo; no entanto, as estações da CPTM não se mostraram um diferencial relevante. Adicionalmente, o índice de acessibilidade se mostrou significativo quando considerado, indicando uma possível flexibilização de premissas fortes da teoria econômica a respeito da configuração urbana das cidades.

Palavras-chave: mercado imobiliário, economia urbana, acesso a oportunidades, econometria espacial.

Classificação JEL: C21, R31, R32

ABSTRACT

Although rapid transit equipment such as subway lines can benefit the population of an urban center, the value attributed to it cannot be directly estimated, as it is not possible for consumers to acquire such structures in the market. Therefore, this paper uses the hedonic prices model approach to estimate indirectly how consumers value proximity to subway (CMSP) and commuter rail (CPTM) in the real estate market of São Paulo, Brazil. Eight models were estimated, under different underlying hypothesis about the urban structure: monocentric, duocentric, or yet relaxing the centrality assumption, with the inclusion of a cumulative accessibility index. Other aspects of the real estate market were also estimated, such as property's attributes and neighborhood criminality. The results indicate that proximity to subway stations results to a positive market premium, whereas commuter rail appear not to be a relevant feature. In addition, the accessibility index revealed itself significant when considered, showing that its inclusion can possibly make strong assumptions from the economic theory regarding the urban form of cities more flexible.

Keywords: housing market, urban economics, access to opportunities, spatial econometrics.

JEL codes: C21, R31, R32

1 INTRODUÇÃO

Nos grandes centros urbanos, a necessidade de deslocamento para acessar o mercado de trabalho, serviços de saúde e demais amenidades dificilmente consegue ser suprida apenas pelos meios de transporte individuais. Bertaud (2018) aponta que a provisão adequada da infraestrutura de transporte beneficia uma região com as economias de aglomeração por dois motivos principais: o mercado de trabalho se expande – trabalhadores podem atingir mais empregos em menos tempo – e as interações humanas tornam-se mais frequentes, o que alavanca a inovação e, portanto, a produtividade. Por outro lado, uma deficiência na provisão dessa infraestrutura pode resultar em uma perda de eficiência econômica da região como um todo, além de menor qualidade de vida para os indivíduos. Os mercados de trabalho se fragmentam, o que diminui a capacidade de especialização econômica, e aumentam os custos de transporte da economia como um todo.

Shoup (2003) nota que investimentos em infraestrutura podem reduzir o custo operacional do serviços públicos e valorizar a região beneficiada em valor superior ao seu custo de implementação. No entanto, a dificuldade em financiá-los impede que esses investimentos ocorram, o que leva à fragmentação das cidades e a maiores custos para prover a infraestrutura urbana. O Projeto Acesso a Oportunidades, do Instituto de Pesquisa Econômica Aplicada (Ipea), revela um padrão de desigualdade espacial nas grandes cidades do país, em que as regiões centrais são dotadas de mais infraestrutura e oportunidades do que as periferias (Pereira *et al.*, 2019). Esse padrão resulta, dentre outros fatores, do histórico de baixo investimento em mobilidade urbana no Brasil, descompassado da rápida urbanização dos centros urbanos (BNDES Setorial, 2015).

Os instrumentos de captura de valor podem financiar investimentos em infraestrutura urbana. Suzuki *et al.* (2015) estudaram algumas experiências como o caso de Tóquio: o governo promove uma operação urbana de desenvolvimento voltado para o transporte (*transit-oriented development* – TOD)¹, em que unifica os terrenos e a venda de um novo empreendimento imobiliário, mais adensado, financia em parte equipamentos públicos.

O Brasil também é pioneiro no desenvolvimento de operações urbanas. Após o surgimento do arcabouço legal para as OUCs (Operações Urbanas Consorciadas), com o Estatuto da Cidade (BRASIL, 2001), o município de São Paulo desenvolveu as suas primeiras operações do tipo – as OUs Água Espriada e Faria Lima –, em que a prefeitura vende ao incorporador imobiliário o direito de construir acima do coeficiente de aproveitamento básico² e os recursos são usados para investimento em infraestrutura urbana. No entanto, os recursos da experiência paulistana não foram investidos em transporte e tiveram efeitos colaterais inesperados, resultantes da redução do CA básico em toda a cidade (Suzuki *et al.*, 2015).

Na literatura acadêmica, a valoração dos equipamentos de transporte coletivo usualmente é feita pelo modelo de preços hedônicos, o qual permite avaliar a demanda implícita por diferentes atributos no mercado imobiliário (Brueckner, 2011). Essa metodologia forneceu evidências diversas: Duncan (2008) identificou que a proximidade a estações de VLT em San Diego (EUA) valorizou imóveis residenciais de forma diferente para casas e apartamentos; em Phoenix (EUA), Atkinson-Palombo (2010) detectou um padrão parecido. No Brasil, Hermann e Haddad (2005) encontraram uma relação ambígua para São Paulo (SP), enquanto Seabra, Silveira Neto e Menezes (2016) concluíram que o metrô do Recife (PE) tem um efeito negativo no valor dos imóveis.

Este trabalho desenvolve um modelo hedônico para o mercado imobiliário residencial de São Paulo (SP), controlando os diferentes aspectos dos imóveis e da vizinhança, objetivando identificar os prêmios de vantagem locacional proporcionados pela infraestrutura de transporte coletivo. Assim, espera-se testar a hipótese de externalidades positivas do transporte público propostas pela teoria econômica. O estudo contribui para a literatura em duas escalas. A nível local, a estratégia econométrica analisa os impactos indiretos das amenidades urbanas. Ao mesmo tempo, as informações obtidas sobre os imóveis e as

¹ O TOD tem como diferencial a integração do planejamento de uso e ocupação do solo com os equipamentos de transporte coletivo, criando espaços convidativos para o pedestre e de alta acessibilidade (Suzuki *et al.*, 2015).

² O coeficiente de aproveitamento (CA) é a razão entre a área construída (ou edificável) computável e a área do lote. Determina o potencial construtivo de um terreno. No caso de São Paulo, a prefeitura determina os limites de CA para cada zona (Prefeitura do Município de São Paulo, 2022).

características da cidade lança um novo olhar para o mercado habitacional paulistano e, por fim, fornece uma estimativa atualizada à luz das expansões da rede de transporte sobre trilhos da cidade.

Na escala nacional, o trabalho é pioneiro em incorporar no modelo hedônico o índice de acessibilidade cumulativa desenvolvido pelo Ipea, o qual mede a porcentagem de empregos da cidade que podem ser acessados em determinado tempo a partir de um ponto. Tal índice é vantajoso pois captura diretamente as vantagens locacionais de um imóvel advindas da acessibilidade, enquanto o uso das distâncias a equipamentos de transporte não é uma medida exata da acessibilidade (Higgins e Kanaroglou, 2016). Além disso, o índice pode refinar a modelagem nos trabalhos aplicados, na medida em que permite abandonar a premissa de centros de emprego atomizados.

As próximas seções são divididas da seguinte maneira: na seção 2, é feita uma revisão do referencial teórico da economia urbana e, na seção 3, da literatura empírica; partindo, então, para a discussão da metodologia adotada, na seção 4. A seção 5 descreve a base de dados e expõe os resultados encontrados; por fim, a seção 6 **Erro! Fonte de referência não encontrada.** conclui o trabalho.

2 REFERENCIAL TEÓRICO

2.1 Modelos de economia urbana

Os primeiros estudos em economia urbana se deram com von Thünen, no século XIX (Morency-Lavoie, 2015). No entanto, a formalização matemática só foi desenvolvida no XX, em especial por Alonso, Muth e Mills. Brueckner (1987) consolidou esse modelo de equilíbrio espacial urbano, conhecido como Modelo AMM. O modelo tem como ponto de partida uma cidade radial com indivíduos homogêneos que consomem moradia (em área) e um bem numerário e trabalham no CBD (*core business district*), um único ponto no centro da cidade. No equilíbrio, os indivíduos escolhem entre morar perto do trabalho ou consumir mais moradia, mas em uma área mais afastada: assim, a cidade tende a ser mais densa e mais valorizada quanto mais perto do CBD.

Um resultado do modelo pertinente para este trabalho é o impacto de uma mudança no custo do transporte, exógeno ao modelo, sobre o equilíbrio de preços e o nível de utilidade. Considerando que o custo de transporte é o custo de oportunidade do deslocamento (Mills, 1967), espera-se que um investimento em sistemas de transporte reduzam-no. É importante destacar que embora o modelo tenha premissas simplificadoras, como estoque elástico de moradias e custo homogêneo de transporte, os impactos de uma redução no custo de transporte sobre o custo de habitação permanecem e são reforçados: agentes econômicos tendem a maximizar a utilidade e lidam com o *tradeoff* entre mais espaço e menores custos de deslocamento, de forma que um investimento em mobilidade que provoque um diferencial de acessibilidade aumenta a pressão competitiva no mercado imobiliário local

Fujita e Ogawa (1982) desenvolveram um modelo em que firmas e famílias interagem entre si na escolha da localização ótima. As famílias escolhem os locais de trabalho e de residência que maximizam o consumo do bem composto. As firmas empregam as famílias e competem com elas (e entre si) pelo consumo de terra: a escolha do local ótimo é inerentemente ligada a um potencial local, que reflete as economias de aglomeração. Quanto maior esse potencial, mais as firmas preferem concentrar-se em um local, dando origem às centralidades. O equilíbrio pode assumir diversas formas, da monocêntrica à acêntrica, a depender do conjunto de parâmetros. Comparando as diferentes formas, a configuração monocêntrica é um equilíbrio somente quando as vantagens locais são altas o bastante para compensar os custos de transporte. O seu oposto é a configuração urbana completamente mista, a qual somente ocorre se os custos de transporte forem tão altos que superam os potenciais locais.

Para dar maior sustentação ao arcabouço teórico, é útil analisar os resultados de estudos empíricos que o analisaram. Bertaud e Malpezzi (2003) estudaram a distribuição espacial em 48 centros urbanos, por meio de dados populacionais ao nível do setor censitário, coletados pelos autores, e de variáveis socioeconômicas, como PIB, obtidas do Banco Mundial. A densidade demográfica foi calculada, para cada setor censitário, dividindo-se a sua população pela área construída (i.e., líquida de parques, lagos etc.). Ainda que muitas características das cidades reais infrinjam as premissas do modelo AMM, o gradiente de densidades exponencialmente decrescente se verifica na maioria dos casos e tende a diminuir frente a

aumentos na renda e na população ou a uma redução nos custos de transporte, assim como suposto na teoria. Os autores destacam que nas cidades classificadas como orientadas para o mercado (Singapura, por exemplo), os gradientes de preço e de densidade são consoantes, refletindo o potencial das regiões centrais. Já naquelas onde intervenções foram mais fortes ao longo da história, a tendência é uma inversão no gradiente de densidades, mas não no gradiente de preços: os locais centrais são os mais valorizados, mas são também menos adensados. Moscou seria o exemplo extremo desse caso, mas Brasília compartilha características semelhantes.

3 ESTUDOS EMPÍRICOS SOBRE AMENIDADES URBANAS

Duncan (2008) analisou a interação da rede de transporte de San Diego (EUA) com as habitações da cidade, com atenção à heterogeneidade dos bairros e moradias: foram estimados dois modelos hedônicos separados, um para casas e outro para apartamentos, e foram incorporadas variáveis binárias para cada bairro da cidade, para captar suas características intrínsecas. A variável de acessibilidade usada foi a distância em milhas a uma estação do sistema de transporte. No entanto, as variáveis ambientais, em sua maioria, foram avaliadas como a proporção da área em um raio de 0,25 milha do imóvel que é dedicado àquela amenidade (por exemplo, praias e estacionamentos). O autor estimou um modelo de mínimos quadrados ordinários (MQO) com erros padrão robustos à heterocedasticidade e encontrou uma pequena dependência espacial, a qual não foi modelada. Os resultados indicaram uma valorização de 17% para apartamentos próximos às estações e de 6%. Isso viabilizaria financeiramente operações de TOD, na medida em que o mercado absorve as vantagens locacionais da proximidade.

Atkinson-Palombo (2010) estudou os efeitos no mercado imobiliário da construção de uma linha de VLT em Phoenix (EUA). A amostra de 10 mil imóveis foi dividida em quatro grupos diferentes, a depender do uso predominante (residencial ou misto) e do tipo da moradia (casa ou apartamento), seguindo o raciocínio de Duncan (2008). Após a construção do VLT, as áreas de uso misto se apreciaram mais, sobretudo os apartamentos, enquanto os bairros residenciais sofreram desvalorizações. Os resultados heterogêneos foram atribuídos às diferentes demografias atraídas pelos tipos de bairros e moradias. Os modelos foram estimados por MQO agrupado e consideraram a valorização entre os períodos antes e após a conclusão das obras. Como a dependência espacial era pequena, não houve modelagem de econometria espacial. Em relação à estrutura urbana, foi considerado um modelo duocêntrico, em que se calcula a distância à centralidade mais próxima do imóvel. A distância às estações de VLT foi medida por uma variável binária, igual a 1 se o imóvel estiver a meia milha de distância de uma estação.

Hermann e Haddad (2005) analisaram o impacto de diferentes amenidades urbanas no mercado de aluguéis de São Paulo (SP), extraído da Pesquisa de Orçamentos Familiares (POF) de 1999 o valor declarado para o aluguel por cerca de 500 famílias. Destacam-se no trabalho a consideração da sintaxe urbana e o processo de escolha das variáveis ambientais. Fundamentados nos modelos AMM e Fujita-Ogawa, os autores estimaram modelos com configurações monocêntricas e duocêntricas, centrados na Sé, o centro tradicional de São Paulo, e no eixo Berrini-Faria Lima, novo centro financeiro da cidade. Já a escolha das variáveis ambientais se deu por uma regressão do aluguel contra as variáveis estruturais dos imóveis (como a área e o número de quartos). Assim, ao projetarem as curvas de nível dos resíduos no mapa da cidade, as regiões em que os resíduos têm valores maiores indicam a ocorrência dessas externalidades, o que permitiu selecioná-las com maior critério. As regressões estimadas por MQO consideraram, ainda, a presença de dependência espacial, mas esta foi rejeitada.

O estudo identificou o modelo monocêntrico com CBD na Sé como o mais adequado. Os principais resultados foram o gradiente de preços partindo do CBD, em conformidade com a teoria, um prêmio de mercado para a proximidade com as estações da CPTM e a não-significância estatística da proximidade com as estações do Metrô, por dois motivos. Primeiro, o seu efeito ambíguo: aumentam a acessibilidade e desenvolvem o comércio, mas produzem poluição sonora e podem aumentar a criminalidade. Esse efeito é análogo à depreciação nos imóveis dos bairros residenciais próximos às estações em Atkinson-Palombo (2010). Segundo, a extensão da rede do Metrô, então limitada ao centro da cidade, pode ter feito com que a variável captasse efeitos de outras variáveis ambientais da região não analisadas por sua baixa disponibilidade

Seabra, Silveira Neto e Menezes (2016) estimaram o impacto das amenidades urbanas em residências da capital pernambucana a partir de cerca de 9,7 mil transações residenciais ocorridas em 2012, disponíveis em uma base de dados da prefeitura referente ao ITBI (Imposto sobre Transações de Bens Imóveis), cobrindo 90% da área urbana. Além de usar uma base abrangente, o estudo separa os impactos diretos e indiretos (efeito transbordamento) de cada atributo. Isso é possível em razão da técnica de econometria espacial adotada frente à presença de dependência espacial positiva e significativa. Assim como Hermann e Haddad (2005), os autores analisaram os resíduos da regressão estrutural para escolher as variáveis ambientais. Consideraram apenas a estrutura monocêntrica, mas em forma quadrática, refletindo a hipótese de que o gradiente de preços em relação ao CBD tem um ponto de máximo a uma certa distância deste, diminuindo a partir de então. Os resultados obtidos foram significativos, validando a hipótese. Variáveis ambientais de destaque foram a proximidade ao rio Capibaribe e à orla do mar, a quantidade de crimes em um raio de 1,5km do imóvel e proximidade às Zonas Especiais de Interesse Social (ZEIS)³, sendo todas significativas e que as duas últimas tendem a desvalorizar o imóvel. Já a proximidade às estações de metrô também se revelou como uma desamenidade urbana. O resultado contraintuitivo pode se dever à baixa penetração da rede nas áreas mais nobres da cidade e também à sua estrutura, totalmente em superfície, que pode provocar externalidades negativas em razão da poluição sonora.

4 METODOLOGIA E BASE DE DADOS

À luz da discussão teórica e empírica, discutiremos o modelo hedônico de preços e a especificação do modelo estimado. O protocolo ideal para a estimação de um modelo hedônico deve partir da estimação tradicional por MQO e, em seguida, analisar a presença de dependência espacial e determinar o modelo que melhor estima os dados. No modelo hedônico, a dependência pode se manifestar se os ofertantes precificarem seus imóveis com base no preço das propriedades vizinhas. Concomitantemente, deve-se prestar atenção para a presença da heterocedasticidade (Almeida, 2012; LeSage e Pace, 2009; Tyszler, 2006; Wooldridge, 2016).

4.1 Modelo hedônico

O modelo hedônico decompõe gastos com um bem em diferentes características, o que possibilita atribuí-las um valor implícito (Malpezzi, 2002). No mercado imobiliário, as características analisadas são os aspectos estruturais da propriedade – área, banheiros, quartos etc. – e as amenidades urbanas, que influenciam (positiva ou negativamente) no valor do imóvel.

Higgins e Kanaroglou (2016) analisaram mais de 130 estudos realizados entre 1970 e 2010 sobre a captura de valor de investimentos em mobilidade urbana na América do Norte. Destes, 106 utilizaram o modelo de preços hedônicos. No entanto, as diferenças na metodologia dessas análises em relação à forma funcional, às variáveis escolhidas, às proxies e a eventuais variáveis omitidas causam uma grande disparidade nos resultados produzidos, potencialmente tornando os modelos heterocedásticos e enviesados, o que dificulta a sua comparação. Nesse sentido, propõem a adoção de uma série de critérios para melhorar a qualidade dos modelos, como: (i) incorporar medidas de acessibilidade relativa em vez do uso da distância como proxy – como adiantado na introdução deste trabalho –, (ii) controlar o modelo para fatores como outras amenidades e características como desenvolvimento orientado para o transporte (TOD), e (iii) adotar controles para dependência espacial e heterogeneidade nos dados, como a medida de autocorrelação espacial do I de Moran, em sua maior parte incorporadas pelos demais trabalhos aqui analisados.

O conjunto de regressores não é formalmente definido pela literatura; contudo, Malpezzi (2002) destaca o número de cômodos, a área construída e qualidade do acabamento, entre as variáveis estruturais. As variáveis externas ao imóvel devem abarcar qualidade da vizinhança e distância a pontos de interesse.

³ ZEIS são “áreas de assentamentos habitacionais de população de baixa renda, surgidos espontaneamente, (...)” ou, ainda, “áreas de Programas Habitacionais de Interesse Social (...) destinadas, prioritariamente, às famílias originárias de projetos de urbanização. (Prefeitura do Recife, [s.d.]”)

Essas variáveis devem ser definidas no contexto do local de estudo. Seguindo a estratégia de Hermann e Haddad (2005) e de Seabra, Silveira Neto e Menezes (2016), uma forma de analisar as externalidades causadas pelas variáveis ambientais é regredir o valor do imóvel contra as variáveis estruturais, deixando no resíduo as variáveis explicativas. Pela distribuição espacial e sinal dos resíduos, é possível comparar se as aparentes externalidades positivas e negativas coincidem com características esperadas para a cidade, por exemplo: bairros nobres, presença de vegetação, proximidade de centros de emprego.

Quanto às variáveis de acessibilidade, a não relevância estatística das estações do metrô paulistano e a desamenidade no caso do Recife podem advir do efeito ambíguo desses equipamentos e do baixo potencial de acessibilidade desses modos de transporte, como os próprios autores destacaram. Passados 17 anos da publicação do trabalho de Hermann e Haddad (2005), a rede do Metrô-SP expandiu de 55,9 quilômetros (CMSP, 2018) para 101 quilômetros (CMSP, 2021), cerca de 80%, e estabeleceu conexões com mais regiões da cidade. Assim, espera-se encontrar neste trabalho maior relevância da rede do Metrô no valor dos aluguéis, em razão do aumento de sua importância como instrumento de acessibilidade.

Higgins e Kanaroglou (2016) debatem ainda sobre medidas de acessibilidade e de uso do solo, centrais neste estudo. Partindo da definição de acessibilidade como o potencial para alcançar oportunidades distribuídas no espaço (Páez *et al.*, 2012), sugerem que o uso da proximidade a estações de transporte coletivo como *proxy* para acessibilidade pode não ser uma boa métrica, por não considerar diretamente os ganhos de acessibilidade advindos do equipamento. Este pode ser o caso se uma linha de metrô não atender às principais zonas de interesse de uma cidade, ou então o acesso via automóvel ou bicicleta for mais competitivo.

Nesse sentido, aqui incorporamos, além das medidas de acessibilidade tradicionais, o índice de oportunidades cumulativas (Pereira *et al.*, 2019). Como destacam Higgins e Kanaroglou (2016), a incorporação de tal medida permite uma análise mais detalhada dos reais ganhos de acessibilidade proporcionados pelos equipamentos de transporte coletivo, na medida em que afere as oportunidades de emprego que este permite alcançar em um intervalo de tempo.

Em relação à forma funcional, múltiplos autores apontam para a log-linear⁴. Esta apresenta vantagens, como o significado intuitivo – os resultados podem ser interpretados como a variação percentual no aluguel frente a um incremento unitário das variáveis dependentes⁵ (Malpezzi, 2002) – e a atenuação da heterocedasticidade (Malpezzi, 2002; Wooldridge, 2016).

4.2 Estratégia econométrica

As regressões realizadas seguem uma forma geral, similar a Hermann e Haddad (2005):

$$P = f(S, E, A), \quad (1)$$

em que P é o valor do aluguel, S é o vetor variáveis estruturais, como área do imóvel e número de quartos, A incorpora as variáveis de acessibilidade – distância aos CBDs, às estações e o índice de oportunidades cumulativas –, e E considera as variáveis ambientais.

Após a análise exploratória de dados, as variáveis ambientais escolhidas foram as distâncias geodésicas às ZEIS 1 e 5, em quilômetros ($dist_zeis1$ e $dist_zeis5$). Incluímos as variáveis de distância aos parques e às unidades de conservação, porém a primeira resultou não significativa. O mesmo resultado se sucedeu com a variável indicativa de distância às ZEIS 3. Para considerar o efeito da criminalidade, foi incorporado o logaritmo natural do número de crimes hediondos cometidos no distrito policial em que o imóvel se localiza:

$$E = dist_conserva + dist_zeis1 + dist_zeis5 + \ln(crime) \quad (2)$$

As variáveis estruturais escolhidas foram o logaritmo da área em m², o n° de banheiros, o andar do apartamento e variáveis binárias para ar-condicionado, chuveiro a gás, presença de mobília, imóvel novo ou reformado, academia no edifício e sauna no edifício.

⁴ Para clarificação, a expressão “log-linear”, se refere à especificação conforme definida por Wooldridge (2016), na qual a variável dependente se encontra na forma logarítmica, mas não a variável independente.

⁵ Essa interpretação pode ser aproximada por $\Delta y = 100(\beta_i)$ ou $\Delta y = 100 \cdot \exp(\beta_i) - 1$, esta última mais precisa.

$$S = \log(\text{area}) + \text{suites} + \text{andar} + \text{tipo} + \text{arcond} + \text{armqua} + \text{gas_chuv} + \text{mobil} + \text{novo_ref} + \text{gas_cano} + \text{quadra} + \text{sauna} \quad (3)$$

Na Tabela 1 encontram-se resumidas as combinações estatisticamente significantes para as variáveis de acessibilidade, dentre as diversas estimadas. A regressão (1) assumiu um modelo duocêntrico, enquanto as regressões (2) e (3) seguem a estrutura monocêntrica, em que os diferentes CBDs propostos são a Praça da Sé – representada pela sua estação de metrô – e a avenida Faria Lima, no seu cruzamento com a av. Horácio Lafer. O primeiro local foi escolhido em razão da sua significância histórica, assim como em Hermann e Haddad (2005), enquanto o segundo está em um ponto central do novo centro financeiro da cidade. As regressões (4) a (8) diferem ao incorporam as oportunidades de trabalho cumulativas que podem ser atingidas por transporte público em 60 minutos, como medida de acessibilidade, em diferentes cenários: complementar às variáveis de distância às estações e modelo monocêntrico (4); complementar àquelas variáveis no modelo duocêntrico (5 e 6); medida de acessibilidade como substituta da variável de distância ao centro (7); e, por fim, variável de acessibilidade cumulativa no lugar de todas as demais variáveis de acessibilidade (8).

Tabela 1: variáveis ambientais nas diferentes regressões analisadas.

Variável	Descrição	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>dist_cbd</i>	Distância ao CBD mais próximo	•			•				
<i>dist_se</i>	CBD = Sé		•			•			
<i>dist_farialima</i>	CBD = Faria Lima			•			•		
<i>dist_metro</i>	Dist. estação do Metrô mais próxima	•	•	•	•	•	•	•	•
<i>dist_cptm</i>	Dist. estação da CPTM mais próxima	•	•	•	•	•	•	•	•
<i>cmatt60</i>	Acessibilidade cumulativa x10				•	•	•	•	•

Fonte: elaboração própria.

A dependência espacial é testada pelo Índice de Moran, amplamente aplicado na literatura. Um sinal positivo no valor do índice indica similaridade: valores altos da variável estão próximos a valores altos; valores baixos são vizinhos de valores baixos. Um sinal negativo indica que valores altos são geralmente circundados por valores baixos, e vice-versa. Neste estudo, o primeiro caso pode indicar que os imóveis se distribuem em *clusters* de acordo com o seu valor. A matriz de ponderação espacial é especificada seguindo o procedimento de Baumont (2004): em primeiro lugar, estima-se o modelo por MQO; em seguida, testa-se a presença de autocorrelação espacial nos resíduos dessa regressão pelo I de Moran usando diferentes matrizes espaciais. Assim, a matriz que gerar o maior índice estatisticamente significativo é a ideal, pois é a que conseguiu melhor captar a dependência espacial.

Caso o I de Moran indique a presença de dependência espacial, é necessário ainda especificar o modelo que melhor represente o verdadeiro processo gerador dos dados. Tyszler (2006) sugere iniciar do modelo SAC: se apenas ρ ou apenas λ for significativo, parte-se respectivamente para os modelos SAR ou SEM. Se ambos ρ e λ o forem, deve-se então estimar o modelo SAC e analisar se ρ permanece significativo: em caso positivo, o modelo SAC está mais próximo da realidade; em caso negativo, o modelo ideal é o SEM. Os três modelos estão descritos abaixo, conforme Almeida (2012):

Quando a dependência se dá apenas na variável dependente, estima-se o modelo SAR – *spatial autoregressive*, ou modelo de defasagem espacial –, definido da seguinte maneira:

$$y = \rho Wy + X\beta + \varepsilon, \quad (4)$$

em que $X\beta$ representa os regressores convencionais do modelo clássico e seus respectivos coeficientes, ε é o termo de erro, Wy é o vetor $n \times 1$ de defasagens espaciais para a variável dependente, y , e ρ é o coeficiente de autorregressão espacial, $-1 < \rho < 1$, de modo a evitar um comportamento explosivo. Se ρ for estatisticamente significativa e diferente de zero, há dependência espacial na variável dependente

Se, porém, a dependência espacial é residual, o modelo adequado é o SEM – *spatial error model*, ou modelo de erro espacial, definido como

$$\begin{aligned} y &= X\beta + \xi, \\ \xi &= \lambda W\xi + \varepsilon \end{aligned} \quad (5)$$

em que $W\xi$ é a defasagem no termo de erro, λ é o coeficiente autorregressivo e os distúrbios da observação i são uma média dos distúrbios de seus vizinhos j somada ao erro aleatório, ε . Assim como no modelo SAR, $-1 < \lambda < 1$.

Uma terceira variante é o modelo SAC – *spatial autorregressive mixed complete* (Tyszler, 2006) –, o qual considera uma relação de dependência espacial tanto no regressando, quanto no termo de erro:

$$y = \rho W_1 y + X\beta + \lambda W_2 \xi + \varepsilon, \quad (6)$$

em que os subscritos em W indicam que as matrizes de ponderação não precisam ser idênticas, λ é o coeficiente de autocorrelação nos erros e ρ e λ situam-se no intervalo $(-1,1)$.

Kelejian e Prucha (1998) sugeriram a estimação por mínimos quadrados em dois estágios, dando origem aos métodos STSLS – *spatial two-stage least squares* –, aplicado ao modelo SAR, e a sua forma geral, GSTSLS (*generalized STSLS*), para o modelo SAC. Usando $W_1 X$ e $W_1^2 X$ como instrumentos, é possível obter estimadores consistentes e que não dependem da premissa de homocedasticidade, ao contrário do estimador por MV.

Além de exigirem métodos de estimação diferentes do modelo linear clássico, os coeficientes dos modelos espaciais não podem ser interpretados da maneira usual. LeSage e Pace (2009) apontam que não é possível interpretar o coeficiente de uma variável i como sendo a derivada parcial do regressando, $\widehat{\beta}_r = \partial y / \partial x_r$: essa interpretação depende da hipótese, válida no modelo clássico, de independência entre as variáveis. Porém, como as variáveis são espacialmente dependentes, uma mudança no valor de x em uma região pode afetar o valor de y em outra região; esta, ainda, pode afetar novamente a variável inicialmente alterada, por um efeito retroalimentador (*feedback*).

As chamadas medidas de impacto permitem medir corretamente a influência das variáveis, dividindo-a em impactos diretos, indiretos e totais e computando o seu valor médio entre as observações. De acordo com LeSage e Pace (2009): o impacto **direto** médio é similar ao coeficiente analisado no modelo clássico, na medida em que mede a mudança na variável y_i diante de uma mudança na variável x_r na própria observação, x_{ir} . Já o impacto **indireto** médio reflete o impacto em y_i frente a uma alteração em x_{jr} , para todas as demais observações j . Por fim, o impacto **total** médio é a média de todas as derivadas parciais de y_i com respeito a todos os x_{jr} , inclusive quando $i = j$. Desta forma, o impacto indireto pode ser calculado como o impacto direto subtraído do impacto total. Para a variável r :

$$\bar{M}(r)_{indireto} = \bar{M}(r)_{total} - \bar{M}(r)_{direto} \quad (7)$$

5 BASE DE DADOS E RESULTADOS

5.1 Base de dados

Os dados utilizados neste trabalho foram tratados no R, linguagem de programação e ambiente para gráficos e computação estatística (R Foundation, 2021), por meio de suas funções nativas e de funções voltadas ao tratamento de dados do pacote *tidyverse* (Wickham *et al.*, 2019). As informações utilizadas, como imóveis, estações de metrô e limites administrativos, são georreferenciadas: assim, o pacote *sf* (Pebesma, 2018) permite analisar suas relações topológicas, como a distância entre um apartamento e a estação mais próxima. O índice de Moran e a matriz de ponderação espacial foram calculados usando o pacote *spdep* (Bivand e Wong, 2018), enquanto os modelos de econometria espacial são estimados a partir do pacote *spatialreg* (Bivand, Millo e Piras, 2021). Os mapas e gráficos da análise exploratória também foram feitos no R, com os pacotes *tmap* (Tennekes, 2018) e *ggplot2* (Wickham, 2016).

Este trabalho utiliza uma base de imóveis anunciados na plataforma web “QuintoAndar”, coletadas em 23/06/2020. A amostra contém aproximadamente 5 mil imóveis à venda naquela data, dos quais cerca de 1,4 mil também estão disponíveis para aluguel. O valor do aluguel encontra-se líquido de impostos e taxas condominiais. As propriedades estavam georreferenciadas na projeção WGS 84, padrão utilizado internacionalmente (no GPS, por exemplo); no entanto, foram convertidas para a projeção SIRGAS 2000 / UTM zone 23S para compatibilização com os demais arquivos espaciais (*shapefiles*).

Do *site* da Secretaria de Segurança Pública de São Paulo (SSP, 2021) foram coletadas as ocorrências registradas no ano de 2019 em cada distrito policial (DP) da cidade. A base de dados é agregada ao nível dos DPs, cujos limites não coincidem com os distritos administrativos da cidade, o gera complicações relacionadas ao problema de mudança de área (Almeida, 2012). Foram testadas duas possibilidades no modelo econométrico: a densidade dos crimes pela área dos DPs e o número absoluto de crimes. Foram testadas diferentes combinações de dados da base, como uma proporção de crimes pela área do distrito. A variável de melhor ajuste foi uma criada a partir da soma dos crimes hediondos informados na base⁶, de acordo com a legislação vigente (BRASIL, 1990).

Do portal “GeoSampa” (“Sistema de Consulta do Mapa Digital da Cidade de São Paulo”, 2021), da prefeitura de São Paulo, foram obtidos os shapefiles dos limites administrativos dos distritos e subprefeituras da cidade, das linhas e estações do Metrô e da CPTM, das ZEIS e dos parques e unidades de conservação. As datas de referência são 01/01/2011 para os distritos e 01/08/2018 para as subprefeituras de acordo com o portal.

A cidade conta com cinco tipos diferentes de ZEIS, das quais três serão analisadas: ZEIS 1, ZEIS 3 e ZEIS 5⁷. A ZEIS 1 equivale aproximadamente às áreas de ocorrência de ocupações informais e empreendimentos de habitação social; o objetivo é usá-la como variável de vizinhança assim como em Seabra, Silveira Neto e Menezes (2016). Já as ZEIS 3 e ZEIS 5 equivalem respectivamente a áreas com imóveis ociosos e a lotes subutilizados, que estejam em regiões bem equipadas de infraestrutura: a intuição por trás da dessas variáveis é analisar se a proximidade a imóveis abandonados/encortiçados e a terrenos baldios é uma desamenidade.

Os parques urbanos e Unidades de Conservação Ambiental (APA) datam de janeiro de 2016, segundo os metadados do portal. O objetivo é incluir uma variável de distância geodésica da residência ao parque e à unidade de conservação mais próximos, em linha reta, como variáveis de aprazibilidade da vizinhança. No entanto, há baixa quantidade dessas amenidades na região estudada, o que pode resultar na baixa significância das variáveis.

Os arquivos das linhas e estações do transporte metropolitano equivalem à totalidade da rede em operação na data de coleta da base de imóveis. A distância geodésica às estações de metrô e de trem mais próximas serão incluídas como variáveis de acessibilidade no modelo.

Por fim, o índice de acessibilidade para a cidade foi coletado no R por meio de um pacote elaborado pelo Ipea, como parte do Projeto Acesso a Oportunidades. O índice é uma estimativa do “tempo mínimo que se leva para acessar a oportunidade mais próxima” (Pereira *et al.*, 2019) – dentre os diferentes índices disponíveis, utilizou-se o de oportunidades de emprego cumulativas atingidas por transporte público em até 60 minutos. O índice é calculado a partir de dados geolocalizados de postos de trabalho e de dados GTFS⁸ disponibilizados pelas autoridades responsáveis pelo transporte público, conforme Pereira *et al.* (2019).

O objetivo é, seguindo a sugestão de Higgins e Kanaroglou (2016), inferir a vantagem locacional de um imóvel mais diretamente, haja vista que as proxies adotadas, como a distância às estações, nem sempre são capazes de captar a acessibilidade. Outro motivo para inserir o índice é a possibilidade de substituir a variável de distância ao CBD – ou às múltiplas centralidades – e, desta forma, diminuir as imprecisões derivadas da simplificação da estrutura espacial da cidade em pontos atômicos de concentração de empregos.

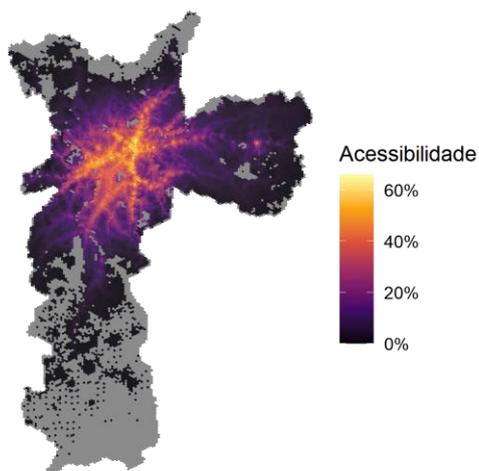
A Figura 1 abaixo representa o índice calculado: a área da cidade foi dividida em uma grade hexagonal, cada hexágono correspondente a poucos quarteirões (Pereira *et al.*, 2019). A imagem evidencia que as regiões centrais de São Paulo estão melhor posicionadas em relação à acessibilidade. O ponto amarelo isolado a leste é a estação multimodal de Itaquera, a linha vertical no centro do mapa é correspondente à linha 1 do Metrô, o eixo no sentido leste-centro equivale ao corredor das linhas 3, 11 e 12 e o eixo centro-sudoeste equivale à linha 4.

⁶ Os crimes considerados foram homicídio doloso, lesão corporal seguida de morte, estupro e latrocínio.

⁷ Para uma definição completa, ver (“Sistema de Consulta do Mapa Digital da Cidade de São Paulo”, [s.d.])

⁸ General Transit Feed Specification (GTFS), ou especificação geral de feed de trânsito, é um formato comum para especificação de itinerários de transporte público usado internacionalmente (Google, 2021).

Figura 1: índice de oportunidades cumulativas.
 Proporção de trabalhos acessíveis
 por transporte público em menos de 60 min.



Fonte: Pereira et al. (2019)

5.2 Análise exploratória dos dados

A base de dados de imóveis conta com diversas informações a respeito dos imóveis. O banco de dados foi filtrado para incluir apenas os imóveis disponíveis para aluguel. Adicionalmente, foi necessário remover imóveis situados no mesmo edifício, o que era impeditivo para o cálculo da matriz de pesos espaciais, bem como uma observação que se encontrava em um ponto da malha hexagonal do índice de acessibilidade cujo valor não estava disponível, o que impedia o uso da variável. O resultado foi um conjunto de 1.260 observações. A Tabela 2 abaixo resume as estatísticas das variáveis utilizadas do banco de dados dos imóveis.

Tabela 2: estatísticas descritivas do banco de dados de imóveis (continua)

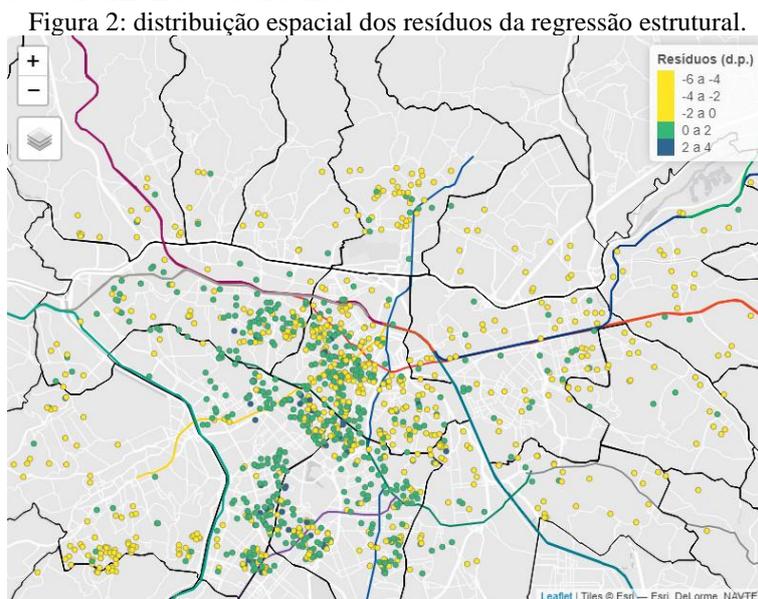
Variável	significado	Mín.	Q1	Mediana	Média	Q3	Máx.
<i>aluguel</i>	Aluguel, R\$	670	1850	2975	3860.40	4999	19000
<i>andar</i>	Nº do andar	0	3	6	7.13	10	30
<i>arcond</i>	Ar-condicionado	0	0	0	0.24	0	1
<i>area</i>	Área (m ²)	12	54	80	106.45	138	450
<i>armqua</i>	Armários no quarto	0	1	1	0.84	1	1
<i>gas_cano</i>	Gás encanado	0	1	1	0.77	1	1
<i>gas_chuv</i>	Chuveiro a gás	0	0	0	0.47	1	1
<i>mobil</i>	Mobiliado	0	0	0	0.30	1	1
<i>novo_ref</i>	Novo ou reformado	0	0	0	0.31	1	1
<i>porteiro</i>	Portaria 24h	0	1	1	0.91	1	1
<i>quadra</i>	Quadra de esportes no ed.	0	0	0	0.33	1	1
<i>sauna</i>	Sauna no edifício	0	0	0	0.24	0	1
<i>suites</i>	Nº de suítes	0	0	1	0.94	1	4
<i>tipo</i>	Tipo do imóvel	1	2	2	1.92	2	2

Fonte: elaboração própria

Em razão da abundância de informações disponíveis no banco de dados, diversas regressões foram estimadas com diferentes combinações das variáveis estruturais, até que se chegasse a um modelo parcimonioso para a regressão estrutural:

$$\begin{aligned} \log(\text{aluguel}) = & \log(\text{area}) + \text{suites} + \text{andar} + \text{tipo} + \text{arcond} \\ & + \text{armqua} + \text{gas_chuv} + \text{mobil} + \text{novo_ref} + \text{gas_cano} \\ & + \text{quadra} + \text{sauna}. \end{aligned} \quad (8)$$

Os coeficientes estimados não são reportados pois essa equação intermediária tem por fim apenas permitir a visualização das externalidades urbanas.



Fonte: elaboração própria usando pacote *tmap* do software R.

A Figura 2 acima representa a distribuição espacial dos imóveis sobreposta nos mapas da cidade e das linhas do Metrô e da CPTM. A cor indica o resíduo da regressão estrutural, em desvios padrão: quanto maior o desvio, maior o impacto esperado das amenidades urbanas sobre o valor dos aluguéis. À primeira vista, é possível notar a concentração das amostras em uma área aproximadamente retangular, centrada na Praça da Sé e indo de Santana a Bosque da Saúde, no sentido norte-sul, e de Jardim Ester Yolanda a Ermelino Matarazzo, no sentido leste-oeste. A densidade de imóveis na base de dados é menor nas regiões mais afastadas e tem uma grande concentração nas regiões mais centrais, sobretudo nas subprefeituras da Sé, da Lapa, de Pinheiros e da Vila Mariana.

O mapa indica um padrão semelhante ao do estudo de Hermann e Haddad (2005) para a mesma cidade: as regiões mais nobres apresentam resíduos majoritariamente positivos, ao passo em que nas amostras mais periféricas predominam resíduos negativos. Dois exemplos na análise são o cluster de resíduos negativos próximo à comunidade de Paraisópolis (classificada como ZEIS 1), no canto inferior esquerdo, e o cluster de resíduos positivos na subprefeitura de Pinheiros: o cálculo I de Moran pode auxiliar a verificar a existência de um padrão na cidade como um todo. Já a subprefeitura da Sé tem um padrão diverso: a região engloba tanto bairros nobres, como Higienópolis e Pacaembu, quanto bairros degradados, como Campos Elíseos, além do centro tradicional da cidade. A distribuição dos resíduos reflete essa característica, como se pode ver no mapa.

Em relação às variáveis de vizinhança que se pretende incluir, a análise visual indica ocorrência de resíduos negativos próximos às ZEIS dos três tipos. Essa relação não é tão clara quanto aos parques e unidades de conservação, haja vista que estes se encontram, em sua maioria, afastados das amostras. A exceção é o parque do Ibirapuera, visível na Figura 2, ao redor do qual se encontram vários imóveis com resíduos positivos. Similarmente, não é possível traçar uma relação visual explícita com os perímetros ambientais.

A respeito das variáveis de acessibilidade, cabe observar que os resíduos também são maiores para imóveis mais próximos às linhas do Metrô e da CPTM. Em relação à estrutura espacial, não parece haver um gradiente muito claro a partir da Praça da Sé, centro tradicional da cidade. Na realidade, a amostra sugere que as possíveis externalidades locais se concentram nas subprefeituras de Pinheiros e Vila Mariana, para onde se deslocou o centro financeiro da cidade nas últimas décadas. A Figura 2 vai na mesma direção, apontando que as regiões com maior índice de acessibilidade, frequentemente, têm maior incidência de resíduos positivos na regressão das variáveis estruturais.

5.3 Resultados

Foram analisadas oito regressões pelo método MQO. O I de Moran foi calculado usando uma matriz de pesos espaciais para os três vizinhos mais próximos. O índice resultou em uma estatística positiva e significativa em todos os cenários: isso indica que há dependência espacial positiva, implicando na modelagem espacial. A Tabela 3 abaixo contém as estatísticas calculadas.

Tabela 3: estatística I de Moran calculada

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I de Moran	0,0536***	0,0549***	0,0482***	0,0387***	0,0318***	0,0099***	0,0365***	0,0443***

Fonte: elaboração própria

Em seguida, adotou-se o procedimento de Tyszler (2006) para especificação do modelo: o SAC foi estimado pelo método da Máxima Verossimilhança e ambos os coeficientes de defasagem na variável dependente (ρ) e autocorrelação nos erros (λ) se mostraram significativos. A estimação do SAR também retornou valores significativos para ρ , fornecendo mais evidências para o uso do modelo completo em comparação com as versões SAR e SEM. As estatísticas para esses parâmetros se encontram na Tabela 4 e na Tabela 5 abaixo.

Tabela 4: parâmetros do modelo SAC estimado por MV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ	0.2761	0.3473	0.0474	0.2945	0.3541	-0.0106	0.3477	0.3475
p-valor	0.0008	0.0000	0.6878	0.0002	0.0000	0.9317	0.0000	0.0000
λ	0.6497	0.5877	0.7360	0.5863	0.4654	0.7898	0.5432	0.5385
p-valor	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000	0.0000

Fonte: elaboração própria.

Tabela 5: parâmetros do modelo SAR estimado por MV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ρ	0.4229	0.4416	0.3800	0.3977	0.4190	0.3666	0.4121	0.3820
p-valor	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Fonte: elaboração própria.

Como os modelos não apresentaram resíduos homocedásticos, o modelo SAC foi também estimado pelo método dos mínimos quadrados espaciais generalizados (GSTSLS): as medidas de impacto encontradas foram similares às aquelas obtidas por MV.

O ANEXO I contém a Tabela 6, com as medidas de impacto para os modelos SAC estimados por GSTSLS. Houve maior volatilidade entre os diferentes cenários do que em comparação com os resultados por MQO e sem dependência espacial, todavia, a significância estatística da maioria das variáveis se manteve e o seu sinal também. As maiores diferenças foram nas variáveis de acessibilidade.

As variáveis estruturais apresentaram resultados em linha com a literatura. Cabe, porém, destacar que algumas variáveis binárias, como as que avaliam a presença de ar-condicionado e chuveiro a gás, estejam atuando como proxy para outras características de imóveis de alto padrão. Sob essa ótica, os impactos indiretos dessas variáveis parecem refletir não exatamente uma apreciação do aluguel quando o apartamento vizinho instala ar-condicionado ou chuveiro a gás, mas sim, o efeito de transbordamento que se vê nos bairros quando construções de padrão mais elevado valorizam toda a região.

Na categoria das variáveis ambientais, os resultados foram diversos. O coeficiente de *dist_conserva* indica que cada quilômetro de distância da unidade de conservação ambiental mais próxima reduz o aluguel entre 2,23% e 5,12% em efeitos diretos. No entanto, é necessário cautela ao interpretar essa variável: as unidades de conservação da amostra se encontram majoritariamente nas bordas da cidade, afastadas de praticamente todas as amostras: é possível que a medida não reflita exatamente a amenidade dessas unidades sob o mercado imobiliário. O afastamento em um quilômetro das ZEIS 5 torna os aluguéis de 6,85% a 11,06% mais altos. É pertinente observar que as ZEIS 5 são fragmentadas (por serem ao nível do lote) e tem seu local de ocorrência mais limitado, sobretudo no centro da cidade e a leste. O sinal e a intensidade de ambas as ZEIS está alinhado com a intuição econômica, bem como com os resultados de Seabra, Silveira Neto e Menezes (2016) para o Recife, no caso das ZEIS 1. A variável *log(crimes)* indica que, na média, aumentar em um ponto percentual o número de vítimas em crimes hediondos implica em uma redução de 0,20% no aluguel em efeitos totais, variando de 0,16% no modelo (3) a 0,22% no modelo (7). Essas estimativas foram significativamente menores nos modelos em que a dependência espacial não foi modelada, o que reforça a pertinência de se estimá-la corretamente quando o interesse é analisar variáveis geolocalizadas.

Em relação às variáveis de acessibilidade, os resultados indicam que a proximidade às estações do Metrô é valorizada pelo mercado imobiliário, preservando sua significância estatística. Os impactos são maiores nas regressões (1) a (3), que não consideram o índice de acessibilidade: quando este é introduzido, o efeito de *dist_metro* reduz em cerca de 25% comparado com as situações anteriores. A variável tem seus maiores valores no cenário (1), considerando a estrutura duocêntrica: os impactos totais são de 6,26% por via direta, representando a redução no aluguel a cada quilômetro em que o imóvel se afasta de uma estação, e de 4,22% por efeitos indiretos, como o efeito de transbordamento de valorização (desvalorização) da vizinhança em geral pela proximidade (distância) com o equipamento. Os resultados obtidos para todos os modelos SAC foram consideravelmente maiores do que nos não-espaciais e, ainda, com menor variabilidade entre os modelos de cada método. Além da diferença em magnitude, a introdução do índice de acessibilidade leva à perda de significância da variável *dist_metro* no modelo (5) estimado por MQO, o que não se verifica nos modelos espaciais. Além disso, a distinção entre os impactos diretos e indiretos permite analisar os efeitos de transbordamento das estações de metrô, o que não é possível se a dependência espacial não for incorporada.

Já a distância às estações da CPTM não teve impacto significativo no valor do aluguel. Cabe considerar que como os dados utilizados só estavam disponíveis para a capital. Em São Paulo, há maior oferta de outros modos de transporte do que no resto da região metropolitana. Assim, a rede da CPTM pode não representar um ganho de acessibilidade considerável na capital, enquanto nas cidades da região metropolitana, é provável que a CPTM represente um diferencial de acessibilidade maior (embora não o possamos demonstrar neste trabalho). Mesmo a estatística tendo sido rejeitada, o racional por trás do sinal é similar ao encontrado na literatura: características da rede da CPTM, como a estrutura totalmente em superfície e estações com menor integração ao nível da rua, podem representar externalidades negativas como o barulho.

As três variáveis de distância ao CBD indicaram a ocorrência do gradiente de preços preconizado pela teoria nos cenários em que a variável de acessibilidade *cmatt60* não foi incluída. A regressão (1), em que foi considerada a distância ao CBD mais próximo (modelo duocêntrico), apresentou as maiores medidas de impacto: cada quilômetro de distância reduz o aluguel em 4,07%, em efeitos totais. Decompondo os efeitos, a distância ao CBD mais próximo tem um peso de 2,30% no aluguel de forma direta, ao passo em que estar situado em uma vizinhança afastada provoca um efeito indireto – pela redução nos valores vizinhos – de 1,77%. Os cenários monocêntricos tiveram valores parecidos, sendo as medidas de impacto total e indireto maiores em módulo quando se considera que o centro da capital se mantém na Sé (2), enquanto o módulo do impacto direto é maior no modelo (3), centrado na Faria Lima.

A inclusão da medida de acessibilidade cumulativa (*cmatt60*), a partir do modelo (4), se revelou significativa em todos os cenários. Como implicação, as variáveis de distância ao CBD e às estações obtiveram resultados menores do que antes da inclusão de *cmatt60*. No entanto, enquanto *dist_metro* manteve-se relevante nas oito alternativas, as variáveis *dist_cbd*, *dist_se* e *dist_farialima* perderam sua significância nos cenários (4) a (8). Por um lado, é possível que as variáveis sejam tão altamente

correlacionadas que se torna difícil distingui-las, o que ocasionaria a perda de significância das variáveis de distância. Por outro lado, a análise dos fatores de inflação da variância (FIV) feita para os modelos estimados por MQO não indicou problemas de multicolinearidade elevada. Os valores obtidos ficaram entre 1 e 3,5 para a maior parte das variáveis: a exceção foi a variável de distância do modelo duocêntrico (*dist_cbd*), cujo FIV foi de 4,5 no modelo (1) e de 6,5 no modelo (4), justamente aquele em que se incorpora o índice *cmatt60*, mas ainda não preocupante.

A diferença entre incorporar ou não outras medidas de acessibilidade é muito grande para as estimativas de *cmatt60*, cujo impacto a total quase dobra entre as alternativas (4) e (8). No primeiro cenário – quando a estrutura duocêntrica é considerada e o índice é incorporado – a cada dez pontos percentuais adicionais de empregos acessíveis em 60 minutos por transporte público, o aluguel torna-se 3,09% maior por efeitos diretos e 2,36% maior por efeitos indiretos. No cenário (7), quando *cmatt60* substitui as medidas de centralidade, mas *dist_metro* e *dist_cptm* estão presentes (apenas a primeira significativa), a mesma variação no índice de acessibilidade implica em aluguéis 7,13% maiores, sendo 3,83% por efeitos diretos e 3,30% indiretamente, por situar-se em uma vizinhança com boa acessibilidade. Quando, porém, todas as demais variáveis de acessibilidade são removidas, o diferencial indicado por *cmatt60* chega a representar um prêmio de 10,80% no valor do aluguel, 5,97% por efeitos diretos e 4,83% por efeitos indiretos.

6 CONSIDERAÇÕES FINAIS

Os modelos hedônicos analisados permitiram identificar que os principais atributos do mercado imobiliário seguem a intuição econômica, preservando a direção dos efeitos, ainda que com variações nos valores, em razão de os diferentes modelos partirem de pressupostos teóricos distintos. Embora a separação entre os aspectos de vizinhança e os unicamente relativos à acessibilidade seja sutil, os resultados dão fortes indícios de que a infraestrutura de transporte coletivo implica em um prêmio de mercado. A relação entre valor do aluguel e distância às estações de metrô se preservou em todos os modelos de forma significativa e positiva: isso fornece ainda mais evidências para as externalidades positivas geradas pelo investimento em mobilidade urbana, na medida em que o atributo é valorizado pelos indivíduos no mercado imobiliário.

A inclusão do índice de acessibilidade cumulativa como variável explicativa no modelo e a consecutiva perda de significância das variáveis de distância ao CBD indicam que a nova medida pode substituir essas variáveis e, assim, aproximar a análise empírica da economia urbana da estrutura espacial observada na prática, relaxando o pressuposto de centros de emprego atômicos. Ao mesmo tempo, a persistência da significância da variável de distância às estações do Metrô mesmo com a inclusão do índice de acessibilidade indica que as variáveis podem ser usadas de forma complementar.

Alguns fatores que contribuíram para a especificação dos modelos analisados foram a abundância de informações estruturais dos imóveis analisados e o alto nível de informações georreferenciadas disponibilizadas pelo município: o controle dessas variáveis torna a relação de causalidade entre acessibilidade e valor do aluguel mais precisa. Além disso, a incorporação da dependência espacial nos modelos permitiu corrigir o viés da estimação pelo modelo clássico e, ainda, distinguir os efeitos diretos dos efeitos de transbordamento, fornecendo maiores evidências a respeito das dinâmicas econômico-espaciais em curso. A estimação por dois estágios (GSTSLS) permitiu, ainda, obter estimadores robustos à heterocedasticidade.

Por outro lado, a escolha por utilizar uma base de anúncios on-line trouxe duas implicações: a limitação espacial da amostra, escassa nas periferias, e a ausência de um valor de equilíbrio de mercado, como seria em uma base de transações. Outro limitador é a natureza de corte transversal das observações, de forma que as estimativas refletem o valor estático dos atributos em apenas um período de tempo.

Expansões futuras que enriqueceriam a análise são a obtenção de dados para diferentes períodos de tempo, permitindo elaborar um painel e estimar, por exemplo, a mudança na valoração dos atributos frente às recentes expansões da rede metroviária; ou, ainda, analisar como as expansões afetam o mercado imobiliário durante a obra. Por fim, a pandemia em curso de Covid-19 catalisou um processo já em andamento de mudança nas interações de indivíduos e empresas com o espaço, por exemplo, com a adesão em massa ao comércio on-line e ao teletrabalho. Nesse cenário incerto de mudança nas dinâmicas

territoriais, os modelos hedônicos associados a métodos de econometria espacial e dados em painel podem ajudar a compreender as tendências, ao apontar quais atributos ganharão e quais perderão valor.

7 REFERÊNCIAS

ALMEIDA, E. **Econometria Espacial Aplicada**. 1. ed. Campinas: Alínea, 2012.

ATKINSON-PALOMBO, C. Comparing the Capitalisation Benefits of Light-rail Transit and Overlay Zoning for Single-family Houses and Condos by Neighbourhood Type in Metropolitan Phoenix, Arizona. **Urban Studies**, v. 47, n. 11, p. 2409–2426, out. 2010.

BAUMONT, C. **Spatial effects in housing price models : do housing prices capitalize urban development policies in the agglomeration of Dijon (1999) ?** [s.l.] Laboratoire d'économie et de gestion (LEG), maio 2004. Disponível em: <<https://hal.archives-ouvertes.fr/hal-01525664>>. Acesso em: 13 jun. 2021.

BERTAUD, A. **Order without design: how markets shape cities**. Cambridge, Massachusetts: MIT Press, 2018.

BERTAUD, A.; MALPEZZI, S. The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition. p. 103, 17 dez. 2003.

BIVAND, R.; MILLO, G.; PIRAS, G. A review of software for spatial econometrics in R. **Mathematics**, v. 9, n. 11, 2021.

BIVAND, R.; WONG, D. W. S. Comparing implementations of global and local indicators of spatial association. **TEST**, v. 27, n. 3, p. 716–748, 2018.

BNDES SETORIAL. BNDES Setorial, n. 41, mar. 2015. **BNDES Setorial**, p. 396, mar. 2015.

BRASIL. 8072. LEI Nº 8.072, DE 25 DE JULHO DE 1990. . 25 jul. 1990.

____. 10,257. LEI Nº 10.257, DE 10 DE JULHO DE 2001. Citation Key: brasil2001. 10 jul. 2001.

BRUECKNER, J. K. Chapter 20 The structure of urban equilibria: A unified treatment of the muth-mills model. *Em: Handbook of Regional and Urban Economics*. [s.l.] Elsevier, 1987. v. 2p. 821–845.

____. **Lectures on urban economics**. Cambridge, Mass: MIT Press, 2011.

CMSP. **Linha do tempo – Metrô 50 anos**. Disponível em: <<http://50anos.metrosp.com.br/index.php/linha-do-tempo/>>. Acesso em: 11 abr. 2021.

____. **Metrô São Paulo | Quem somos - saiba sobre transporte metropolitano**. Disponível em: <<http://www.metro.sp.gov.br/metro/institucional/quem-somos/index.aspx>>. Acesso em: 11 abr. 2021.

DUNCAN, M. Comparing Rail Transit Capitalization Benefits for Single-Family and Condominium Units in San Diego, California. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2067, n. 1, p. 120–130, jan. 2008.

FUJITA, M.; OGAWA, H. Multiple equilibrium and structural transition of non-monocentric urban configurations. p. 36, 1982.

GOOGLE. **Visão geral da GTFS estática | Google Transit estático**. Disponível em: <<https://developers.google.com/transit/gtfs?hl=pt-br>>. Acesso em: 7 set. 2021.

HERMANN, B. M.; HADDAD, E. A. Mercado imobiliário e amenidades urbanas: a view through the window. **Estudos Econômicos (São Paulo)**, v. 35, n. 2, p. 237–269, jun. 2005.

- HIGGINS, C. D.; KANAROGLOU, P. S. Forty years of modelling rapid transit's land value uplift in North America: moving beyond the tip of the iceberg. **Transport Reviews**, v. 36, n. 5, p. 610–634, 2 set. 2016.
- KELEJIAN, H.; PRUCHA, I. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating A Spatial Autoregressive Model with Autoregressive Disturbances. **The Journal of Real Estate Finance and Economics**, v. 17, p. 99–121, 1 fev. 1998.
- LESAGE, J. P.; PACE, R. K. **Introduction to spatial econometrics**. Boca Raton: CRC Press, 2009.
- MALPEZZI, S. Hedonic Pricing Models: A Selective and Applied Review. *Em*: O'SULLIVAN, T.; GIBB, K. (Eds.). **Housing Economics and Public Policy**. Oxford, UK: Blackwell Science Ltd, 2002. p. 67–89.
- MILLS, E. S. An Aggregative Model of Resource Allocation in a Metropolitan Area. **The American Economic Review**, v. 57, n. 2, p. 197–210, 1967.
- MORENCY-LAVOIE, F. **Von Thünen on the Bybanen**. Bergen: Norwegian School of Economics, 18 set. 2015.
- PÁEZ, A.; SCOTT, D. M.; MORENCY, C. Measuring accessibility: positive and normative implementations of various accessibility indicators. **Journal of Transport Geography**, v. 25, p. 141–153, nov. 2012.
- PEBESMA, E. Simple features for r: Standardized support for spatial vector data. **The R Journal**, v. 10, n. 1, p. 439–446, 2018.
- PEREIRA, R. H. M. *et al.* Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras – 2019. <http://www.ipea.gov.br>, 2019.
- PREFEITURA DO MUNICÍPIO DE SÃO PAULO. **Coefficiente de Aproveitamento**. Site governamental. Disponível em: <<https://gestaourbana.prefeitura.sp.gov.br/coeficiente-de-aproveitamento-ca/>>. Acesso em: 6 jul. 2022.
- PREFEITURA DO RECIFE. **Quais são os tipos de ZEIS?** Disponível em: <<https://licenciamento.recife.pe.gov.br/node/920>>. Acesso em: 8 jun. 2021.
- R FOUNDATION. **R: The R Project for Statistical Computing**. Disponível em: <<https://www.r-project.org/>>. Acesso em: 13 jun. 2021.
- SEABRA, D. M. DA S.; SILVEIRA NETO, R. D. M.; MENEZES, T. A. DE. Urban amenities and Real State Value: An Empirical Analysis for Recife City. **Economia Aplicada**, v. 20, n. 1, p. 143, 7 mar. 2016.
- SHOUP, D. C. Is under-investment in public infrastructure an anomaly? *Em*: **Methodology for Land and Housing Market Analysis**. London: Taylor & Francis e-Library, 2003. .
- Sistema de Consulta do Mapa Digital da Cidade de São Paulo**. Disponível em: <http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx>. Acesso em: 17 maio. 2021.
- SSP. **Dados Estatísticos do Estado de São Paulo**. Site governamental. Disponível em: <<http://www.ssp.sp.gov.br/Estatistica/Pesquisa.aspx>>. Acesso em: 31 out. 2021.
- SUZUKI, H. *et al.* **Financing Transit-Oriented Development with Land Values: Adapting Land Value Capture in Developing Countries**. [s.l.] The World Bank, 2015.
- TENNEKES, M. tmap: Thematic maps in R. **Journal of Statistical Software**, v. 84, n. 6, p. 1–39, 2018.

TYSZLER, M. **Econometria espacial: discutindo medidas para a matriz de ponderação espacial**. São Paulo: Escola de Administração de Empresas de São Paulo, 2006.

WICKHAM, H. **ggplot2: Elegant graphics for data analysis**. [s.l.] Springer-Verlag New York, 2016.

____. Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019.

WOOLDRIDGE, J. **Introdução à econometria: uma abordagem moderna**. 6. ed. São Paulo: Cengage Learning, 2016.

ANEXO I
Regressão

Tabela 6: medidas de impacto do modelo SAC

(continua)

	(1)			(2)			(3)			(4)		
	Direto	Indireto	Total	Direto	Indireto	Total	Direto	Indireto	Total	Direto	Indireto	Total
log(areas)	0.5154***	0.3687***	0.8841***	0.5172***	0.4503***	0.9675***	0.5091***	0.2514*	0.7605***	0.5136***	0.3811***	0.8947***
banh	0.0736***	0.0526**	0.1262***	0.0714***	0.0621***	0.1335***	0.0742***	0.0367*	0.1109***	0.0752***	0.0558***	0.131***
andar	0.0061***	0.0043*	0.0104**	0.006***	0.0052**	0.0112**	0.0061***	0.003*	0.0091**	0.006***	0.0044*	0.0104**
arcond	0.1499***	0.1072**	0.2571***	0.1513***	0.1317***	0.2831***	0.1486***	0.0734*	0.222***	0.1487***	0.1103***	0.2591***
gas_chuv	0.0921***	0.0659**	0.1579***	0.0917***	0.0798**	0.1715***	0.0901***	0.0445*	0.1346***	0.0921***	0.0683**	0.1604***
mobil	0.1141***	0.0816**	0.1958***	0.1147***	0.0998***	0.2145***	0.1132***	0.0559*	0.1691***	0.1148***	0.0852***	0.2000***
novo_ref	0.0899***	0.0643**	0.1541***	0.0908***	0.079**	0.1698***	0.0895***	0.0442*	0.1336***	0.0883***	0.0655**	0.1538***
academia	0.0843***	0.0603**	0.1447***	0.0854***	0.0744**	0.1598***	0.0827***	0.0409*	0.1236***	0.0833***	0.0618**	0.1451***
sauna	0.0784**	0.0561*	0.1345**	0.0795**	0.0692*	0.1486**	0.0783**	0.0387'	0.1170*	0.078**	0.0579*	0.136**
dist_conserva	-0.0307***	-0.0219**	-0.0526***	-0.0233***	-0.0203**	-0.0437***	-0.0247***	-0.0122*	-0.0369***	-0.0268***	-0.0199**	-0.0467***
dist_zeis1	0.0222	0.0159	0.0381	0.025	0.0217	0.0467	0.0304'	0.015	0.0453'	0.0263	0.0195	0.0458
dist_zeis5	0.0492***	0.0352***	0.0843***	0.0561***	0.0488***	0.1049***	0.0444**	0.0219*	0.0663**	0.0469***	0.0348***	0.0816***
log(crime)	-0.1136***	-0.0813**	-0.1949***	-0.1113***	-0.0969**	-0.2083***	-0.1105***	-0.0546*	-0.1651***	-0.1167***	-0.0866**	-0.2032***
dist_metro	-0.0646***	-0.0462**	-0.1107***	-0.0541***	-0.0471**	-0.1013***	-0.0736***	-0.0363*	-0.1099***	-0.0474***	-0.0352**	-0.0826***
dist_cptm	0.0141	0.0101	0.0243	0.0057	0.005	0.0107	0.0031	0.0015	0.0047	0.0124	0.0092	0.0216
dist_cbd	-0.0233*	-0.0166*	-0.0399*							-0.0125	-0.0093	-0.0218
dist_se				-0.016*	-0.0139*	-0.0299*						
dist_farialima							-0.0185*	-0.0091*	-0.0277*			
cmatt60										0.0304*	0.0226*	0.053*

Nota: *** significante a 0,1%, ** significante a 1%, * significante a 5%, ' significante a 10%.

Fonte: elaboração própria.

Tabela 4: medidas de impacto do modelo SAC

(conclusão)

	(5)			(6)			(7)			(8)		
	Direto	Indireto	Total	Direto	Indireto	Total	Direto	Indireto	Total	Direto	Indireto	Total
log(areal)	0.5158***	0.4079***	0.9237***	0.5103***	0.3314**	0.8417***	0.5131***	0.4273***	0.9404***	0.5143***	0.3959***	0.9102***
banh	0.0738***	0.0583***	0.1321***	0.0757***	0.0491**	0.1248***	0.0751***	0.0625***	0.1376***	0.0768***	0.0591***	0.1359***
andar	0.0059***	0.0047**	0.0106**	0.006***	0.0039*	0.0098**	0.0059***	0.0049**	0.0108**	0.0058**	0.0045*	0.0103**
arcond	0.1496***	0.1183***	0.2679***	0.1482***	0.0962**	0.2445***	0.149***	0.1241***	0.2731***	0.1477***	0.1137***	0.2614***
gas_chuv	0.0922***	0.0729***	0.165***	0.091***	0.0591*	0.1501***	0.0908***	0.0756***	0.1664***	0.0907***	0.0698**	0.1606***
mobil	0.1151***	0.0911***	0.2062***	0.1145***	0.0743**	0.1888***	0.1153***	0.096***	0.2114***	0.1186***	0.0913***	0.2099***
novo_ref	0.0891***	0.0704**	0.1595***	0.088***	0.0572*	0.1452***	0.0887***	0.0738**	0.1625***	0.087***	0.0669**	0.1539***
academia	0.085***	0.0672**	0.1523***	0.0823***	0.0534*	0.1357***	0.0828***	0.0689**	0.1517***	0.0788***	0.0607**	0.1395***
sauna	0.0795**	0.0629**	0.1424**	0.0779**	0.0506*	0.1284**	0.0779**	0.0649*	0.1429**	0.0807**	0.0621**	0.1428**
dist_conserva	-0.0235***	-0.0186**	-0.0422***	-0.0232***	-0.0151*	-0.0382***	-0.0207***	-0.0173**	-0.038***	-0.0128**	-0.0098*	-0.0226*
dist_zeis1	0.0239	0.0189	0.0429	0.0321*	0.0208'	0.0529*	0.0347*	0.0289*	0.0636*	0.0343*	0.0264*	0.0607*
dist_zeis5	0.055***	0.0435***	0.0985***	0.0438***	0.0284*	0.0722***	0.0432***	0.036***	0.0791***	0.0448***	0.0345***	0.0793***
log(crime)	-0.113***	-0.0894***	-0.2024***	-0.1174***	-0.0762*	-0.1937***	-0.1205***	-0.1004***	-0.2209***	-0.121***	-0.0931***	-0.2141***
dist_metro	-0.0374**	-0.0296*	-0.0670**	-0.0507***	-0.0329*	-0.0837***	-0.0462***	-0.0385**	-0.0847***			
dist_cptm	0.0088	0.0070	0.0157	0.0068	0.0044	0.0112	0.0058	0.0049	0.0107			
dist_cbd												
dist_se	-0.0120'	-0.0095'	-0.0216'									
dist_farialima				-0.0077	-0.005	-0.0127						
cmatt60	0.0314**	0.0248*	0.0562*	0.0335**	0.0218*	0.0553**	0.0376***	0.0313**	0.0689***	0.058***	0.0446***	0.1026***

Nota: *** **significante a 0,1%**, ** **significante a 1%**, * **significante a 5%**, ' **significante a 10%**.

Fonte: elaboração própria