

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO
MESTRADO EM ADMINISTRAÇÃO

HENRIQUE PINTO RAMOS

**Um estudo sobre a previsibilidade dos preços de imóveis de Porto Alegre -
evidências dos mercados de venda e de locação.**

Porto Alegre

Abril/2016

CIP - Catalogação na Publicação

Pinto Ramos, Henrique

Um estudo sobre a previsibilidade dos preços de imóveis de Porto Alegre - evidências dos mercados de venda e de locação. / Henrique Pinto Ramos. -- 2016. 73 f.

Orientador: Marcelo Scherer Perlin.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Escola de Administração, Programa de Pós-Graduação em Administração, Porto Alegre, BR-RS, 2016.

1. Imóveis. 2. Preços Hedônicos. 3. Econometria Espacial. 4. Previsão. 5. Model Confidence Set. I. Scherer Perlin, Marcelo, orient. II. Título.

HENRIQUE PINTO RAMOS

**Um estudo sobre a previsibilidade dos preços de imóveis de Porto Alegre –
evidências dos mercados de venda e de locação.**

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre em
Administração pelo Programa de Pós-
Graduação em Administração da Universidade
Federal do Rio Grande do Sul.

**Orientador: prof^o Marcelo Scherer Perlin,
PhD**

BANCA EXAMINADORA:

Prof. Dr. Guilherme Kirch – PPGA/UFRGS

Prof. Dr. Flávio Augusto Ziegelmann – PPGA/PPGE/UFRGS

Prof. Dr. Guilherme Ribeiro de Macêdo – PPGA/PPGE/UFRGS

Prof. Dr. Marcelo Scherer Perlin (orientador) – PPGA/UFRGS

Porto Alegre

Abril/2016

RESUMO

Esta dissertação apresenta um estudo sobre o mercado imobiliário de Porto Alegre que tem como objetivo verificar a formação de preços dos mercados de venda e de locação. Este objetivo é cumprido de duas formas: a) analisando como o tempo que um imóvel está no mercado afeta seu preço e b) comparando como diferentes modelos econométricos preveem os preços de imóveis da capital gaúcha. Para o primeiro tópico, encontram-se evidências de que os proprietários reduzem, em média, 9% o valor de locação para atrair clientes à medida que o tempo passa. Isto não se concretiza da mesma forma no mercado de venda, uma vez que há uma valorização à medida que o imóvel está no mercado. Para o segundo tópico, foram estimados cinco modelos (ingênuo, MQO, SAR, SEM e combinação) para averiguar a performance de previsão fora da amostra dos preços de imóveis. Embora as estatísticas descritivas dos erros não permitam análises conclusivas a respeito dos modelos, a estimação do *model confidence set* de Hansen *et al.* (2011) mostra o modelo ingênuo como a melhor série de previsões para os mercados de venda e de locação.

Palavras-chave: Imóveis, Preços Hedônicos, Econometria Espacial, Previsão, *Model Confidence Set*

ABSTRACT

This thesis presents a study on the real estate market of Porto Alegre in order to verify the price formation in both sales and rental markets. This objective is accomplished in two ways: a) analyzing how time on market of a property impacts its price and b) comparing how different econometric models predict real estate prices. For the first topic, I show evidence that owners reduce on average 9% of rental value of available homes in order to attract customers as time on market goes by. This pattern is not present in the sales market, in which an appreciation occurs as time passes, indicating an inflation compensation. For the second topic, five models (naive, OLS, SAR, SEM and a combination) were estimated to determine out-of-sample forecasting accuracy. Although traditional error measures do not allow to choose the best forecasting series, the model confidence set procedure of Hansen et al. (2011) shows the naïve model as being the best forecasting model.

Keywords: Real Estate, Hedonic Prices, Spatial Econometrics, Forecasting, Model Confidence Set.

ABREVIATURAS UTILIZADAS

- ADEMI** – Associação de Dirigentes de Empresas do Mercado Imobiliário
- EMBRAESP** – Empresa Brasileira de Estudos de Patrimônio de São Paulo
- IPEAD** – Instituto de Pesquisas Econômicas e Administrativas
- MCS** – *Model Confidence Set*
- MQO** – Mínimos Quadrados Ordinários
- POF** – Pesquisa de Orçamentos Familiares
- SAR** – *Spatial Autoregressive Model*
- SECOVI** – Sindicato dos Empregados em Empresas de Compra, Venda, Locação e Administração de Imóveis Residenciais e Comerciais
- SEM** – *Spatial Error Model*
- SINDUSCON** – Sindicato da Indústria da Construção Civil

ÍNDICE DE TABELAS

Tabela 1 - Descrição dos trabalhos realizados no Brasil.....	20
Tabela 2 - Matriz de dois vizinhos mais próximos	28
Tabela 3 - Variáveis utilizadas	37
Tabela 4 - Período de estimações dentro e fora da amostra	38
Tabela 5 - Estatísticas descritivas.....	48
Tabela 6 - Tempo de publicação e variação nos preços	49
Tabela 7 - Definição dos k vizinhos	51
Tabela 8 - Teste LM	52
Tabela 9 - Estimções in-sample.....	53
Tabela 10 - Erro absoluto médio (%).....	54
Tabela 11 - Erros de previsão em relação aos bairros	56
Tabela 12 - Erros de previsão por bairro	58
Tabela 13 - p-values para estimação do MCS	59

ÍNDICE DE FIGURAS

Figura 1 – Dinâmica do processo espacial	25
Figura 2 – Configuração de vizinhança por contiguidade	27
Figura 3 – Modelo SAR	40
Figura 4 – Modelo SEM	41

SUMÁRIO

1.	INTRODUÇÃO	10
2.	OBJETIVOS	14
3.	REVISÃO DA LITERATURA	15
3.1.	Preços hedônicos	15
3.2.	Tempo no mercado.....	22
3.3.	Econometria espacial	25
3.3.1.	Matrizes de dependência espacial	26
3.3.2.	Testes para identificação de dependência espacial	29
3.4.	Previsão de preços de imóveis	32
4.	DADOS E VARIÁVEIS UTILIZADAS	35
4.1.	Descrição das variáveis	36
5.	MÉTODO	38
5.1.	Impacto do tempo no mercado.....	38
5.2.	Modelos para previsão	38
5.3.	Testes de previsão	43
6.	RESULTADOS	47
6.1.	Estatísticas descritivas	47
6.2.	Impacto do tempo no mercado.....	49
6.3.	Definição da matriz de pesos espaciais.....	50
6.4.	Calibragem dos modelos de previsão	52
6.5.	Previsão fora da amostra	54
6.6.	Model Confidence Set	59
7.	CONSIDERAÇÕES FINAIS	61
7.1.	Limitações e sugestões de pesquisas futuras	62
	REFERÊNCIAS	63
	APÊNDICE	71

1. INTRODUÇÃO

Os mercados financeiros permitem o estabelecimento de contratos entre compradores e vendedores de valores mobiliários, como ações e títulos de dívidas, além de mercadorias e outros tipos de acordos. Dentre os benefícios dos mercados financeiros estão a captação de capital e a transferência de risco entre participantes, em que ambos os fatores facilitam a entrada de investidores que não teriam acesso a estas oportunidades sem a existência deste sistema. No entanto, os mercados financeiros não contêm todas as oportunidades de investimento disponíveis, havendo opções reais como projetos privados e outras possíveis aplicações para um agente com capital. Dentre estas possibilidades, o investimento em imóveis é uma alternativa de certa forma popular (EXAME), uma vez que um proprietário pode locar um imóvel para um inquilino, e/ou revende-lo após um período de valorização e realizar um ganho de capital. Especificamente no Brasil, este tipo de investimento teve um grande crescimento nesta década. Como exemplo, no Rio de Janeiro os preços se valorizaram 128% de 2008 a 2014 (ZAP IMÓVEIS), dentro de um cenário econômico que motivou a abertura de muitos Fundos de Investimentos Imobiliários para aproveitar o bom momento do mercado imobiliário. Em 2005, havia 63 fundos registrados na CVM com um patrimônio líquido de aproximadamente R\$ 2,5 bilhões. Em 2015, apesar do revés destes fundos e do mercado em geral, o patrimônio líquido girava em torno de R\$ 61 bilhões em 261 fundos (ANBIMA).

Isto mostra como o investimento em imóveis cresceu nos últimos anos, porém, ainda assim, pode-se questionar como precificar os ativos imobiliários. Isso porque, um imóvel pode ser considerado tanto como um bem de consumo quanto um ativo financeiro. Da mesma forma que estes bens oferecem utilidades a quem o possui, como a moradia e todas as amenidades que estão no entorno de sua localização, há atributos comuns a ativos financeiros, como oferta, demanda, preço, risco e retorno. Esta dupla utilidade caracteriza a complexidade na precificação do mercado de imóveis. Gonzáles e Formoso (2000) ressaltam algumas dificuldades na estimação destes preços: a falta de informação de seus agentes sobre os seus mecanismos de funcionamento e uma alta heterogeneidade das características entre os diferentes imóveis. Além desses aspectos, acredita-se que outros fatores não-financeiros como campanhas de marketing das incorporadoras e imobiliárias pode afetar os preços de imóveis (PINTO e ZAMPIER, 2013).

Entende-se que o valor do imóvel é a soma da utilidade de suas características intrínsecas, como tamanho, número de quartos, número de vagas na garagem, e de atributos ligados a vizinhança: a localização, violência, e demais características socioeconômicas da região. Embora intuitivo, este raciocínio só se tornou claro a partir dos estudos sobre bens heterogêneos de Lancaster (1966) e posterior trabalho de Rosen (1976), em que se definem preços hedônicos como os preços derivados das diferentes características e utilidades presentes em um determinado bem. A partir dos estudos de Witte *et al.* (1979), vários trabalhos surgiram para testar a importância de determinados atributos no valor de um imóvel para precificá-lo corretamente (CHESHIRE e SHEPPARD, 2004; DUBIN e SUNG, 1990; GONZÁLES E FORMOSO, 1995; NELSON, 1978). No entanto, é prática comum do mercado imobiliário a comparação dos valores de imóveis de bairros próximos para a definição do seu preço de venda, o que sugere a necessidade de estudos sobre as melhores formas de estimar estes preços (ELHORST, 2000; GONZÁLES E FORMOSO, 2000; PACE *et al.*, 1998).

No Brasil, há poucos estudos abordando esta temática, concentrados em capitais do sudeste do país como São Paulo (ALVES *et al.* 2012; FÁVERO *et al.* 2008; HERMANN e HADDAD, 2005; SANT'ANA, 2006), Rio de Janeiro (FERREIRA NETO, 2002, ROZENBAUM e MACEDO-SOARES, 2007) e Belo Horizonte (FURTADO, 2007, MACEDO, 1998). Destes, o trabalho de Sant'Ana (2006) é o único no qual se realizam previsões fora da amostra. Localmente, há poucos estudos na cidade de Porto Alegre/RS que analisam a formação de preços do mercado imobiliário local (GONZÁLES E FORMOSO, 1994) e nenhum estudo focado na previsão destes valores, tampouco na comparação entre modelos. Por isso, este trabalho tem o objetivo de contribuir para a expansão desta literatura, à medida em que são comparadas previsões de preços de imóveis fora da amostra com cinco modelos diferentes nos mercados de venda e de locação de Porto Alegre nos anos de 2012 e 2013. Foi utilizada uma base de dados representativa para venda (15081 observações) e para locação (1579 observações), não havendo conhecimento de trabalhos que abordem esta temática para ambos os mercados. Os modelos estimados consistem em uma estratégia ingênua, uma estimação por MQO, dois modelos espaciais (SAR e SEM) e uma combinação de todas as previsões. Embora as estatísticas descritivas para as previsões fora da amostra não permitam a definição de um modelo mais apropriado, a construção do *Model Confidence Set* de Hansen *et al.* (2011)

mostra que o modelo ingênuo apresenta um melhor desempenho para a previsão tanto no mercado de venda quanto no de locação.

Liu (2012) assevera que a previsão é uma das maiores aplicações para a estimação de modelos de preços de imóveis, em que possíveis clientes, construtoras, imobiliárias, sindicatos, órgãos de classe e agentes públicos podem aproveitar-se destas predições. Estes últimos podem utilizar as ferramentas deste estudo para aprimorar a definição de políticas macroeconômicas que evitem a ocorrência das chamadas bolhas imobiliárias (GLAESER *et al*, 2008; HIMMELBERG *et al*, 2005). Um estudo sobre a previsão dos preços de imóveis também pode contribuir para a literatura que estuda as altas e baixas nos preços de imóveis acima das expectativas, denominados *booms* e *busts* (AGNELLO e SCHUKENETCH, 2011; MUELLBAUER e MURPHY, 1997; MANKIW e WEIL, 1989; MILES, 2007). Para praticantes do mercado, o estudo é relevante tanto para quem demanda quanto para quem oferta imóveis. Para quem busca comprar ou locar, sejam investidores profissionais sejam moradores, é relevante conhecer o potencial de valorização de um imóvel (LIU, 2012). No lado da oferta, o estudo é pertinente para construtoras e imobiliárias uma vez que, usando os métodos descritos nesta dissertação, é possível analisar oportunidades de investimento com VPL positivo e, assim, orientar compradores. Além disso, para órgãos de classe como SINDUSCON-RS e SECOVI-RS o estudo é útil pois propicia uma ferramenta para análise sobre a valorização (ou desvalorização) do mercado imobiliário como um todo, bem como informações relevantes sobre rumo que o mercado pode tomar.

Além da previsão de preços, este é o primeiro estudo no Brasil que analisa o impacto do tempo que um imóvel está à venda ou à locação na alteração dos respectivos preços. As estatísticas descritivas mostram que um imóvel que tem seu preço reajustado durante o período de anúncio tem um corte de, em média, 9% em seu aluguel. Isto é consonante com a hipótese de demanda por liquidez dos proprietários, a medida em que, ao encontrar um inquilino, o locador reduz custos como condomínio e adquire proteção contratual contra inflação. Estas evidências não se mostram tão fortes no mercado de compra, que parece se comportar de forma a reajustar os preços para repor perdas inflacionárias.

O presente trabalho visa contribuir para a literatura de duas formas: a) expandindo o conhecimento em modelos de previsão de imóveis, através de uma comparação entre

modelos de previsão, b) entendendo o impacto do tempo no mercado na variação de preços de imóveis. Além da expansão da literatura, estas contribuições devem impactar atuantes do mercado, tanto clientes finais quanto intermediários, a ter definições mais claras sobre a formação de preços de imóveis. Para uma melhor organização, o trabalho foi dividido da seguinte forma: a seção 2 apresenta os objetivos e a seção 3 faz uma revisão da literatura necessária para compreender a formação de preços de imóveis. A seção 4 descreve os dados e variáveis utilizadas, a seção 5 apresenta os métodos utilizados, a seção 6 apresenta os resultados e a seção 7 conclui esta dissertação com as considerações finais.

2. OBJETIVOS

O objetivo geral desta dissertação será o de verificar a formação de preços de imóveis na cidade de Porto Alegre.

Para isso, delimitam-se os seguintes objetivos específicos, que contribuirão para o cumprimento do objetivo geral:

- Realizar uma análise descritiva dos dados de imóveis de Porto Alegre;
- Analisar o impacto do tempo no mercado nos preços de imóveis;
- Escolher as matrizes de ponderação que melhor retratam a dependência espacial dos dados;
- Comparar a eficiência dos modelos na previsão fora da amostra.

3. REVISÃO DA LITERATURA

Para melhor apresentação da revisão da literatura, esta seção foi subdividida em quatro partes. Na seção 3.1 será explicado o que são os modelos hedônicos e como eles se aplicam ao mercado imobiliário, bem como um breve detalhamento do que já foi produzido na literatura brasileira. A seção 3.2 expõe trabalhos e teorias a respeito do impacto do tempo no mercado no preço de um imóvel. Na seção 3.3, será feita uma revisão bibliográfica sobre econometria espacial e, por fim, na 3.4 será mostrado como a literatura vêm trabalhando a previsão de preços de imóveis.

3.1. Preços hedônicos

O mercado imobiliário caracteriza-se pela complexidade dos bens transacionados (GONZÁLEZ E FORMOSO, 2000). Da mesma forma que os consumidores obtêm benefícios pelos atributos intrínsecos ao próprio imóvel (tamanho, número de quartos e garagens), o consumidor também estará negociando de acordo com um mercado implícito de amenidades urbanas, sendo estes atributos que não estão relacionados ao imóvel em si, mas são adquiridos no “pacote”. Por exemplo, localização, acessibilidade e criminalidade do bairro são características do entorno do imóvel que são levadas em conta na precificação deste tipo de ativo (LANCASTER, 1966; ROSEN, 1976; WITTE *et al*, 1979). Essa visão é definida inicialmente por Lancaster (1966), asseverando que os consumidores buscam a utilidade nas características de um bem, não no bem em si. O mesmo autor cita que as características de um bem na sua totalidade podem ter uma utilidade diferente do que os mesmos atributos de forma separada. Trazendo isso para a realidade do mercado imobiliário, as definições de Lancaster ajudam a compreender como um imóvel com a mesma metragem e mesmo número de quartos pode ter um preço diferente de acordo com, por exemplo, a localização ou a criminalidade na vizinhança. Dado isso, a presença, ou ausência, de amenidades urbanas pode interferir no bem-estar de quem demanda imóveis e, com isso, no valor que estarão dispostos a pagar por estes ativos (HERMANN e HADDAD, 2005).

Rosen (1976) aprofunda o trabalho de Lancaster (1966) e define preços hedônicos como os preços implícitos de atributos que são ofertados ao mercado a partir dos valores dos

produtos finais. Exemplificando, a estimação dos preços hedônicos visa mostrar como um atributo, como a violência do bairro, altera os preços de venda de imóveis com diferentes níveis de criminalidade em diferentes bairros. Basicamente, trata-se do “preço” que se está disposto a pagar pela violência daquele bairro, mesmo que não se trate de uma unidade mensurável explicitamente por um indivíduo.

Em suma, os modelos hedônicos focam em mercados em que um bem genérico pode ter diversos atributos ligados a ele, sendo relevante investigar como o preço deste bem varia em diferentes configurações dos pacotes de atributos (EPPLÉ, 1987). Witte *et al.* (1979) citam que esta heterogeneidade dos atributos dos imóveis é um grande obstáculo na análise do mercado imobiliário. Isso porque, a estimação de equações de equilíbrio entre oferta e demanda não é tão facilmente calculada, uma vez que as características dos imóveis abarcam atributos intrínsecos e as amenidades urbanas. Essa complexidade é descrita por Hermann e Haddad (2005):

Embora possamos supor que exista uma oferta, uma demanda e um preço de equilíbrio para as diferentes amenidades, não podemos coletar esses dados diretamente no mercado. Não existe explicitamente um mercado de compra e venda de trânsito, ou um mercado de compra e venda de poluição (p. 238).

Por esta razão, a definição de quais pacotes de atributos entrarão no modelo estrutural que estimará os preços de imóveis é uma questão importante na literatura. O intuito destes estudos é definir o valor marginal das características intrínsecas e extrínsecas do imóvel. Alguns artigos encontraram resultados relevantes para a importância da proximidade de centros históricos e comerciais (FURTADO, 2007), a poluição do ar (CHONG *et al.*, 2003 ; NELSON, 1978), a proximidade de escolas (BRASINGTON e HAURIN, 2006; CHESHIRE e SHEPPARD, 2004; CLAPP *et al.*, 2008; GIBBONS e MACHIN, 2008), a criminalidade do bairro (GIBBONS e MACHIN, 2008; RONDON e ANDRADE, 2003), a proximidade de estações de trem (BOWES e IHLANFELDT, 2001), o barulho devido à proximidade de aeroportos (ESPEY e LOPEZ, 2000; POPE, 2008) e até mesmo a vista que os apartamentos possuem (BENSON *et al.*, 1998). Além destes trabalhos, há uma grande quantidade de *papers* que exploram modelos hedônicos sem o objetivo de testar um atributo em específico, mas sim analisar de forma conjunta as diferentes características dos imóveis e de seus entornos (ALVES *et al.*, 2011, BALA *et al.*, 2014; BALTAGI e BRESSON, 2010; DUBIN e SUNG, 1990; GONZÁLES e FORMOSO, 1995). No entanto, não há um consenso na literatura sobre quais fatores devem entrar no modelo

estrutural (MACEDO, 1998). Essa arbitrariedade na definição dos modelos também pode ser atribuída a disponibilidade de dados sobre as amenidades dos imóveis, uma vez que os dados desejados podem não estar disponíveis, ou serem de má qualidade (BUTLER, 1982).

No Brasil, há alguns estudos aplicando os modelos hedônicos, principalmente em capitais estaduais da região Sudeste. Na cidade do Rio de Janeiro, Rozenbaum e Macedo-Soares (2007) propõem um modelo para a construção de um índice de preços de imóveis baseado em lançamentos imobiliários (antes do início da construção) do bairro Barra da Tijuca. Foram consideradas variáveis explicativas hedônicas como a distância ao centro, distância ao shopping mais próximo e uma *dummy* caso o lançamento fosse próximo à Avenida das Américas (beira-mar). Embora a técnica econométrica utilizada não esteja clara, os autores encontraram um R^2 ajustado na regressão de 0,953 tendo o preço por metro quadrado como variável dependente. Na mesma cidade, Ferreira Neto (2002) também realiza um estudo com modelos hedônicos para lançamentos imobiliários utilizando variáveis de vizinhança como quantidade de hospitais no bairro, de escolas, postos de gasolina, taxa de criminalidade por bairro, além de características próprias dos imóveis. O objetivo do estudo de Ferreira Neto (2002) foi estabelecer a forma funcional que minimizaria o erro de previsão para fora da amostra¹. Foram realizadas regressões por MQO nas formas linear, semi-logarítmica e dupla logarítmica. A forma dupla logarítmica foi a que teve um menor erro de previsão, chegando a um nível aproximado de 25%.

Na cidade de Belo Horizonte, Macedo (1998) compara modelos tradicionais e modelos de econometria espacial (SAR e SEM) e encontra que a proximidade entre bairros e a adjacência entre eles é um fator importante para explicar a variação dos preços de imóveis². O autor encontra fortes evidências estatísticas sobre a necessidade de utilizar uma abordagem econométrica espacial para o caso estudado. Rondon e Andrade (2003) fazem regressões por MQO para verificar a influência da criminalidade nos aluguéis da cidade, e encontram que, dependendo do bairro, o valor de aluguel de um apartamento com as mesmas características físicas pode variar em até 27% o valor do imóvel. Neste estudo, a forma semi-logarítmica foi a escolhida para as estimações e não foram utilizadas técnicas de econometria espacial para

¹ Foram utilizadas como medidas de previsão o Erro Quadrático Médio e o Erro Absoluto Médio.

² Os modelos espaciais são detalhados nas seções 3.3 e 5.2.

averiguar a correlação espacial³. Ainda na capital mineira, Furtado (2007) testa a necessidade de estimar espacialmente um modelo de preços hedônicos. Primeiramente, o autor estima modelos via MQO sem considerar a correlação espacial entre a localização dos imóveis. Após isso, ele verifica, e confirma, a necessidade da modelagem espacial em que as equações são estimadas por um modelo SAR por Máxima Verossimilhança.

Dentre os trabalhos realizados no Brasil, a maior parte tem como objeto de estudo a cidade de São Paulo. Um *paper* relevante para a literatura local é o de Hermann e Haddad (2005), em que os autores utilizam um modelo hedônico considerando duas métricas de localização: a monocêntrica, em que é inserida nas estimações a distância do imóvel até o centro da cidade e a a duocêntrica, que considera tanto a distância do centro da cidade quanto a distância do centro comercial⁴. Estes autores realizaram testes para a existência de correlação espacial e descobriram que, à medida que vão sendo inseridas variáveis relacionadas às amenidades nas estimações, a correlação espacial é reduzida. Neste caso, os autores inserem informações do bairro como poluição do ar, quantidade relativa de áreas verdes, população moradora de favelas e distância à estação de metrô mais próxima. Este é um *insight* relevante, uma vez que, caso haja dados detalhados disponíveis, a correlação espacial pode ser mitigada, ou reduzida. Isto vai ao encontro do citado por Butler (1982) sobre a relevância de dados de qualidade para as estimações utilizando preços de imóveis.

Sant'ana (2006) realiza um estudo na cidade de São Paulo com uma base de dados de lançamentos imobiliários do ano 1985 até o ano 2001. O autor estima um modelo hedônico via MQO que considera a distância entre os empreendimentos até o centro da cidade e chega a um R² ajustado variando de 41% a 49%. No entanto, a correlação espacial não foi tratada no modelo. Em outro estudo, Fávero *et al.* (2008) usam variáveis sociodemográficas como controle para cada distrito e consideram uma ampla lista de características do imóvel e amenidades urbanas⁵. Utilizando a forma funcional logarítmica, os autores estimam a elasticidade-preço das variáveis explicativas tanto para a oferta quanto para a demanda, seguindo o trabalho de Rosen (1974). Alves *et al* (2011), também realizam um estudo em São

³ No entanto, na conclusão do artigo os autores sugerem a estimação com essas abordagens.

⁴ Para mais detalhes sobre as configurações monocêntrica e duocêntrica, ver BRUECKNER (1987) e FUJITA E OGAWA (1982).

⁵ Vide Quadros 2 a 6 de Fávero *et al* (2008). Pela contagem do autor desta dissertação, foram utilizadas 26 variáveis, entre intrínsecas e extrínsecas ao imóvel.

Paulo semelhante aos anteriores com uma base mensal de dados de imóveis dos anos de 2001 a 2008. Os autores consideram como variáveis de controle, além das características físicas, amenidades e distâncias ao centro histórico e ao centro comercial, o *spread* de juros bancários e o prazo de financiamento imobiliário. Foi encontrada significância estatística para a primeira variável, mas não para a segunda.

Em Porto Alegre, González e Formoso (1994) realizam um estudo com os preços de aluguel de 504 imóveis no ano de 1992, utilizando em seu modelo hedônico as características físicas e as distâncias ao centro e aos principais *shoppings centers* como variáveis explicativas. Os autores também realizaram regressões (via MQO) em subgrupos com imóveis de tamanho pequeno, médio e grande. Dependendo das especificações, os modelos chegaram a um coeficiente R^2 de mais de 80%. No estado do Rio Grande do Sul como um todo, Braga (2010) utiliza dados de imóveis transacionados para analisar a importância de aspectos socioeconômicos para a determinação dos respectivos valores em 290 municípios. Embora o foco da análise não tenha sido em modelos hedônicos, o estudo citado contribui para evidenciar que características socioeconômicas tem um peso na formação do preço de imóveis.

Por fim, é possível observar que, nos casos brasileiros, a preponderância é de estimações via MQO. Nota-se que a ocorrência da modelagem espacial bem como a forma funcional (linear, logarítmica ou dupla logarítmica) não encontram padrões na literatura brasileira. A fim de melhor explicitar como foram realizados os estudos utilizando modelos hedônicos no país, foi elaborada a Tabela 1 para comparar as diferentes abordagens.

Tabela 1- Descrição dos trabalhos realizados no Brasil

Autores	Localização do Estudo	Período utilizado	Observações	Fonte dos Dados de Imóveis	Métodos empregados
Rozenbaum e Macedo-Soares (2007)	Rio de Janeiro	2002 a 2005	Não cita	Ademi-RJ	Modelo hedônico sem tratamento para dependência espacial. Estimação por MQO.
Ferreira Neto (2002)	Rio de Janeiro	2000	120	Ademi-RJ	Modelo hedônico sem tratamento para dependência espacial. Estimação por MQO.
Macedo (1998)	Belo Horizonte	out-1995	53	IPEAD	Modelo hedônico com tratamento para dependência. Estimação por MQO e por Máxima Verossimilhança.
Rondon e Andrade (2003)	Belo Horizonte	abr-2002	1303	IPEAD	Modelo hedônico sem tratamento para dependência espacial. Estimação por MQO.
Furtado (2007)	Belo Horizonte	jun-2005	3196	Coleta manual em site imobiliário	Modelo hedônico com tratamento para dependência espacial. Estimação por MQO e por Máxima Verossimilhança.
Hermann e Haddad (2005)	São Paulo	1999	497	POF	Modelo hedônico com tratamento para dependência espacial. Estimação por MQO.
Sant'ana (2006)	São Paulo	jan-1985 a jan-2001	4717	EMBRAESP	Modelo hedônico sem tratamento para dependência espacial. Estimação por MQO.
Fávero et al. (2008)	São Paulo	2004	1860	Coleta manual em jornais e sites imobiliários	Modelo hedônico sem tratamento para dependência espacial. Estimação por MQO.
Alves et al (2011)	São Paulo	jan-2001 a mar-2008	1254	EMBRAESP	Modelo hedônico com tratamento para dependência espacial. Estimação por MQO Empilhado.
González e Formoso (1994)	Porto Alegre	jun-1992	504	Coleta manual em jornais e imobiliárias	Modelo hedônico sem tratamento para dependência espacial. Estimação por MQO.

Fonte: Elaborado pelo autor.

Analisando a tabela, nota-se que seis dos dez estudos citados não fazem nenhum tratamento da dependência espacial (GONZÁLES E FORMOSO, 1994; FERREIRA NETO, 2002; RONDON E ANDRADE, 2003; SANT'ANNA, 2006; ROZENBAUM E MACEDO-SOARES, 2007; FÁVERO *et al.*, 2008). Além disso, oito *papers* utilizam a estimação por MQO, o que pode gerar estimativas viesadas e inconsistentes⁶, ao passo que em apenas dois estudos (MACEDO, 1998; FURTADO, 2007) estimam os modelos por Máxima Verossimilhança. Em um primeiro momento, estes erros na estimação e na omissão da dependência espacial dos modelos poderiam ser atribuídos à disseminação do conhecimento, uma vez que a literatura e os avanços computacionais poderiam ter restringido o uso de modelos mais sofisticados. No entanto, este argumento parece não ser válido, uma vez que alguns estudos mais antigos, como o de Macedo (1998), levam em consideração a dependência espacial e realizam a estimação por Máxima Verossimilhança, ao passo que estudos mais recentes cometem alguns destes equívocos na definição dos modelos.

Outro ponto relevante na Tabela 1 diz respeito ao número de observações, apresentando certa heterogeneidade neste aspecto. Alguns dos *papers* citados possuem uma amostra relativamente pequena (o estudo de Macedo (1998) analisa 53 observações, o de Ferreira Neto (2002), 120), o que evidencia o já citado problema da disponibilidade de dados de qualidade⁷. Em contrapartida, alguns trabalhos dispõem de uma base de dados maior, como ocorre em Sant'anna (2006), Furtado (2007), Fávero *et al.* (2008), Alves *et al.* (2011), em que há mais de mil observações.

No entanto, junto ao número de observações, é necessário analisar o horizonte contemplado pela amostra. Nota-se que a maioria dos estudos contempla um curto prazo de tempo, em que há alguns casos cujo os estudos limitam-se a apenas um mês de observações (GONZÁLES E FORMOSO, 1994; MACEDO, 1998; RONDON E ANDRADE, 2003; FURTADO, 2007) ou a um ano (FERREIRA NETO, 2002; HERMANN E HADDAD, 2005; FÁVERO *et al.*, 2008). Como exceção, há os estudos de Sant'anna (2006), que possui um horizonte temporal de dezesseis anos e de Alves *et al.* (2011), contemplando um período de oito.

⁶ Para mais detalhe sobre este viés, recomenda-se Almeida (2012).

⁷ Alguns estudos utilizam informações coletadas manualmente de sites de imobiliárias ou jornais, como é o caso de Gonzáles e Formoso (1994), Furtado (2007) e Fávero *et al.* (2008).

Dado este panorama das pesquisas realizadas no país, vislumbra-se uma oportunidade de expandir a literatura local utilizando técnicas de econometria espacial, pouco utilizadas nos estudos citados previamente. No presente trabalho, serão aplicadas estas técnicas a uma base de dados com um grande número de observações com o foco na previsão fora da amostra e no impacto do tempo no mercado. Entende-se que a comparação entre diferentes modelos econométricos aplicados a uma grande amostra contribuirá para a literatura. Tais métodos serão explicitados na seção 5.

3.2. Tempo no mercado

O mercado imobiliário tende a ser pouco líquido devido a algumas características específicas. Como os imóveis são bens heterogêneos (têm uma alta variedade de características intrínsecas ou de vizinhança), o número de compradores para um único imóvel pode ser pequeno, uma vez que há diversas opções. Além da alta disponibilidade de recursos financeiros (no caso de compra), há uma série de fricções que diminuem a liquidez deste mercado: a necessidade de aprovação de financiamento, o custo de procurar o imóvel, custos de transação, de corretagem e de mudança, o que implica em preços voláteis e grandes diferenças no *bid-ask spread* (HAN, 2014; KRAINER, 2001). Diferentemente do mercado de ações, em que os preços dos ativos são definidos pelo mercado, um proprietário deve definir discricionariamente o valor inicial em seu anúncio. Nisso, incorre-se em um *trade-off*: ao divulgar no mercado um preço acima do considerado justo, há chances de o vendedor realizar um lucro maior. No entanto, é possível que este negócio demore mais tempo para se concretizar, pois os compradores estarão buscando outros imóveis semelhantes a um menor valor. De forma análoga, ao oferecer o imóvel com desconto, abaixo do valor justo, o imóvel pode ser vendido rapidamente, porém, diminuindo o lucro do vendedor⁸. Por isso, o tempo no mercado (em inglês, *time on market* - TOM) é uma variável estudada na literatura estrangeira pois pode fornecer informações relevantes sobre o comportamento dos vendedores e da liquidez do mercado de forma geral (TAYLOR, 1999; KNIGHT, 2002).

⁸ O trabalho de Miller (1978) é um dos primeiros encontrados na literatura abordando esta dinâmica.

Vendedores que apresentam um alto custo de oportunidade de seu capital, devem escolher um preço de venda inicial do imóvel que seja mais atrativo para compradores, a fim de que a transação ocorra rapidamente. Já um proprietário com um menor custo de oportunidade (ou, menor taxa de desconto), poderá aguardar mais tempo para que receba ofertas melhores por seu ativo, ou, definir um preço alto e aguardar o tempo necessário para realizar o lucro esperado. Miller (1978) ressalta que, além destes efeitos, espera-se que com um maior tempo no mercado, o vendedor reajuste os preços a fim de repor perdas inflacionárias, o que é confirmado em sua análise empírica. No entanto, Taylor (1999) deriva um modelo teórico no qual, à medida que um imóvel permanece no mercado, este ativo pode ser visto como sendo de baixa qualidade. Segundo este modelo, pode haver alguma informação a respeito do imóvel que não está sendo disponibilizada aos possíveis compradores. Para evitar esta percepção, os proprietários terão cuidado em definir o preço inicial de oferta do imóvel: o vendedor tenderá a colocar um preço inicial mais baixo para evitar que o imóvel fique muito tempo no mercado e, com isso, estigmatizado. Isto sugere que um imóvel ainda disponível no mercado e um preço abaixo da média pode tratar-se de uma residência com algum problema, ou apresentar baixa qualidade⁹. Chen e Rutherford (2012), a partir de um modelo Logit, encontram que imóveis com maior (menor) qualidade são vendidos de forma mais rápida (demorada), indo ao encontro da teoria de Taylor (1999).

Embora estes trabalhos estudem a diferença entre o preço inicial e o de venda concretizada, há estudos que analisam como o preço do imóvel varia entre estes períodos, ou seja, o período em que ele está sendo anunciado no mercado. Como na maioria dos mercados o preço de oferta não é fixo, um vendedor pode lançar o imóvel a um determinado preço, observar a resposta do mercado e, após isso, ajustar o preço de venda. Resumindo os trabalhos previamente citados, Yavas e Yang (1995) definem um modelo teórico para analisar esta oscilação de preços: com base em outros anúncios e valores de venda de imóveis próximos e semelhantes ao seu, o vendedor terá sua própria noção de valor do imóvel, dada como P_S . A partir disso, ele definirá P_L como o preço anunciado do imóvel e a venda ocorrerá se o preço máximo que um comprador estiver disposto a pagar por um imóvel (P_B) for igual ou maior a P_L , considerando que os atributos procurados no imóvel são iguais aos oferecidos.

⁹Em suma, aproxima-se de um problema de seleção adversa. Para uma revisão do tema, sugere-se Akerlof (1978).

Se $P_L > P_B$ nenhuma transação ocorre naquele período, porém há oportunidade de o vendedor rever o preço de venda e aguardar novas ofertas. No entanto, a definição do preço inicial do imóvel tem importância pois, caso o vendedor tenha demanda por liquidez, ou, por algum motivo o seu custo de oportunidade for alto, ele terá pressa em alterar os preços para que mais compradores sejam atraídos. A partir disso, se há um reajuste nos preços, o vendedor estará sinalizando a alteração no valor mínimo que estaria disposto a vender o imóvel. Consonante a isto, Knight (2002) encontra evidências de que, ao ajustar os preços dos imóveis para baixo, a venda ocorre em menor tempo. Em contrapartida, imóveis os quais não tiveram seu preço reajustado levaram mais tempo para serem comercializados. Além disso, há evidências de que os imóveis que estão há mais tempo no mercado são vendidos com um desconto maior em relação ao seu preço de venda inicial. Isto corrobora com a hipótese de Taylor (1999) em que os imóveis são estigmatizados à medida que eles permanecem no mercado. Outros estudos mais recentes como, Beracha e Seiler (2014), Carrillo (2013), Merlo *et al.* (2015) e Smith *et al.* (2015) complementam esta temática. Além desta visão, é possível que a queda no preço do anúncio reflita uma demanda por liquidez por parte dos vendedores.

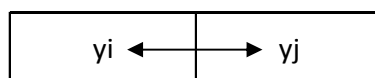
Embora as evidências mais recentes sejam a favor da teoria de Taylor (1999), para o mercado de venda de imóveis, não há evidências na literatura estrangeira sobre esta dinâmica para imóveis à locação. Isto se deve, provavelmente, à dificuldade em se obter este tipo de dados. No entanto, acredita-se que o mercado de locação tenha motivações mais favoráveis à uma redução no valor com um maior tempo no mercado. Considerando um proprietário que se depara com a escolha entre colocá-lo à venda ou à locação, se o seu custo de oportunidade for alto, talvez ele opte por alugar o imóvel para obter algum retorno sobre seu patrimônio. Isso porque, o processo de venda muitas vezes pode ser demorado devido à definição do comprador, da liberação de financiamento (se necessário) e de outros custos como o de transação. Além disso, um locador está protegido contra perdas inflacionárias, já que os contratos de aluguel geralmente incluem uma cláusula de reajuste anual indexada a um indicador de preços. Como a renda de um investimento em aluguel caracteriza-se como um dividendo, é esperado que um locador se preocupe em encontrar rapidamente um locatário para que possa começar a receber a renda proveniente, caso contrário, o locador arca com custos de manutenção do imóvel. Dado isso, também é esperado que um locador que não

teve obtido sucesso ao divulgar seu imóvel no mercado reduza o preço de locação com maior intensidade.

3.3. Econometria espacial

Almeida (2012) define a econometria espacial como um ramo que analisa, estima, prevê e testa modelos teóricos influenciados pelos efeitos espaciais. Segundo o autor, estas técnicas econométricas começaram a surgir no início da década de setenta e vêm se desenvolvendo até então. A importância da econometria espacial surge pelo fato das estimações por MQO, muito utilizadas na econometria tradicional, muitas vezes não encontrarem estimadores eficientes e consistentes na presença de efeitos espaciais (ALMEIDA, 2012). Tais efeitos referem-se à dependência espacial e à heterogeneidade espacial. Em um contexto de dados em corte transversal, a dependência espacial ocorre quando indivíduos, domicílios, empresas ou bairros não são mais independentes entre si e a fonte desta dependência está nas relações de proximidade entre as unidades de medida. Exemplificando: o valor de um imóvel em um bairro, y_i , dependerá de bairros próximos a ele, digamos, y_j . Desta forma, há o conceito de multidirecionalidade do processo espacial, em que a variável y_i impacta y_j , mas também é impactada por y_j . Exemplificando no mercado imobiliário gaúcho, é como se o preço de um imóvel no bairro Centro, impactasse no valor de um imóvel no bairro Cidade Baixa¹⁰ e vice-versa. De forma geral, a Figura 1 retrata esta dinâmica.

Figura 1 – Dinâmica do processo espacial



Fonte: Adaptado de Almeida (2012).

Por outro lado, a heterogeneidade espacial trata da instabilidade da ocorrência dos fenômenos ao longo das regiões. Segundo Almeida (2012), isso pode ocorrer tanto nos coeficientes estimados quanto no padrão do erro aleatório, violando a hipótese de homoscedasticidade. Tyszler (2006) exemplifica o caso em que estados ricos e pobres estão

¹⁰ Estes dois bairros fazem fronteira geográfica um com o outro.

inseridos em diferentes regiões de um mesmo país, mas o coeficiente a ser estimado será apenas um para o país inteiro. Isso pode levar a instabilidade nos coeficientes estimados, uma vez que a delimitação geográfica dos bairros, estados ou regiões podem não ser as melhores medidas para mensurar, ou agrupar, determinadas unidades de medida.

De forma geral, um modelo básico de econometria espacial (SAR – *Spatial Autoregression Model*) tem a seguinte forma estrutural:

$$y = \rho W y + X \beta + \varepsilon, \quad (1)$$

em que y refere-se a variável dependente, ρ é o coeficiente de dependência espacial a ser estimado, W é a matriz de dependência espacial e X refere-se a variáveis explicativas exógenas com os respectivos coeficientes β , e ε refere-se ao termo de erro aleatório. A matriz W procura refletir um arranjo espacial das interações resultantes do fenômeno a ser estudado (ALMEIDA, 2012) em que regiões mais próximas interagem mais do que regiões mais distantes. A questão primordial é definir o que usar para medir esta conectividade entre as regiões para, só então, realizar as estimações do modelo (GONZÁLES e FORMOSO, 2000; PLAIGIN, 2006; TYZSLER, 2006). Além disso, é possível antever problemas de endogeneidade neste modelo, uma vez que a variável y está nos dois lados da equação, devido à dependência espacial. A literatura mostra que a presença de autocorrelação espacial viesia os estimadores por MQO (ALMEIDA, 2012; ELHORST, 2000; FURTADO, 2007; LESAGE, 1999). Ou seja, as hipóteses de Gauss-Markov são violadas, não gerando estimadores BLUE¹¹. A seção seguinte mostra como especificar esta dependência espacial através de matrizes de ponderação.

3.3.1. Matrizes de dependência espacial

Ao analisar dados com dependência espacial, uma questão que o pesquisador se depara é sobre como modelar corretamente tipo de interação. Calcular a matriz de variância-covariância entre todas as unidades espaciais (por exemplo, bairros de uma cidade) acarretaria em um problema computacional difícil de resolver, além de pouco prático (ALMEIDA, 2012). Para solucionar isto, utiliza-se uma matriz W que evidencia um arranjo para

¹¹ Sigla para *Best Linear Unbiased Estimator* – Melhor estimador linear não viesado (tradução nossa).

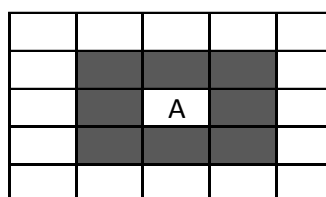
a ocorrência de todas as interações espaciais entre todas as regiões. Os componentes dessa matriz são os termos w_{ij} , que definem a interação entre as diferentes regiões i e j . Por relacionar todos os pares de observações, a matriz W é quadrada de tamanho n por n . É importante a forma de definição desta matriz, pois a escolha dela muitas vezes é arbitrária, uma vez que ainda não é conhecido nenhum teste formal para defini-la (PLAIGIN, 2006). Além disso, os resultados das estimações de um modelo espacial podem ser sensíveis a escolha das matrizes de pesos (ALMEIDA, 2012). Elas podem ser de âmbito de distância geográfica, bem como socioeconômica. Pelo escopo deste trabalho, serão focadas as matrizes de proximidade geográfica, primeiramente evidenciadas na literatura por Moran (1948) e Geary (1954).

Uma tipologia de matriz W pode ser baseada na contiguidade, na qual duas regiões, estados, cidades ou qualquer outra unidade espacial, serão consideradas vizinhas caso tenham uma fronteira geográfica em comum (LESAGE, 1999). Com base neste conceito, receberão o valor um em caso positivo e zero caso contrário. Por convenção da literatura, considera-se $w_{ii} = 0$, em que a observação não é vizinha de si própria (ALMEIDA, 2012; TYSZLER, 2006):

$$w_{ij} = \begin{cases} 1, & \text{se } i \text{ e } j \text{ são vizinhos} \\ 0, & \text{se } i \text{ e } j \text{ não são vizinhos} \end{cases} \quad (2)$$

A Figura 2 mostra uma tipografia dos vizinhos da observação “A” com base na contiguidade. A áreas hachurada refere-se aos vizinhos:

Figura 2 – Configuração de vizinhança por contiguidade



Fonte: ALMEIDA, 2012

Almeida (2012) mostra que, além da contiguidade das regiões, é possível definir a matriz de ponderação a partir da distância geográfica da região até um ponto de corte definido previamente. Neste caso, é adotada uma matriz dos k vizinhos mais próximos, $w_{ij}(k)$:

$$w_{ij}(k) = \begin{cases} 1, & d_{ij} \leq d_i(k) \\ 0, & d_{ij} > d_i(k) \end{cases} \quad (3)$$

em que $d_i(k)$ refere-se à distância de corte para a região i a fim de que esta região tenha k vizinhos. Desta forma, essa distância de corte varia entre as regiões. É importante ressaltar que, neste caso não está sendo considerada a contiguidade, ou fronteira, entre as regiões, mas sim, o número de vizinhos que a região terá. Almeida (2012) exemplifica esta tipologia com as regiões do país na tabela abaixo, considerando os dois vizinhos mais próximos para as macrorregiões do país.

Tabela 2 - Matriz de dois vizinhos mais próximos

Regiões	N	NE	CO	SE	S
N	0	1	0	0	0
NE	1	0	0	0	0
CO	1	1	0	1	1
SE	0	0	1	0	1
S	0	0	1	1	0

Fonte: Adaptado de Almeida (2012).

Desta maneira, corrige-se um potencial problema que é o de alguma região não ter vizinhos, ou tê-los em muito pouca quantidade (principalmente no caso de fronteiras/bordas de cidades), pois todas as unidades espaciais terão o mesmo número de vizinhos. Além deste tipo de matriz, Tyszler (2006) mostra que é possível adaptar esta tipologia para evitar a utilização de variáveis binárias, baseada na distância normal ou inversa:

$$w_{ij}(k) = \begin{cases} 1/d_{ij}, & d_{ij} \leq d_i(k) \\ 0, & d_{ij} > d_i(k) \end{cases} \quad (4)$$

Desta maneira, são dados pesos às distâncias reais entre as regiões. Isto pode ser relevante, uma vez que, ao usar *dummies*, todas as inter-relações têm o mesmo impacto, o que não ocorre nesta configuração. Tyszler (2006) realiza simulações de Monte Carlo para comparar diferentes matrizes de ponderação e chega ao resultado de que, dados pressupostos de normalidade e homoscedasticidade dos erros, as matrizes de contiguidade geram resultados piores na estimação dos coeficientes do que matrizes de distâncias quando o número de observações aumenta. Uma prática na aplicação de matrizes de ponderação

espacial é a normalização em cada linha. Isso permite a interpretação dos pesos como percentuais de impacto espacial, já que, após a normalização, os pesos estarão contidos no intervalo zero ou um (TYSZLER, 2006). No entanto, é importante ressaltar que o decaimento de impacto à medida que a distância geográfica aumenta entre regiões pode ser perdida com a normalização.

Almeida (2012) ressalta a dificuldade em escolher um dos modelos da matriz W e o número de vizinhos. O autor cita Baumont (2004) que sistematiza um procedimento para reduzir a arbitrariedade da escolha do número de k vizinhos a serem levados em conta na matriz de ponderação. Primeiramente, é estimado o modelo clássico de regressão linear por MQO, testam-se os resíduos para autocorrelação espacial pela estatística I de Moran¹², usando diferentes valores de k vizinhos. Baumont (2004) usou matrizes variando $k =$ de 1 até 20, escolhendo a matriz que gerou a maior estatística significativa do I de Moran. Portanto, isto evidencia o modelo com maior a dependência espacial possível.

3.3.2. Testes para identificação de dependência espacial

Para a correta estimação dos modelos espaciais é necessário saber que tipo de autocorrelação está implícita nos dados. Estimar um modelo com dependência espacial quando este efeito não existe pode acarretar em estimativas viesadas dos coeficientes e/ou uma complexidade de estimação maior que a necessária.

De forma geral, podem ser definidos dois tipos de testes para a dependência espacial: os testes difusos e os testes focados (ALMEIDA, 2012). Os primeiros não informam o tipo de autocorrelação espacial predominante no modelo, apenas testam a hipótese nula de não dependência espacial nos resíduos. Sendo assim, se esta hipótese é rejeitada, não se sabe qual o tipo de dependência que ocorre. No caso dos testes focados, são fornecidas indicações de tipo predominante de dependência espacial. Com isso, esta categoria de testes tem por

¹² Este procedimento será detalhado na subseção seguinte.

objetivo discriminar o processo gerador de autocorrelação espacial. A maioria das abordagens utilizadas nestes casos são testes de Multiplicadores de Lagrange¹³.

O teste difuso I de Moran é um teste simples para detecção de dependência espacial nos resíduos da regressão. A forma sugerida por Almeida (2012) é dada por:

$$I = \frac{n}{S_0} \left(\frac{e'W e}{e'e} \right) \quad (5)$$

em que e são os resíduos da estimação por MQO e $S_0 = \sum_i \sum_j w_{ij}$, representa um fator de normalização. Caso a matriz W esteja normalizada na linha, $S_0 = n$ uma vez que a soma das linhas será igual a unidade. Portanto, o teste I de Moran pode ser reescrito como:

$$I = \left(\frac{e'W e}{e'e} \right) \quad (6)$$

Na hipótese nula, os resíduos não são correlacionados, ou seja, eles são independentes e distribuídos aleatoriamente no espaço. A inferência do teste é feita por meio de uma variável $z(I)$ que segue assintoticamente uma distribuição normal:

$$z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \quad (7)$$

o valor esperado de I é dado por:

$$E(I) = \frac{tr(MW)}{n - k} \quad (8)$$

e a variável M :

$$M = I_n - X(X'X)^{-1}X'. \quad (9)$$

Por fim, a variância da estatística I é definida como:

$$Var(I) = \frac{\{tr(MWM') + tr(MW)^2 + [tr(MW)]^2\}}{\{(n - k)(n - k + 2)\}}. \quad (10)$$

¹³ Para uma breve revisão do Multiplicadores de Lagrange, sugere-se a leitura de PORTUGAL (1995) e de CRAMER (1986) para uma revisão mais aprofundada.

Este é um teste de fácil implementação computacional, porém requer que os resíduos sejam normais e homoscedásticos. Além disso, o teste é sensível à escolha da matriz de ponderação W e não identifica o tipo de autocorrelação espacial. Almeida (2012) ressalta que, em casos de heteroscedasticidade, há maior erro do tipo 1 (rejeitar a hipótese nula quando ela é verdadeira). Como há estes problemas em sua estimação e o procedimento não indica qual o processo de dependência espacial que ocorre nos dados, sugere-se não basear a decisão de um modelo apenas neste teste.

Os testes focados informam o tipo de dependência espacial que toma forma quando a hipótese nula de independência é rejeitada. Um exemplo é teste do Multiplicador de Lagrange, MLp. Ele verifica a hipótese nula $H_0: \rho = 0$ contra a hipótese alternativa $H_1: \rho \neq 0$. Ou seja, está sendo testada a hipótese de estarmos tratando de um SAR. Para verificar isso, são estimados o vetor escore e a matriz de informação através do cálculo do logaritmo da máxima verossimilhança, que assume a forma de:

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \sum \ln \det(I_n - \rho W) + \frac{(y - \rho W y - X\beta)'(y - \rho W y - X\beta)}{2\sigma^2} \quad (11)$$

Após isso, é necessário derivar o vetor escore dessa função e avaliá-lo em $\rho = 0$. Portanto:

$$d_p = -tr (I_n - \rho W)^{-1} W + \frac{e' W y}{\sigma^2} \quad (12)$$

$$d_p = \frac{e' W y}{\sigma^2}$$

Após isso, computa-se a matriz de informação:

$$Inf = \begin{bmatrix} tr [W^2 + W'W] + \frac{[WX\beta]'[WX\beta]}{\sigma^2} & \frac{(X'WX\beta)'}{\sigma^2} \\ \frac{(X'WX\beta)'}{\sigma^2} & \frac{X'X}{\sigma^2} \end{bmatrix}. \quad (13)$$

Por fim, a estatística de teste é dada por:

$$ML_{\rho} = \frac{\left(\frac{e'Wy}{s^2}\right)^2}{\left\{\frac{(WX\hat{\beta})'MWX\hat{\beta}}{s^2} + tr[W'W + W^2]\right\}} \quad (14)$$

em que s^2 é a estimativa para variância do erro (σ^2) e $\hat{\beta}$ são as estimativas por MQO dos coeficientes. Este teste segue uma distribuição qui-quadrado com graus de liberdade iguais ao número de restrições sendo testadas. Almeida (2012) ressalta que, por se tratar de um teste assintótico, essa estatística é mais apropriada para grandes amostras e que o teste assume que a especificação do modelo foi feita corretamente. Através de simulações de Monte Carlo, Anselin e Rey (1991) mostram que este teste é robusto contra erros não normais e contra heteroscedasticidade. Outra vantagem é a facilidade computacional de sua aplicação. No entanto, uma desvantagem é a falta de poder que implica na frequente rejeição da hipótese nula quando ela é verdadeira. Desta forma, recomenda-se o uso dos dois testes citados nesta seção para obter uma maior confiabilidade nas estimativas.

3.4. Previsão de preços de imóveis

A definição de preços de imóveis impacta direta, ou indiretamente, diversos aspectos de uma economia. A um nível doméstico, muitas vezes um imóvel é um dos maiores componentes da riqueza de uma família (LIU, 2012). Em mercados como o Reino Unido, há evidências que um aumento no preço de imóveis impacta positivamente o consumo de famílias proprietárias de imóveis (CAMPBELL e COCCO, 2007), o que se relaciona com restrições financeiras e a percepção de riqueza das famílias. Além deste trabalho, há diversos outros estudando esta temática (ADELINO *et al*, 2015; BROWING *et al*, 2013; CASE *et al*, 2013).

Tendo conhecimento disto, é relevante saber se, de fato, é possível prever os preços de imóveis. Em um artigo seminal para os estudos do mercado imobiliário, Case e Shiller (1988) demonstram que os índices de preços de imóveis em quatro cidades dos Estados Unidos seguiam um processo autorregressivo de primeira ordem. Após isso, os autores realizaram um estudo evidenciando a previsibilidade dos preços dos imóveis por meio de estimações via MQO (CASE e SHILLER, 1990). Neste contexto, Pace *et al*. (1998) utilizam um modelo espaço-temporal para prever os preços de imóveis do condado de Fairfax, Estados

Unidos. O modelo adotado utilizava 12 regressores, em comparação com um modelo estimado por MQO usando variáveis *dummy* com 27 variáveis. Os modelos foram estimados usando 70.822 observações entre os anos de 1969 a 1991. Com o modelo espaço-temporal, houve uma redução de 37% nos erros de previsão dentro da amostra, reduzindo-os de 14,78% para 9,26%¹⁴. Além disso, os resíduos apresentaram baixa correlação no modelo espacial-temporal, ao contrário do modelo tradicional, estimado por MQO. Em seguida, Pace *et al.* (2000) estendem o modelo utilizado em Pace *et al.* (1998) para uma aplicação com uma amostra de 5243 observações do município de Baton Rouge, Estados Unidos, durante o período de 1985 a 1992. O estudo trouxe resultados semelhantes ao anterior, em que o modelo espaço-temporal reduziu a soma dos quadrados dos resíduos em 8% chegando ao valor de 7,4% para previsões fora da amostra.

Liu (2012) usa o modelo de Pace *et al.* (1998) não considerando a classificação de bairros dos imóveis, mas de subgrupos definidos por corretores de imóveis atuantes no mercado. Essa identificação visa diminuir a heterogeneidade dos preços, uma vez que há a hipótese de que a delimitação dos bairros não seja um agrupamento adequado para representar os preços de imóveis (HERMANN E HADDAD, 2005)¹⁵. De forma semelhante a Pace *et al.* (1998), o autor estima um modelo por MQO e outro usando o modelo espaço-temporal em uma base de dados de 437.734 transações imobiliárias da região Randstad, Holanda. Os resultados apontam um pseudo R^2 de 0.87 no modelo STAR em comparação a 0.77 no modelo MQO. Para a previsão fora da amostra, foram reservados 20% da amostra para a estimação, a qual o erro absoluto mediano encontrado para o modelo por MQO foi de 13,4% em comparação a 9,4% do modelo espaço-temporal¹⁶.

Um ponto que merece atenção é o estudo de Bourassa *et al.* (2007), em que, contrariando a literatura, os autores chegam a melhores resultados na previsão de preços utilizando uma estimação por MQO com 4.880 transações imobiliárias de Auckland, Nova Zelândia. Ao comparar projeções fora da amostra entre estimações por MQO, SAR e modelos

¹⁴ Considerando o Erro Absoluto Mediano. Para mais detalhes, ver Tabela 4 de Pace *et al.* (1998, p.27).

¹⁵ Almeida (2012) define o Problema da Unidade de Área Modificável (MAUP, em inglês), que consiste no fato de que o zoneamento utilizado para classificar e delimitar uma região pode não representar a sua dinâmica espacial. Sendo assim, dependendo da escala utilizada, pode-se gerar autocorrelação espacial artificial.

¹⁶ Para maiores detalhes, ver a Tabela 4 de Liu (2012, p.360).

geoestatísticos¹⁷, em algumas especificações, os autores chegam a um erro de previsão menor nas estimações via MQO. Por exemplo, no modelo MQO, foram encontrados um erro médio absoluto e um erro percentual médio de 53.168 e 17,8%, respectivamente, contra 60.534 e 18,1% no modelo SAR. No entanto, ao adicionar nos modelos as classificações referentes a submercados¹⁸ estas medidas são alteradas em favor do modelo SAR. Os erros médios absolutos e percentuais encontrados foram de 49.074 e 14,6% contra 49.901 e 15% nas estimações por MQO.

É importante ressaltar que a base de dados utilizada por Bourassa *et al.* (2007) continha um alto detalhamento sobre as características dos imóveis, como o tempo de construção, a qualidade do acabamento interno, qualidade do bairro, acessibilidade, se o imóvel apresentava alguma vista, número de garagens, dentre outras características. Hermann e Haddad (2005) encontraram que, com uma maior disponibilidade de dados sobre as variáveis de vizinhança disponíveis, a dependência espacial nos modelos tendia a desaparecer nas estimações. Estes dois resultados podem estar relacionados, uma vez que as estimações por MQO de Bourassa *et al.* (2007) tiveram uma boa performance com uma base de dados mais detalhada. Uma vez que as bases de dados disponíveis sobre preços de imóveis são as mais variadas possíveis, é interessante testar estes diferentes modelos para previsão, uma vez que a literatura não chegou a resultados conclusivos a favor de um, ou de outro.

¹⁷ Estes modelos consideram a localização exata dos pontos baseados nas coordenadas de longitude e latitude. Como esta característica não se apresenta nos dados utilizados neste trabalho, não serão detalhados estes procedimentos de estimação.

¹⁸ Nesta literatura, este termo refere-se à identificação dos imóveis não pelos bairros, mas por regiões delimitadas por avaliadores e participantes do mercado. Para maiores detalhes, ver Bourassa *et al.* (2007) e Liu (2012)

4. DADOS E VARIÁVEIS UTILIZADAS

Para as estimações dos modelos de previsão serão usadas duas bases de dados separadas: uma de apartamentos residenciais à venda e outra para apartamentos residenciais disponíveis para locação, ambas contemplando a cidade de Porto Alegre e retiradas do portal Penselmóveis¹⁹. Um algoritmo em MatLab[®] foi utilizado para fazer a coleta automatizada dos dados para estes tipos de imóveis à venda (locação) entre o período de Outubro de 2012 (Novembro de 2012) a Novembro de 2013 em intervalos de tempo irregulares de, aproximadamente, 15 dias²⁰. Era disponível uma base de dados com frequência maior dos dados, porém optou-se por reduzir a amostra pois os intervalos eram demasiados irregulares, o que poderia impactar na confiabilidade dos resultados, além de questões de viabilidade computacional²¹. Para facilitar a referência a estas bases, ela será referida no trabalho como “base restrita”. Foram removidos o primeiro e o último percentil de observações referentes ao valor por metro quadrado útil e total, área total área útil e do valor do imóvel. Esta base contempla apenas imóveis que aparecem uma única vez na amostra. As bases de dados contêm 15081 observações para venda e 1579 para locação, tendo todos os valores monetários deflacionados para Outubro de 2012, mês inicial das observações à venda²².

Para a verificar o impacto do tempo no mercado no valor do imóvel foi utilizada uma base de dados com maior frequência e maior número de observações do que a usada para os modelos de previsão. Trata-se da mesma base, porém com mais períodos em que foram coletados dados e, conseqüentemente, mais observações. O uso desta base maior se faz necessário pois o interesse é, principalmente, nos imóveis que se repetem ao longo da amostra e que tem alguma alteração no preço do anúncio. Como a parcela de imóveis que tem seu valor alterado durante o período do anúncio é relativamente pequena em relação ao total, isso tornaria os resultados pouco confiáveis utilizando poucas observações. Esta base será referida neste trabalho como “base completa”. Ao total, serão utilizadas 15178 (2355) observações de 1345 (207) imóveis repetidos para venda (locação) que têm seu preço alterado

¹⁹ O portal Penselmóveis recentemente foi adquirido pelo Zap Imóveis (www.zapimoveis.com.br)

²⁰ Agradeço ao prof^o Marcelo Perlin, autor do algoritmo, e orientador deste trabalho, por ter cedido esta base de dados.

²¹ As listas com as datas exatas utilizadas nas bases de dados deste trabalho constam no Apêndice.

²² Os modelos de previsão também foram estimados com as bases de dados completas gerando resultados semelhantes.

no período da amostra²³. Serão considerados os efeitos do tempo no mercado no preço nominal dos imóveis e, posteriormente, será realizada uma comparação com os efeitos inflacionários.

Em ambas as bases foram removidas observações de bairros que não estivessem presentes em todos os períodos da amostra. Além disso, os nomes dos bairros foram ajustados para que estivessem em conformidade com a lista oficial disponibilizada pela prefeitura de Porto Alegre²⁴. Algumas subdivisões de bairros acabam tornando-se referência para os anunciantes, embora não sejam bairros em si. Por exemplo, o bairro Intercap refere-se a uma área específica do bairro Partenon, assim como o Jardim Dona Leopoldina refere-se a uma região do bairro Rubem Berta. Para realizar esta tarefa, as referências que não constam na lista oficial foram pesquisadas na plataforma Google Maps® e nos sites da Prefeitura de Porto Alegre, a fim de coletar informações sobre a qual bairro oficial pertencem as indicações nos anúncios. Para evitar uma classificação errônea, bairros não-oficiais em que não foram encontradas referências foram removidos da amostra.

4.1. Descrição das variáveis

A variável dependente neste estudo será o preço do imóvel, medida amplamente usada em estudos semelhantes a este (HERMANN e HADDAD, 2005; FURTADO, 2007; LIU, 2012; PACE *et al*, 2000). As variáveis explicativas para os modelos de previsão, são a área total, o número de quartos, o número de vagas na garagem disponíveis, o tempo no mercado do imóvel, além de *dummies* para cada bairro²⁵. Estas variáveis são comumente usadas em estudos do mercado imobiliário (ALVES *et al*, 2012; LIU, 2012; MACEDO, 1998; PACE *et al*.

²³ Cada observação na base de dados possui um código de identificação. Com isto é possível identificar os imóveis que se repetem ao longo da amostra.

²⁴ Há um projeto de lei para a alteração dos bairros de Porto Alegre, porém até a entrega desta dissertação, ainda não fora aprovado. Para visualizar a lista de bairros oficiais de Porto Alegre: http://proweb.procempa.com.br/pmpa/prefpoa/observatorio/usu_doc/historia_dos_bairros_de_porto_alegre.pdf.

²⁵ Também foram testadas especificações incluindo a distância a pontos de interesse como *shoppings*, escolas particulares, praças e parques. Outras características do imóvel obtidas na descrição do anúncio, como a presença de churrasqueiras, piscinas, o fato do imóvel ser uma cobertura e criminalidade do bairro também foram testadas. Além disso foram estimados modelos incluindo índices do mercado como IVG-R, FIPE-ZAP e INCC. Embora estas variáveis e especificações sejam utilizadas na literatura, elas não apresentaram coeficientes significativos nas estimações e/ou não contribuíram no poder preditivo dos modelos.

1998; PACE *et al.* 2000; SANT'ANA 2006). Embora não tenha sido encontrado nenhum estudo brasileiro que utilizasse a variável tempo no mercado, ela é estudada na literatura estrangeira (BERACHA e SELIER, 2014; KNIGHT, 2002; MERLO *et al.* 2015). Além destas variáveis intrínsecas ao imóvel, foi adicionada como regressora a taxa de financiamento imobiliário, divulgada pelo Banco Central em frequência mensal. Alves *et al.* (2012) mostram que as taxas de financiamento podem afetar o preço dos imóveis. A lista das variáveis utilizadas consta na tabela abaixo:

Tabela 3- Variáveis utilizadas

Variável	Descrição
Preço	Preço do imóvel à venda ou locação
Tempo no mercado	Dias em que o anúncio está no portal
Área Total	Área total do imóvel
Nº Dormitórios	Número de dormitórios do imóvel
Nº Garagens	Número de vagas na garagem do imóvel
Tx. Financ.	Taxa de financiamento imobiliário
<i>Dummy</i> bairro	<i>Dummy</i> para o bairro do imóvel

Fonte: autor

Além destas variáveis, será criada uma matriz de pesos espaciais. O cálculo desta matriz é feito a partir de informações de georreferenciamento da cidade de Porto Alegre, disponíveis no portal ObservaPOA²⁶. Utilizando softwares como QGIS® e GeoDa®, é possível exportar as coordenadas para que possa ser construída uma série coma localização de cada bairro. Desta forma, é possível calcular as distâncias entre bairros e, com isso, gerar as matrizes de dependência espacial.

²⁶ www.observapoa.com.br

5. MÉTODO

Esta seção tratará da abordagem a ser utilizada na parte empírica.

5.1. Impacto do tempo no mercado

Para avaliar como o tempo no mercado afeta o preço de imóveis à venda e à locação, serão analisadas estatísticas descritivas relativas aos imóveis que se repetem durante a base de dados utilizada. Conforme a seção 3.2, espera-se que o tempo que um imóvel está disponível no mercado tenha impactos diferentes entre os mercados de venda e de locação. Para avaliar isso, será comparado o percentual de imóveis com reajuste positivo e negativo, bem como o percentual deste ajuste médio. Além disto, a variável tempo no mercado também se encontra nos modelos de previsão para que sua significância seja testada.

5.2. Modelos para previsão

A estimação para imóveis à venda (locação) usará aproximadamente 10% (40%) dos dados para a calibragem dos modelos e o restante para previsão fora da amostra. A calibragem contempla o período de Outubro (Novembro) de 2012 a meados Maio de 2013. Da segunda quinzena de Maio a Novembro de 2013 é realizada a previsão fora da amostra, utilizando os coeficientes estimados no período de calibragem. O mês de Maio de 2013 divide as estimações dentro e fora da amostra pois é o ponto mediano no horizonte temporal das bases de dados. Embora a distribuição das observações seja assimétrica (há mais observações no final da amostra) e o horizonte temporal curto, optou-se por privilegiar um período maior de previsão ao invés de um equilíbrio entre observações no período de calibragem e no período de previsão. A Tabela 4 resume os dados para as estimações.

Tabela 4 - Período de estimações dentro e fora da amostra

	Venda	Aluguel
Observações <i>in sample</i>	1796	614
Observações <i>out of sample</i>	13285	965
Total de observações	15081	1579
Período <i>in sample</i>	10/2012 a 05/2013	11/2012 a 05/2013
Período <i>out of sample</i>	05/2013 a 11/2013	05/2013 a 11/2013

Fonte: autor

Será utilizado um modelo ingênuo de previsão de preços como *benchmark*. Esta estratégia consiste em calcular a média do preço por metro quadrado para cada bairro dentro do período de calibragem e, posteriormente, utilizar estas médias para prever um imóvel com base em sua área. Exemplificando: supondo que a média do bairro Floresta para aluguel seja de R\$ 20,00/m² durante o período de calibragem, a estratégia de previsão seria multiplicar este valor pela metragem do imóvel que está sendo previsto. Nota-se que, esta estratégia não considera qualquer valorização, ou desvalorização, nos bairros e dos imóveis em geral, pois a medida de previsão será a mesma para todo o período fora da amostra. A estratégia pode ser resumida pela seguinte equação:

$$\hat{y}_i = a_{total} * \bar{y}, \quad (15)$$

em que \hat{y} refere-se à previsão do imóvel i , a_{total} a área total do imóvel previsto e \bar{y} , a média do bairro durante o período da amostra. Embora esta estratégia não seja documentada na literatura, acredita-se que ela reflete o comportamento de um investidor/comprador não sofisticado, que buscará informações passadas dos preços dos imóveis como previsores do valor futuro.

Como comparação ao modelo ingênuo, será estimado o modelo tradicional por MQO utilizando a matriz de covariância de White (1980) para controle da heteroscedasticidade. Não serão incluídas matrizes de dependência espacial, apenas as variáveis Área Total, Nº de Dormitórios, Nº de Garagens, uma constante e *dummies* para cada bairro. A variável Tempo no Mercado não foi incluída no modelo pois a sua relevância consiste na comparação do imóvel individualmente ao longo do tempo, não na previsão do mercado como um todo. A equação a ser estimada será:

$$y = \alpha + X\beta + D + \varepsilon, \quad (16)$$

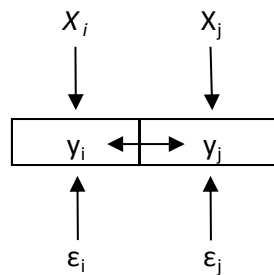
sendo y a matriz $n \times 1$ contendo a variável dependente, α a constante, X a matriz de regressores, β refere-se a matriz de coeficientes estimados, D é a matriz contendo *dummies* para cada um bairros utilizados nas estimações, com exceção ao bairro Centro Histórico, e ε refere-se aos resíduos da regressão.

Como comparação ao modelo MQO e ingênuo, serão estimados os modelos espaciais SAR e SEM, em que, no primeiro, as variáveis dependentes de regiões vizinhas interagem entre si. O modelo SAR a ser estimado pode ser definido como:

$$y = \alpha + \rho Wy + X\beta + \varepsilon, \quad (17)$$

em que ρ (*rho*) é o coeficiente de dependência espacial, W é a matriz $n \times n$ contendo as interações espaciais. Wy é um vetor n por 1 e o restante da notação permanece a mesma. Almeida (2012) comenta a restrição de que $|\rho| < 1$ para manter a estrutura de estacionariedade da série. A interpretação do parâmetro ρ é intuitiva, se o valor for positivo, temos uma correlação espacial positiva, isto é: um valor alto de y em regiões vizinhas aumenta o valor de y em outras regiões e a relação contrária caso o coeficiente seja negativo. Caso o coeficiente ρ não seja estatisticamente diferente de zero, pode-se considerar que não há correlação espacial no modelo, tratando-se de um modelo tradicional. A Figura 3 mostra a dinâmica do modelo SAR.

Figura 3 – Modelo SAR



Fonte: ALMEIDA (2012).

De acordo com Almeida (2012), a estimação pode ser realizada por log-máxima verossimilhança conforme a equação abaixo.

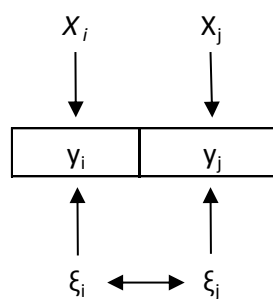
$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \sum \ln |I_n - \rho W| - \frac{\varepsilon' \varepsilon}{2\sigma^2} \quad (18)$$

Cabe ressaltar que, a construção da matriz W será feita de modo que cada observação tenha um número k de vizinhos, calculados a partir da distância euclidiana entre os pontos. A definição deste número se dará por uma adaptação do procedimento de Baumont (2004), explicitado na seção 3.2.1. Na estimação original maximiza-se a dependência espacial do

modelo, porém isto não é atrelado ao objetivo de previsão. Por isso, será escolhido o número de vizinhos que minimiza o Erro Percentual Médio da estimação dentro do período de calibragem.

No modelo SEM, estima-se o caso em que um conjunto de variáveis que não está na equação estrutural apresentam dependência espacial e afetam a variável dependente. Ou seja, há variáveis omitidas (ou um erro de medida), que estão dentro do termo de erro, embora a variável dependente não seja correlacionada espacialmente. A Figura 4 ajuda a entender melhor o modelo:

Figura 4 – Modelo SEM



Fonte: ALMEIDA (2012).

Pode-se formalizar o modelo como:

$$\begin{aligned}
 y &= X\beta + \xi \\
 \xi &= \lambda W\xi + \varepsilon,
 \end{aligned}
 \tag{19}$$

em que λ (lambda) é o parâmetro de erro autorregressivo espacial que acompanha a defasagem $W\xi$. Neste caso, o valor do erro autoregressivo é uma média dos erros nas regiões vizinhas mais um componente de erro aleatório, expresso por ε . Aqui também se faz necessária a condição de estacionariedade em que $|\lambda| < 1$. Caso essa condição seja violada, haveria um comportamento explosivo, pois os choques das regiões vizinhas não se dissipariam ao longo do espaço (ALMEIDA, 2012). A estimação deste modelo também será feita pela função de log-verossimilhança²⁷, definida:

²⁷ Kelejian e Prucha (1999) apontam que, mesmo não havendo normalidade dos resíduos, a estimação por Quase Máxima Verossimilhança ainda gera estimativas consistentes.

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |I_n - \lambda W| - \frac{1}{2\sigma^2} (y - X\beta)' (I_n - \lambda W)' (I_n - \lambda W) (y - X\beta) \quad (20)$$

Além das previsões geradas pelos modelos apresentados, será criada uma série baseada na combinação das previsões. Embora exista uma discussão na literatura em que se questiona se a ponderação de diferentes previsões pode oferecer melhores resultados do que uma simples média entre elas, foi realizada uma simples média aritmética entre todas as estimativas dos modelos ingênuo, MQO, SAR e SEM (GENRE *et al.*, 2013)²⁸.

Para evidenciar as diferenças entre os modelos, serão reportadas as previsões para os imóveis individualmente e para a média por metro quadrado dos bairros. Isso porque, um investidor ou possível comprador/locador pode também estar interessado em comparar os preços entre bairros de uma cidade. Para mensurar a eficácia destas previsões, serão usadas medidas como o erro percentual médio e mediano e erro percentual absoluto médio e mediano. Também será apresentada a amplitude dos erros de previsão com base nos erros percentuais máximos e mínimos para os bairros. As definições das estatísticas descritivas dos erros constam nas equações abaixo:

$$EPM_{\text{Médio}} = \frac{(\sum_i^N (\bar{y}_i - \hat{y}_i) / \bar{y}_i)}{n_{\text{bairros}}} \quad (21)$$

$$EPAM_{\text{Médio}} = \frac{(\sum_i^N |\bar{y}_i - \hat{y}_i| / \bar{y}_i)}{n_{\text{bairros}}} \quad (22)$$

$$EPM_{\text{Mediano}} = \frac{\text{mediana}(\bar{y}_i - \hat{y}_i)}{n_{\text{bairros}}} \quad (23)$$

$$EPM_{\text{Mediano}} = \frac{\text{mediana}|\bar{y}_i - \hat{y}_i|}{n_{\text{bairros}}} \quad (24)$$

$$\text{Amplitude} = \max(\bar{y}_i - \hat{y}_i / \bar{y}_i) - \min(\bar{y}_i - \hat{y}_i / \bar{y}_i) \quad (25)$$

²⁸ Também foram testadas combinações dos modelos utilizando como pesos os coeficientes de uma regressão entre o preço do imóvel dentro da amostra e as previsões geradas pelos modelos. No entanto, os resultados não foram satisfatórios, e, por isso, não são evidenciados aqui.

Nas equações acima, \bar{y}_i refere-se a média do bairro i , \hat{y}_i a média prevista do bairro i , $n_bairros$, ao número de bairros utilizados na previsão. De forma a conferir mais robustez à previsão de dados, serão realizados testes para comparar as previsões. Estes procedimentos são descritos na próxima seção.

5.3. Testes de previsão

É de interesse para *practitioners* e pesquisadores em geral definir qual modelo, dentre uma série de opções disponíveis, apresenta um melhor desempenho para prever um determinado objeto de estudo. Embora algumas medidas de erro como o Erro Absoluto Médio e a Média da Raiz dos Erros Quadráticos (*root mean squared errors* - RMSE, na sigla em inglês) ofereçam indícios sobre a acurácia de uma previsão, pode-se questionar se estes critérios levam a uma escolha ótima.

A partir disso, a literatura desenvolveu diversos testes para esta problemática. Diebold e Mariano (1995) criaram um dos primeiros testes, denominados como EPA (*equal predictive ability*). Neste trabalho, os autores testam a hipótese nula de que a diferença entre duas funções de perda (por exemplo, RMSE) é igual a zero. A equação abaixo define as variáveis.

$$\begin{aligned}d_t &= L_{1t} - L_{2t} \\ \mu &= E[d_t] \\ H_0: \mu &= 0\end{aligned}\tag{26}$$

Em que L_{it} refere-se a função perda de uma previsão de um determinado modelo i no tempo t , d_t a diferença dentre as funções perda, μ o valor esperado desta diferença. Caso os modelos prevejam de forma semelhante, a diferença entre as previsões não deve ser estatisticamente diferente de zero. Conforme Hansen *et al.* (2011) comentam, a utilização da diferença entre funções perda apresenta boas características, uma vez que, mesmo que a série que está sendo prevista apresente raiz unitária ou heteroscedasticidade, isto afetará as previsões de forma semelhante, fazendo com que a diferença entre elas provavelmente não apresente estes efeitos.

Embora o modelo de Diebold e Mariano (1995) seja de grande utilidade, ele responde a pergunta “estes modelos preveem da mesma forma?”, não sendo seu objetivo apresentar evidências se um modelo pode ser considerado melhor do que o outro. Com este intuito, Hansen (2005) desenvolveu o que se chamou na literatura como testes SPA – *superior predictive ability*. Este tipo de teste visa comparar uma série de previsões em relação a um *benchmark*. Neste procedimento, testa-se a hipótese nula de que a diferença entre os erros de uma série de modelos e um modelo *benchmark* é menor do que zero.

$$\begin{aligned}
 d_{kt} &= L_{0t} - L_{kt} & k &= 1, 2, \dots, m \\
 \mu_k &= E[d_{kt}] & k &= 1, 2, \dots, m \\
 H_0: \mu_t &\leq 0
 \end{aligned}
 \tag{27}$$

Em que a notação se mantém a mesma da equação 27, porém o teste abarca k previsões distintas em relação ao *benchmark*. Se o modelo de referência prevê melhor, a diferença entre ele e qualquer outro modelo deverá ser menor ou igual a zero. Desta fora, o teste responde a pergunta: “algum modelo prevê melhor que o *benchmark*?”, através de um teste de hipóteses composto. Hansen (2005) deriva estatísticas de teste em que a hipótese alternativa é de que pelo menos um dos modelos de comparação apresentam erros de previsão menores do que a referência. Este tipo de teste é útil principalmente quando há motivações teóricas para definir um modelo como *benchmark*, como ao testar a Hipótese dos Mercados Eficientes.

Caso o pesquisador não tenha evidências teóricas sobre qual modelo deve ser considerado o de referência, Hansen *et al.* (2011) desenvolveram um procedimento para testar a eficiência de uma série de previsões sem definir um *benchmark*. Considere um conjunto (set) M_0 com todas as previsões disponíveis e avaliadas a partir de uma função de perda $L_{i,t}$. A hipótese inicial do procedimento é a de que todas as previsões M_0 fazem parte do conjunto de melhores previsões (M^*), sendo necessário realizar testes para confirmar esta hipótese. À medida em que a hipótese inicial é rejeitada, são excluídas, sequencialmente, as piores previsões e o teste é realizado até que a hipótese nula não seja rejeitada. Para chegar a este resultado, definem-se as performances relativas dos modelos como $d_{ijt} = L_{i,t} - L_{j,t}$ para todo $i, j \in M_0$, em que $L_{i,t}$ refere-se a uma função perda da previsão i no tempo t . Defina-se $M^* \equiv \{i \in M_0: E[d_{ijt}] \leq 0 \forall j \in M_0\}$ como um subgrupo de objetos cujo a função perda

de i é menor do que a de j . Os autores denominam este subgrupo como *superior objects*. Portanto, as hipóteses do teste são:

$$\begin{aligned} H_{0M}: E[d_{ij,t}] &= 0 && \text{para todo } i, j \in M \\ H_{1M}: E[d_{ij,t}] &\neq 0 && \text{para algum } i, j \in M \end{aligned} \quad (28)$$

em que, caso a hipótese nula seja rejeitada, remove-se o modelo com pior performance preditiva. Inicialmente, define-se $M_0 = M$ e, à medida que a hipótese nula é rejeitada, os modelos são removidos de M . Este é um teste de equivalência a um nível α , em que o procedimento é repetido para cada par ij até que não se rejeite a hipótese nula, restando um conjunto de modelos com habilidade preditiva estatisticamente igual, sendo este o *Model Confidence Set*. Especificamente, calcula-se uma estatística t para cada previsão i , de forma que:

$$\begin{aligned} t_i &= \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}} && \text{para todo } i \in M \\ \bar{d}_i &= \bar{L}_i - \bar{L} \end{aligned} \quad (29)$$

em que \bar{L}_i é a média da função perda para a previsão i e \bar{L} é a média de todas as funções contidas em M , \bar{d}_i é o desvio da média da função perda i em relação à média de todas as demais. Para calcular $\widehat{var}(\bar{d}_i)$, realiza-se *bootstrap* em bloco²⁹. A partir da estatística t , computa-se a estatística de teste³⁰:

$$t_{max} = \max_{i \in M} t_i \quad (30)$$

O próximo passo é definir um critério de exclusão dos modelos. Os autores definem este critério como:

$$e_M = \arg \max_i t_i \quad (31)$$

Desta forma, é removido do grupo de previsões M aquela que mais contribui para o aumento da estatística t (ou seja, para a rejeição da hipótese nula). O critério e_M seleciona a

²⁹ Este procedimento considera a formação de blocos no processo de amostragem de forma a minimizar a incidência de dependência nas estimativas do *bootstrap*. Para uma revisão do tema, sugere-se Politis e Romano (1994).

³⁰ A derivação das propriedades assintóticas do teste encontra-se em Hansen *et al.* (2011).

previsão que apresenta a maior perda em relação à média de todos os modelos (HANSEN *et al*, 2011). Para tornar a compreensão mais clara, demonstra-se um passo-a-passo do procedimento para os testes. Calculam-se:

- As funções perda $L_{i,t}$ para todo $i = 1, \dots, m$ e $t = 1, \dots, n$.
- As médias de perda para cada modelo $\bar{L}_i \equiv \frac{\sum_{t=1}^n L_{i,t}}{n}$
- As variáveis correspondentes por bootstrap $L_{b,i,t}^* = L_{i,\tau_{b,t}}$ para $b=1, \dots, B$, $i=1, \dots, m$ e $t=1, \dots, n$, em que B refere-se ao número de amostragens. A partir disso, calculam-se as médias a partir de $\bar{L}_{b,i}^* = \frac{\sum_{t=1}^n L_{b,i,t}^*}{n}$
- Os desvios da variável gerada pelo *bootstrap* em relação à média amostral da função de perda: $\bar{\zeta}_{b,i}^* \equiv \bar{L}_{b,i}^* - \bar{L}_i$.
- As médias de erros amostrais e via bootstrap: $\bar{L} = \frac{1}{m} \sum_{i=1}^m \bar{L}_i$ e $\zeta_b^* = \frac{1}{m} \sum_{i=1}^m \zeta_{b,i}^*$
- A variância das estimativas por bootstrap: $\widehat{var}(\bar{d}_i) \equiv \frac{1}{B} \sum_{b=1}^B (\zeta_{b,i}^* - \zeta_b^*)^2$
- A estatística t para cada i e o seu máximo: $t_i \equiv \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}}$; $T_{max} = \max_{i \in M} t_i$
- A estimativa para a distribuição empírica de T_{max} : $T_{D,b}^* = \max_{i \in M} t_{b,i}^*$ para $b=1, 2, \dots, B$
e $t_{b,i}^* \equiv \frac{(\zeta_{b,i}^* - \zeta_b^*)}{\sqrt{\widehat{var}(\bar{d}_i)}}$

A partir disso, calcula-se o *p-value* para o teste:

$$\hat{p}(m) = \frac{1}{B} \sum_{b=1}^B 1_{\{T_{max} > T_{max,b}^*\}} \quad (32)$$

Se $\hat{p}(m) < \alpha$, em que α refere-se ao nível de significância do teste, então a hipótese nula é rejeitada e a previsão i é removida de M . Repete-se o processo até que a hipótese nula não seja rejeitada, gerando o *model confidence set*.

6. RESULTADOS

Esta seção será subdividida em: I) estatísticas descritivas, II) impacto do tempo no mercado, III) definição da matriz de pesos espaciais, IV) calibragem dos modelos de previsão, V) previsão para fora da amostra e VI) estimação do *model confidence set*.

6.1. Estatísticas descritivas

A Tabela 5 mostra as estatísticas descritivas das variáveis escolhidas utilizando as bases de dados restritas para venda e para locação. Em ambas as categorias, nota-se que a distribuição de preços tem assimetria positiva, em que as médias, tanto de preço quanto de preço por metro quadrado, são superiores às respectivas medianas. Percebe-se que as médias de área total e de número de dormitórios para venda é superior às respectivas medidas de aluguel. Isto é esperado, pois espera-se que uma família procurando um imóvel para comprar provavelmente irá adquirir um apartamento ou casa maior. Além disso, nota-se um alto desvio-padrão, tanto para venda quanto para locação de imóveis. Isto também está de acordo com as expectativas, uma vez que há uma grande amplitude de preços demonstrada pelos valores mínimos e máximos. Um exemplo disso são os valores máximos de venda e de aluguel, que são, pelo menos, dez vezes maiores do que os respectivos valores mínimos.

Tabela 5 - Estatísticas descritivas

Painel A - Venda

		Média	Mediana	Desvio-padrão	Mínimo	Máximo	Nº obs	Nº de bairros
Preço	R\$	369.598,10	R\$ 292.377,94	R\$ 258.918,76	R\$ 61.981,11	R\$ 1.782.962,84	15081	18
Preço por m ²	R\$	3.173,87	R\$ 3.033,91	R\$ 890,99	R\$ 434,60	R\$ 7.943,29	15081	
Área total		114,75	98,00	66,00	22,00	432,00	15081	
Nº de dormitórios		2,18	2,00	0,79	1,00	5,00	15081	
Nº de vagas na garagem		0,90	1,00	0,83	0,00	5,00	15081	
Tempo no mercado (dias)		16,95	9,00	20,22	1,00	90,00	15081	
Taxa de Financ. Imob.		12,60%	12,34%	0,61%	11,93%	14,00%	15081	

Painel B - Aluguel

		Média	Mediana	Desvio-padrão	Mínimo	Máximo	Nº obs	Nº de bairros
Preço	R\$	1.704,80	R\$ 1.334,98	R\$ 1.261,54	R\$ 422,90	R\$ 8.449,30	1579	15
Preço por m ²	R\$	17,73	R\$ 16,45	R\$ 5,92	R\$ 7,24	R\$ 43,22	1579	
Área total		94,51	81,00	53,28	26,00	358,00	1579	
Nº de dormitórios		1,76	2,00	0,81	1,00	5,00	1579	
Nº de vagas na garagem		0,56	0,00	0,64	0,00	4,00	1579	
Tempo no mercado (dias)		13,67	7,00	18,36	1,00	90,00	1579	
Taxa de Financ. Imob.		12,79%	12,84%	0,58%	11,93%	14,00%	1579	

*Preços deflacionados para valores de Outubro de 2012.

Fonte: autor.

6.2. Impacto do tempo no mercado

Conforme definido na seção 5, para todos os imóveis que tiveram variação em seu preço anunciado, foi regredida a série de preços nominais contra uma constante e seu tempo no mercado. A Tabela 6 reporta as estatísticas descritivas.

Tabela 6 - Tempo de publicação e variação nos preços

Variável	Venda	Locação
Ajuste médio	0,009	-0,096
Inflação média	0,008	0,015
Nº observações - 1 ajuste	1148	175
Nº observações - 2 ajustes	168	25
Nº observações - >=3 ajustes	39	7
% com reajuste > 0	0,597	0,077
% com reajuste < 0	0,403	0,923
Tempo médio até o primeiro reajuste	38,076	38,250
Total de imóveis repetidos	1345	207
Total de observações	15178	2355

Fonte: autor

Notam-se diferenças entre os mercados de venda e de locação. Isto torna-se explícito ao analisar os percentuais de observações com reajustes maiores e menores do que zero. Na venda, 40,3% dos imóveis tiveram seu preço nominal reajustado para baixo, em comparação a 92,3% no mercado de locação. Isto também é ilustrado pela média percentual de ajuste, que, no mercado de aluguel, foi de -9,6%, enquanto no mercado de venda, este valor se aproxima de 1%.

As evidências encontradas no mercado de aluguel parecem estar de acordo com os trabalhos de Taylor (1999), Knight (2002), Merlo *et al.* (2015), os quais encontraram que um maior tempo no mercado implica em um reajuste para baixo no valor do imóvel. Este efeito pode ser atribuído ao custo de capital de um investidor/proprietário para o mercado de aluguel ser mais alto, bem como sua demanda por liquidez. Isso porque, a locação de um imóvel é um processo mais rápido e necessita de um montante de capital menor do que para venda, portanto, aqueles que necessitam de uma renda constante podem optar por este tipo de investimento. Baseado nas informações da Tabela 6, acredita-se que, para acelerar o processo de locação, os proprietários diminuem o valor de seus imóveis para atrair inquilinos e, com isso, ter liquidez e minimizar os custos incorridos com despesas como o condomínio.

Como os contratos de aluguel geralmente incluem uma cláusula para reposição de inflação, os preços no período de anúncio não são reajustados para corrigi-la, mas sim para que o imóvel seja locado mais rapidamente. Isso fica evidente pelo fato do reajuste médio nos preços nominais ser negativo, já que a perda real (considerando a inflação) seria maior ainda. Além disso, os proprietários podem não querer que o imóvel permaneça muito tempo no mercado, de forma a ser criado um estigma de má qualidade (TAYLOR, 1999).

No mercado de venda, embora haja uma tendência para valorização dos imóveis com o passar do tempo, os resultados são menos conclusivos. Isso porque, há uma parcela relativamente alta (40%) que apresenta um ajuste negativo. Como o fluxo de caixa gerado pela venda de um imóvel só ocorre uma vez e, portanto, não está protegido de perdas inflacionárias, pode-se supor que vendedores reajustam seus preços para repor estas perdas e aumentar seu ganho de capital. Isto é evidenciado na Tabela 6: o reajuste médio é maior do que a inflação média evidenciada no período que os imóveis estiveram sendo anunciados.

6.3. Definição da matriz de pesos espaciais

Para escolher o número k de vizinhos a serem utilizados nos modelos de previsão, a Tabela 7 mostra os resultados para o erro absoluto médio (em %), estatística I de Moran e seu respectivo p -value na construção da matriz W ³¹.

³¹ Foram testados até 100 vizinhos sem alteração dos resultados reportados. Para manter a exposição concisa, são mostrados apenas os 30 primeiros.

Tabela 7 - Definição dos k vizinhos

Vizinhos	Venda			Aluguel		
	EAM (%)	I de Moran	p-value	EAM (%)	I de Moran	p-value
2	0,2130	0,3722	0,000	0,2526	0,1509	0,000
3	0,2104	0,3544	0,000	0,2508	0,1507	0,000
4	0,2133	0,3990	0,000	0,2524	0,1951	0,000
5	0,2150	0,4017	0,000	0,2497	0,1625	0,000
6	0,2203	0,3882	0,000	0,2499	0,1553	0,000
7	0,2214	0,3664	0,000	0,2497	0,1503	0,000
8	0,2202	0,3438	0,000	0,2509	0,1767	0,000
9	0,2212	0,3377	0,000	0,2492	0,1505	0,000
10	0,2192	0,3244	0,000	0,2490	0,1480	0,000
11	0,2184	0,3240	0,000	0,2496	0,1430	0,000
12	0,2141	0,3165	0,000	0,2496	0,1402	0,000
13	0,2147	0,3253	0,000	0,2500	0,1539	0,000
14	0,2139	0,3273	0,000	0,2495	0,1558	0,000
15	0,2149	0,3348	0,000	0,2474	0,1436	0,000
16	0,2157	0,3293	0,000	0,2471	0,1425	0,000
17	0,2143	0,3330	0,000	0,2472	0,1439	0,000
18	0,2152	0,3317	0,000	0,2470	0,1393	0,000
19	0,2135	0,3114	0,000	0,2467	0,1331	0,000
20	0,2145	0,3179	0,000	0,2464	0,1272	0,000
21	0,2153	0,3227	0,000	0,2463	0,1256	0,000
22	0,2148	0,3294	0,000	0,2454	0,1138	0,000
23	0,2135	0,3337	0,000	0,2452	0,1040	0,000
24	0,2130	0,3325	0,000	0,2451	0,1060	0,000
25	0,2161	0,3314	0,000	0,2455	0,1077	0,000
26	0,2181	0,3320	0,000	0,2464	0,1043	0,000
27	0,2195	0,3363	0,000	0,2469	0,1057	0,000
28	0,2196	0,3333	0,000	0,2476	0,1083	0,000
29	0,2186	0,3308	0,000	0,2476	0,1028	0,000
30	0,2187	0,3306	0,000	0,2473	0,0991	0,000

*Nº de observações:1796 (venda) e 614 (aluguel)

Fonte: autor.

Segundo o critério adotado, o número de vizinhos escolhidos para as previsões dos modelos espaciais é de 3 e de 24 para venda e para aluguel, respectivamente. A utilização do desempenho dentro da amostra como *proxy* para o desempenho fora dela parece ser mais apropriado do que o critério de Baumont (2004). Nota-se que este critério levaria a escolhas diferentes para o número de vizinhos da matriz W, a qual seria feita com base no índice de Moran que maximiza a dependência espacial. Embora as diferenças sejam pequenas, a Tabela

7 mostra que os números de vizinhos que maximizam a dependência espacial geram erros maiores dentro da amostra.

Para tornar mais robusta a seleção dos vizinhos, foi realizado o teste do Multiplicador de Lagrange, MLp, para as matrizes W. Conforme definido na seção 3.3.2, este procedimento visa testar a hipótese nula de que não há dependência espacial, estimando a significância do coeficiente ρ . A Tabela 8 mostra os resultados:

Tabela 8 - Teste LM

	k = 3 (Venda)	k = 24 (Aluguel)
Estatística LM	1139,44***	107,96***
Valor crítico a 1%	6,635	6,635
Nº variáveis	6	6
Nº observações	1796	614

***: $p < 0.01$

Fonte: autor.

Nota-se que, em ambos os mercados, há evidência de que os coeficientes de dependência espacial são diferentes de zero com alto nível de significância, justificando o uso de modelos espaciais para a previsão de preços.

6.4. Calibragem dos modelos de previsão

Esta seção busca saber como os diferentes atributos de um imóvel formam o seu preço e, a partir esta estimação, realizar a previsão de preços. Com base nos modelos descritos na seção 5.2, são apresentados os resultados para venda (locação) no período de Outubro de 2012 (Novembro de 2012) a Maio de 2013. O restante do período (de Maio a Novembro de 2013) foi reservado para previsão. A Tabela 9 reporta os coeficientes para as estimações por MQO (*dummies* de bairro suprimidas) e pelos modelos SAR e SEM.

Tabela 9 - Estimacões in-sample

Variável	Venda			Aluguel		
	MQO	SAR	SEM	MQO	SAR	SEM
Constante	11,859*** (0,000)	8,882*** (0,000)	11,853*** (0,000)	5,240*** (0,000)	4,678*** (0,000)	5,493*** (0,000)
Área Total	0,005*** (0,000)	0,005*** (0,000)	0,005*** (0,000)	0,006*** (0,000)	0,007*** (0,000)	0,006*** (0,000)
Nº Dormitórios	0,139*** (0,000)	0,118*** (0,000)	0,141*** (0,000)	0,073*** (0,008)	0,075*** (0,009)	0,062** (0,022)
Nº Garagens	0,180*** (0,000)	0,143*** (0,000)	0,169*** (0,000)	0,191*** (0,000)	0,222*** (0,000)	0,202*** (0,000)
Taxa Financ. Imob.	-0,022* (0,053)	-0,024*** (0,000)	-0,021*** (0,000)	0,079** (0,027)	0,092** (0,057)	0,072*** (0,000)
Tempo de Publicaçã	0,000 (0,928)	0,000 (0,984)	0,000 (0,187)	-0,000 (0,527)	-0,000 (0,232)	-0,000 (0,583)
Rho\Lambda	- -	0,247*** (0,000)	0,367*** (0,000)	- -	0,065*** (0,000)	0,691*** (0,000)
R ² médio	0,791	-	-	0,736	-	-
Nº obs	1796	1796	1796	614	614	614
Log-Verossimilhança	-	519,720	543,323	-	42,348	64,201

*Níveis de significância: 1%: ***, 5%: ** e 10%: *. P-values em parênteses.

Fonte: autor

Tanto para compra quanto para locaçã os modelos apresentam a maioria dos coeficientes com sinais e significâncias esperadas. A área total eleva os preços dos imóveis à venda e à locaçã, bem como o número de vagas disponíveis na garagem. Isto está de acordo com resultados de outros estudos que já analisaram estas variáveis (ALVES *et al.*, 2012; HERMANN e HADDAD, 2008; MACEDO 1998). Estes coeficientes mantêm-se significativos e com valores próximos em todos os modelos estimados e nas duas categorias analisadas (venda e locaçã). A variável que mede o tempo em que um imóvel está sendo comercializado não apresentou coeficientes significativos, contrariando a literatura. Seguindo Alves *et al.* (2012), que analisa impactos financeiros na precificaçã de imóveis paulistas, a taxa de financiamento imobiliário apresenta resultados distintos entre os mercados. No mercado de venda, há uma relação negativa entre esta variável e o preço anunciado do imóvel. Isto é esperado, uma vez que, com a diminuiçã da taxa de financiamento imobiliário, os imóveis tornam-se mais acessíveis a compradores com restriçã financeira. Como isto deve levar a um aumento na demanda, espera-se uma elevaçã nos preços. A relação é inversa no mercado de aluguel: com o aumento do custo do financiamento para compra, os interessados em

comprar um imóvel, mas que não podem arcar com o custo do financiamento, podem ter que migrar para o mercado de locação.

Além dos resultados já expostos, nota-se que em ambos os modelos espaciais os coeficientes de dependência espacial são positivos e significativos para os mercados de venda e de locação. No entanto, as interpretações destes coeficientes são diferentes: com valores positivos no modelo SAR (0,247 e 0,065 para venda e locação, respectivamente), imóveis vizinhos com alto preço impactam positivamente seus respectivos preços. O contrário também é válido, em que imóveis com baixo preço tendem a impactar negativamente o preço de seus vizinhos. No modelo SEM, os coeficientes apresentaram valores maiores do que no modelo SAR (0,367 e 0,691, venda e locação, respectivamente), evidenciando que, além da dependência espacial na variável dependente, há outras variáveis não modeladas que impactam o preço anunciado dos imóveis. Isto pode ser considerado esperado, uma vez que variáveis demográficas como renda média por bairro, escolaridade dos habitantes do bairro, dentre outras, não estão disponíveis.

6.5. Previsão fora da amostra

A exposição dos resultados se dividirá em duas análises: da habilidade dos modelos em prever o valor de cada imóvel individualmente e das previsões para os preços por metro quadrado por bairro. Para a primeira análise, a Tabela 10 mostra as estimativas para o Erro Absoluto Médio Percentual para os cinco modelos testados para venda e para locação.

Tabela 10 - Erro absoluto médio (%)

Modelo	Venda	Aluguel
Ingênuo	19,20%	28,27%
MQO	18,53%	25,31%
SAR	21,63%	27,27%
SEM	22,04%	25,46%
Comb	18,38%	25,41%
Obs	13285	965

Fonte:autor.

No mercado de venda, ao prever o valor específico de cada imóvel, a combinação de estimações gerou erros de previsão percentuais menores (18,38%) do que os demais modelos.

Nota-se que a diferença é pequena em relação ao modelo estimado por MQO (18,53%). Para o mercado de aluguel, os modelos MQO e a combinação de previsões apresentaram os menores erros (25,31% e 25,41%, respectivamente). Assim como no mercado de venda, a diferença entre estes modelos é pequena. Além disso, em ambos os tipos de imóveis o *benchmark* apresentou desempenhos inferiores às demais estimações.

Utilizando os coeficientes estimados na seção anterior, são reportadas, na Tabela 11, as estatísticas descritivas para os erros de previsão dos modelos ingênuo, MQO, SAR, SEM e a combinação das previsões. Os percentuais reportados são relativos ao erro percentual para a previsão do metro quadrado por bairro.

Tabela 11 - Erros de previsão em relação aos bairros

	Venda					Aluguel				
	Ingênuo	MQO	SAR	SEM	Comb	Ingênuo	MQO	SAR	SEM	Comb
EA Médio (%)	7,80%	9,88%	13,87%	15,64%	9,25%	6,11%	7,58%	9,86%	11,81%	7,17%
E Médio (%)	7,77%	9,08%	1,48%	1,09%	4,86%	-3,58%	-1,03%	-6,14%	-1,45%	-3,05%
EA Mediano (%)	6,29%	7,98%	11,28%	11,96%	7,94%	2,80%	7,11%	6,88%	12,00%	6,66%
E Mediano (%)	6,29%	7,98%	4,89%	6,79%	7,63%	-0,75%	3,69%	-2,10%	2,84%	-1,61%
E % Máximo	20,45%	20,34%	22,15%	24,14%	17,11%	6,04%	10,15%	11,41%	16,92%	8,19%
E % Mínimo	-0,30%	-5,31%	-60,76%	-67,25%	-29,24%	-19,59%	-13,45%	-40,36%	-32,60%	-24,26%
Amplitude	20,75%	25,65%	82,91%	91,39%	46,35%	25,63%	23,60%	51,78%	49,51%	32,45%
Observações	15081	15081	15081	15081	15081	1579	1579	1579	1579	1579
Nº bairros	18	18	18	18	18	15	15	15	15	15
Nº vizinhos (k)			3	3				24	24	

*Período de previsão: Maio a Novembro de 2013

Fonte: autor

Notam-se resultados semelhantes entre os tipos de mercados. Para a venda, individualmente nenhum modelo apresentou resultados consistentemente melhores do que a estratégia ingênua. Os modelos espaciais, SAR e SEM, geraram erros absolutos médios acima dos 10% (13,87% e 15,64%, respectivamente), apresentando desempenho inferior à estratégia ingênua (7,80%). O comportamento das previsões é semelhante em relação ao erro absoluto mediano, em que os modelos espaciais apresentaram valores acima de 10% e a estratégia ingênua obteve o menor valor (6,29%). Nota-se que o modelo MQO apresentou erros absolutos médios (9,88%) e medianos (7,98%) maiores do que a estratégia ingênua, além de apresentar uma amplitude dos erros maior (25,65%) do que na estratégia ingênua (20,75%). A combinação de previsões apresenta um desempenho levemente melhor que a o modelo MQO (diferença de 0,63% no erro absoluto médio). Embora a combinação de previsões apresente, neste aspecto, um desempenho melhor, a amplitude do erro ainda assim mostrou-se considerável entre os bairros (46,35%) sendo mais do que o dobro do valor do *benchmark*. Nota-se que a estratégia ingênua fez uma subvalorização, tendo em vista que os erros não são simetricamente distribuídos ao redor de zero: o erro máximo é de 20,45% enquanto o mínimo de -0,31%. Isto é esperado, já que se trata de um modelo que não contempla um fator de valorização dos imóveis. Nas abordagens espaciais constatou-se o contrário: uma sobrevalorização, à medida em que os erros mínimos possuem valores considerados altos (-60,76% e -67,25% para os modelos SAR e SEM).

Os resultados para o mercado de aluguel apresentam uma menor discrepância entre os modelos. Em termos de erros absolutos médios e medianos, a estratégia ingênua apresentou valores mais baixos (6,11% e 2,80%) do que os demais modelos econométricos. No entanto, em comparação com a previsão de venda, nota-se que as diferenças entre as medidas de erro entre os modelos são menores. O modelo SAR, por exemplo, apresentou estimativas de erros absolutos médios (9,86%) mais próximos do modelo ingênuo e mais baixos do que o modelo MQO (7,58%) e SEM (11,81%). Em relação aos demais modelos econométricos, a combinação de todas as previsões minimizou os erros absolutos médios e medianos para 7,17% e 6,66%. No entanto, o desempenho ainda é inferior ao modelo ingênuo. A combinação de previsões também não reduz o a amplitude dos erros (32,45%), em comparação com o menor valor (23,60% no modelo

ingênuo). Detalhando os dados expostos previamente, a Tabela 12 mostra os erros de previsão para o preço por metro quadrado segmentados pelos respectivos bairros.

Tabela 12 - Erros de previsão por bairro

Painel A - Venda								
	Média Real	Ingênuo	MQO	SAR	SEM	Comb	Obs	
Auxiliadora	R\$ 3.558,59	6,50%	2,52%	2,08%	19,36%	7,62%	810	
Azenha	R\$ 3.165,18	1,42%	7,28%	15,11%	6,74%	7,64%	330	
Bela Vista	R\$ 3.987,79	1,73%	-5,31%	15,15%	21,42%	8,25%	926	
Boa Vista	R\$ 3.748,55	-0,30%	5,35%	22,15%	24,14%	12,83%	596	
Cavanhada	R\$ 2.616,60	9,31%	7,60%	-13,33%	-20,46%	-4,22%	856	
Centro Histórico	R\$ 3.052,02	0,77%	8,36%	7,69%	6,83%	5,91%	1240	
Cidade Baixa	R\$ 3.387,44	2,51%	7,31%	17,91%	13,29%	10,26%	1038	
Cristal	R\$ 2.808,75	3,75%	8,53%	-6,43%	-9,22%	-0,84%	991	
Cristo Redentor	R\$ 2.940,25	9,21%	6,92%	-10,81%	-1,69%	0,91%	919	
Farroupilha	R\$ 3.055,82	2,61%	14,48%	11,66%	6,72%	8,87%	116	
Floresta	R\$ 3.119,24	7,80%	16,40%	1,65%	10,62%	9,12%	619	
Jardim Botânico	R\$ 3.489,75	13,27%	10,69%	17,21%	15,91%	14,27%	568	
Jardim Carvalho	R\$ 2.569,68	5,37%	-1,85%	-7,92%	-16,56%	-5,24%	290	
Jardim do Salso	R\$ 2.688,75	20,45%	18,79%	-10,14%	-7,16%	5,49%	252	
Jardim Itú-Sabará	R\$ 2.950,88	15,76%	15,31%	-2,15%	-8,61%	5,08%	1007	
Passo D'Areia	R\$ 3.516,19	16,19%	20,34%	16,71%	15,22%	17,11%	1390	
Rubem Berta	R\$ 2.080,83	6,09%	4,97%	-60,76%	-67,25%	-29,24%	304	
Vila Ipiranga	R\$ 3.347,58	17,43%	15,78%	10,90%	10,36%	13,61%	1033	
Painel B - Aluguel								
Auxiliadora	R\$ 19,72	5,34%	4,76%	4,88%	13,21%	7,05%	43	
Bela Vista	R\$ 21,48	-0,75%	-11,36%	0,08%	15,44%	0,85%	71	
Bom Fim	R\$ 19,40	6,04%	4,24%	1,67%	5,89%	4,46%	40	
Cavanhada	R\$ 12,77	-16,07%	-8,00%	-40,36%	-32,60%	-24,26%	22	
Centro Histórico	R\$ 16,17	2,39%	6,91%	-6,88%	-9,54%	-1,78%	139	
Cidade Baixa	R\$ 18,29	2,80%	10,15%	7,97%	5,72%	6,66%	37	
Cristo Redentor	R\$ 14,01	-10,70%	-10,84%	-28,88%	-23,85%	-18,57%	30	
Floresta	R\$ 15,42	-0,56%	6,04%	-7,10%	-4,84%	-1,61%	29	
Menino Deus	R\$ 17,04	-14,12%	-10,55%	-4,72%	-1,80%	-7,80%	101	
Partenon	R\$ 15,75	-6,95%	-10,35%	-13,64%	-12,26%	-10,80%	58	
Passo D'Areia	R\$ 17,64	0,26%	7,11%	0,05%	2,84%	2,57%	73	
Petrópolis	R\$ 18,34	-2,71%	3,69%	-2,10%	5,74%	1,16%	148	
Rio Branco	R\$ 19,60	-19,59%	-13,45%	1,83%	12,00%	-4,81%	69	
Santana	R\$ 15,40	2,13%	0,64%	-16,35%	-14,59%	-7,04%	55	
Vila Ipiranga	R\$ 20,09	-1,19%	5,62%	11,41%	16,92%	8,19%	50	

Fonte: autor.

No Painel A (venda), nota-se que as previsões não mantêm um padrão, nem apresentam um modelo mais favorável para a maioria dos bairros. No entanto, é possível

perceber erros discrepantes para os modelos espaciais (SAR e SEM) no bairro Rubem Berta, variando de -60,76% a -67,25%. Uma característica que pode levar a este resultado é a distância do centro histórico. Acredita-se que este baixo desempenho possa ser atribuído a falta de vizinhos realmente próximos do bairro Rubem Berta. Como a construção da matriz W força a utilização dos 3 vizinhos mais próximos, mesmo que um vizinho seja, na verdade, um bairro distante, ele será considerado na estimação, o que pode gerar estimativas incorretas. No Painel B (aluguel), também nota-se um erro maior para os modelos espaciais em bairros distantes do centro da capital, como Cavahada e Cristo Redentor (erros de -40,36% a -32,60%).

Com base nos dados das Tabelas 10,11 e 12 não é possível inferir se algum modelo é, de forma consistente, superior aos outros em termos de previsão. Embora algumas métricas sejam favoráveis ao modelo ingênuo, outras métricas não o são. Por isso, a seção seguinte realizará testes econométricos para se ter uma melhor compreensão da eficácia dos modelos.

6.6. Model Confidence Set

A Tabela 13 mostra os *p-values* para a estimação do *Model Confidence Set* das previsões de preços de imóveis para venda e para locação, utilizando o erro quadrático como função perda. Para a construção das estatísticas de teste foram utilizadas 10.000 amostragens.

Tabela 13 - p-values para estimação do MCS

	Venda					Aluguel				
	Ingênuo	MQO	SAR	SEM	Comb	Ingênuo	MQO	SAR	SEM	Comb
$\alpha = 0,1$	1,000	0,000	0,000	0,000	0,000	1,000	0,007	0,001	0,382	0,382
$\alpha = 0,05$	1,000	0,000	0,000	0,000	0,000	1,000	0,006	0,001	0,376	0,376
$\alpha = 0,01$	1,000	0,000	0,000	0,000	0,000	1,000	0,006	0,001	0,380	0,380
Amostragens	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000

Fonte: autor.

Nota-se que os resultados se alteram pouco devido à mudança dos níveis de significância testados (α). Para o mercado de venda, em todos os níveis de confiança do

teste apenas o modelo ingênuo está contido no *MCS*. No mercado de locação, nota-se que, a um nível de significância de 1%, o procedimento define os modelos ingênuo, SEM e a combinação como os melhores modelos para previsão.

Estes resultados confirmam as evidências anteriores em que, o modelo ingênuo e a combinação de previsões mostraram-se como os modelos que minimizaram os erros de previsão do metro quadrado dos bairros. Embora os resultados para as previsões individuais direcionem para conclusões diferentes, atribui-se isto a amplitude dos erros de previsão. No mercado de aluguel, além do modelo ingênuo, a combinação de previsões está no *MCS* nos três níveis de significância. Este resultado está de acordo com Stock e Watson (1999) e Koop e Korobilis (2012), que encontram melhores resultados para a previsão da inflação americana a partir de uma combinação de diferentes modelos.

7. CONSIDERAÇÕES FINAIS

Utilizando dados de anúncios de imóveis para os mercados de venda e de locação de Porto Alegre, o presente trabalho teve como objetivo analisar a formação de preços destes mercados. Isto foi realizado de duas maneiras: comparando diferentes modelos para previsão fora da amostra e estimando o efeito do tempo no mercado no preço dos imóveis. Para o primeiro tópico, utilizou-se uma abordagem ingênua como *benchmark* em comparação a uma estimação por MQO, dois modelos espaciais e uma combinação das demais estimativas. Nota-se que, para a previsão dos preços de imóveis individualmente, os erros de previsão fora da amostra podem ser considerados altos, em cerca de 25% tanto para venda quanto para locação. Porém, ao calcular o preço por metro quadrado dos bairros, estes erros se vêm reduzidos para patamares de 7% a 9%. Dentre os modelos utilizados, dependendo da métrica utilizada, um modelo apresenta resultados melhores frente a outro, não tornando clara a escolha de um método mais adequado. Para solucionar isso, utilizou-se a abordagem de *Model Confidence Sets* de Hansen *et al.* (2011). Dado um número m de diferentes previsões, este procedimento testa a hipótese de cada previsão fazer parte do grupo de melhores previsões. Tanto para o mercado de venda quanto para o de locação, apenas o modelo ingênuo estava incluída neste grupo em todos os níveis de confiança do teste.

Para o segundo tópico, os resultados sugerem que imóveis que possuem o seu preço reajustado apresentam diferentes tendências entre o mercado de venda e de aluguel. No mercado de venda, os resultados são mistos: em cerca de 60% dos casos, o reajuste é feito de forma a aumentar o preço de anúncio do imóvel, enquanto os restantes 40% diminuem seu valor. Há evidências de que este reajuste é feito de forma a compensar perdas atribuídas à inflação. No entanto, para o mercado de aluguel os resultados são distintos: de todos os imóveis que tiveram seu preço reajustado durante a amostra, mais de 90% diminuíram o valor de oferta. Acredita-se que esta explicação se deva a demanda por liquidez dos proprietários, pois caso o imóvel esteja vago, os custos de manutenção são gastos pelo proprietário até que o imóvel seja locado. Este trabalho é o primeiro estudo brasileiro que aborda este assunto, apesar de uma extensa literatura estrangeira.

7.1. Limitações e sugestões de pesquisas futuras

Algumas limitações circundam a literatura que analisa o mercado imobiliário. A localização considerada nas estimações que usaram distâncias foi a dos centroides dos bairros até os pontos definidos³². Isto limita a utilização de algumas técnicas econométricas que fazem uso imprescindível da localização do imóvel, como por exemplo, os modelos geoestatísticos e incluem um ruído nas estimações, pois todos os imóveis de um bairro serão considerados em uma mesma localização (o centroide).

Embora os resultados para o objetivo de previsão apresentem *insights* relevantes para os estudos da área, técnicas mais atuais como modelos espaciais em painel (ELHORST, 2014; LESAGE, 1999) e modelos espaço-temporais (LIU, 2012) não puderam ser aplicados nas bases de dados disponíveis devido ao curto horizonte temporal e irregularidade na coleta dos dados. Além disso, com o aprimoramento de algoritmos de coleta de dados em portais imobiliários, acredita-se que coletando dados em uma frequência equilibrada é possível construir uma série temporal adequada e, com isso, aplicar estes métodos mais recentes na estimação e previsão dos preços de imóveis.

Para o impacto do tempo de mercado, acredita-se que as evidências apresentadas sejam de ordem inicial nos estudos desta área. De maneira semelhante a Knight (2002) e Chen e Rutherford (2012), o uso de modelos de escolha binária para determinar que características tornam o imóvel mais propenso a ter seu valor reajustado pode ser utilizado com os dados desta pesquisa. Além disso, há *papers* trabalhando estratégias e táticas para a inserção de um valor de oferta adequado no momento de divulgar um imóvel (YAVAS e YANG, 1995; BERACHA e SEILER, 2014; CARDELLA e SEILER, 2016). Estas estratégias têm dois objetivos não excludentes: tornar o processo de venda/locação mais rápido e maximizar a lucratividade do proprietário do imóvel. Tais estudos podem ser conduzidos de forma a analisar estes fenômenos pouco estudados no mercado imobiliário brasileiro.

³²Uma vez que o endereço exato não era disponível. Problema também documentado em FURTADO (2007).

REFERÊNCIAS

- Adelino, M.; Schoar, A.; Severino, F. **House prices, collateral, and self-employment**. Journal of Financial Economics. 2015
- Agnello, L.; Schuknecht, L. **Booms and busts in housing markets: determinants and implications**. Journal of Housing Economics, Vol. 20, pp.171-190. 2011
- Akerlof, George A. **The market for "lemons": Quality uncertainty and the market mechanism**. The Quarterly Journal of Economics, p. 488-500, 1970.
- Almeida, E. **Econometria Espacial Aplicada**. Campinas, SP. Editora Alínea, 2012.
- Alves, D. C., Yoshino, J. A., Pereda, P. C., Amrein, C. **Modelagem dos Preços de Imóveis Residenciais Paulistanos**. Revista Brasileira de Finanças, Rio de Janeiro, Vol. 9 (2), pp. 167-187. 2011.
- ANBIMA. **Fundos de Investimento Imobiliário** Disponível em: <<http://portal.anbima.com.br/informacoes-tecnicas/estatisticas/fundos-de-investimento-imobiliario/Pages/estatisticas.aspx/>>. Acesso em 26 de Fevereiro de 2016.
- Anselin, L.; Rey, S. **Properties of tests for spatial dependence in linear regression models**. Geographical analysis, Vol. 23 , pp. 112-131. 1991
- _____.; Florax, R. J. **Small sample properties of tests for spatial dependence in regression models: Some further results**. In New directions in spatial econometrics. Springer Berlin Heidelberg pp. 21-74). 1995
- Bala, A P; Peeters, D; Thomas, I, **Spatial issues on a hedonic estimation of rents in Brussels**, Journal of Housing Economics, Vol 25, pp. 104-123, 2014.
- Baller, R D.; Anselin, L.; Messner, S F.; Deane, G.; Hawkins, D F. **Structural Covariates of US County Homicide Rates: Incorporating Spatial Effects**. Criminology, Vol. 39, pp. 561–588. 2001.
- Baltagi, B. H.; Bresson, G, **Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris**, Journal of Urban Economics, Vol 69, Issue 1, pp 24-42, 2011.

Baumont, C. **Spatial effects in housing price models: do housing prices capitalize urban development policies in the agglomeration of Dijon (1999)?** , 2004.

Beenstock M.; Feldman,D.; Felsenstein, D. **Testing for Unit Roots and Cointegration in Spatial Cross-Section Data**. Spatial Economic Analysis, Vol.7, pp. 203-222, 2012

Benson, E. D.; Hansen, J. L., Schwartz., A. L.; Smersh, G. T. **Pricing residential amenities: the value of a view**. Journal of Real State Finance and Economics, v. 16, n. 1, p. 55-73, 1998.

Beracha, E; Selier, M J. **The effect of listing price strategy on transaction selling prices**. The Journal of Real Estate Finance and Economics, v. 49, n. 2, p. 237-255, 2014.

Bourassa, S.; Cantoni, E.; Hoesli, M. **Spatial dependence, housing submarkets, and house price prediction**. Journal of Real Estate Finance and Economics, Vol. 35(2), pp. 143–160. 2007

Bowes, D. R.; Ihlanfeldt, K. R. **Identifying the impacts of rail transit stations on residential property values**. Journal of Urban Economics, Vol 50, pp. 1-25, 2001.

Braga, L.F.T.V. **Valoração de imóveis no Rio Grande do Sul: uma análise a partir de regressão espacial**. Dissertação de Mestrado do Programa de Pós-Graduação em Economia – Universidade Vale do Rio dos Sinos, São Leopoldo, 2010.

Brasington, D; Haurin, DR. **Educational Outcomes and House Values: A Test of the Value Added Approach**. Journal of Regional Science, Vol. 46, pp. 245–268. 2006.

Browning, M.; Gørtz, M.; Leth-Petersen, S. **Housing wealth and consumption: a micro panel study**. The Economic Journal, v123, n568, p.401-428. 2013

Brueckner, J. **The structure of urban equilibria: a unified treatment of the muth-mills model**. Handbook of Regional and Urban Economics. Vol. 2, pp. 821-845, 1987.

Butler, RV. **The specification of hedonic indexes for urban housing**. Land Economics, Vol. 58, pp.96-108, 1982.

Cameron, AC.; Trivedi, PK. **Microeconometrics: methods and applications**. Cambridge University Press, 2005.

Campbell, JY; Cocco, JF. **How do house prices affect consumption? Evidence from micro data**. Journal of Monetary Economics, v. 54, p.591-621, 2007.

Cardella, E; Seiler, MJ. **The effect of listing price strategy on real estate negotiations: An experimental study.** *Journal of Economic Psychology*, v. 52, p. 71-90, 2016.

Carrillo, PE. **To sell or not to sell: Measuring the heat of the housing market.** *Real Estate Economics*, v. 41, n. 2, p. 310-346, 2013.

Case, KE.; Quigley, JM.; Shiller, RJ. **Wealth effects revisited: 1975-2012.** National Bureau of Economic Research. No. w18667. 2013

Case, KE.; Shiller, RJ. **The efficiency of the market for single-family homes.** *The American Economic Review*, Vol. 79, pp. 125-137. 1988

_____. **Forecasting prices and excess returns in the housing market.** *Real Estate Economics*, Vol. 18, pp. 253-273. 1990.

Chen, J; Rutherford, RC. **Quality & Time-on-the-Market in Residential Markets.** *The Journal of Real Estate Finance and Economics*, v. 44, n. 3, p. 414-428, 2012.

Cheshire, P; Sheppard, S. **Capitalising the Value of Free Schools: The Impact of Supply Characteristics and Uncertainty.** *The Economic Journal*, 114, pp. 397–424. 2004

Chong, WK., Phipps, TT., Anselin, L. **Measuring the benefits of air quality improvement: a spatial hedonic approach,** *Journal of Environmental Economics and Management*, Vol 45, Issue 1, pp. 24-39, 2003.

Clapp, J; Nanda, A; Ross, S. **Which School Attributes Matter? The Influence of School District Performance and Demographic Composition on Property Values.** *Journal of Urban Economics*, Vol. 63, pp. 451–466. 2008

Cramer, J. S. **Econometric Applications of Maximum Likelihood Methods,** Cambridge University Press, Cambridge. 1986

Diebold, F. X., Mariano, R. S. **Comparing predictive accuracy.** *Journal of Business & Economic Statistics*. 2005.

Dubin, R A; Sung, C H. **Specification of hedonic regressions: Non-nested tests on measures of neighborhood quality,** *Journal of Urban Economics*, Vol 27, Issue 1, pp. 97-110, 1990.

Elhorst, JP. **Dynamic Models in Space and Time.** Research Report 00C16, University of Groningen, Research Institute SOM (Systems, Organizations and Management). 2000

_____. **Spatial panel data models.** In: Spatial Econometrics. Springer Berlin Heidelberg, p. 37-93. 2014

Epple, D. **Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products.** Journal of Political Economy, Vol. 95 (1), pp. 59-80, 1987.

Exame. **Por que o brasileiro adora investir na poupança e em imóveis?** Disponível em <<http://exame.abril.com.br/videos/seu-dinheiro-na-tv/por-que-o-brasileiro-adora-investir-na-poupanca-e-em-imoveis/>> (acesso em 26/02/2016)

Fávero, L. P. L.; Belfiore, P. P., Lima, G. A. **Modelos de precificação hedônica de imóveis residenciais na região metropolitana de São Paulo: uma abordagem sob as perspectivas da demanda e da oferta.** Estudos Econômicos (São Paulo), Vol. 38, pp. 73-96. 2008

Ferreira Neto, E. **Estimação do preço hedônico: uma aplicação para o mercado imobiliário da cidade do Rio de Janeiro.** Dissertação de Mestrado da Escola de Pós-Graduação em Economia da Fundação Getúlio Vargas, Rio de Janeiro, 2002

Fujita, M.; Ogawa, H. **Multiple equilibria and structural transition of nonmonocentric urban configuration.** Regional Science and Urban Economics, Vol. 12, pp. 161-196, 1982.

Furtado, B. **Mercado imobiliário e a importância das características locais: uma análise quantílico-espacial de preços hedônicos em Belo Horizonte.** Análise Econômica Vol.48. pp.71-98. 2007.

Genre, V., Kenny, G., Meyler, A., Timmermann, A.. **Combining expert forecasts: Can anything beat the simple average?** International Journal of Forecasting, 29(1), 108-121. 2013.

Gibbons, S; Machin, S. **Valuing School Quality, Better Transport and Lower Crime: Evidence from House Prices.** Oxford Review of Economic Policy, Vol.24, pp. 99–119. 2008

Glaeser, E.L.; Gyourko, J.; Saiz, A. **Housing supply and housing bubbles,** Journal of Urban Economics, Vol 64, pp. 198-217, 2008.

González, M. A. S.; Formoso, C. T. **Estimativa de Modelos de Preços Hedônicos para Localização Residencial em Porto Alegre.** Revista Produção. Vol 5, pp. 65-77. 1995

_____. **Análise conceitual das dificuldades na determinação de modelos de formação de preços através de análise de regressão.** Revista Engenharia Civil – UM (Universidade do Minho), Portugal, n. 8, pp. 65-75, 2000

Han, L. **The Microstructure of Housing Markets: Search, Bargaining, and Brokerage** Prepared for Handbook of Regional And Urban Economics, Volume 5 Duranton, Henderson, and Strange (eds.). 2014.

Hansen, P.R. **A test for superior predictive ability.** Journal of Business & Economic Statistics, 2005

_____; Lunde, A.; Nason, J. M. **The Model Confidence Set.** Econometrica, 79(2), 453-497. 2011

Hermann, B. M.; Haddad, E. A. **Mercado Imobiliário e Amenidades Urbanas: A View Through the Window.** Estudos Econômicos, Vol. 35. 2005.

Himmelberg, C; Mayer, C; Sinai, T. **Assessing High House Prices: Bubbles, Fundamentals and Misperceptions.** Journal of Economic Perspectives, Vol. 19, pp. 67-92. 2005

Kelejian, HH.; Prucha, IR. **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, Vol. 17, pp. 99-121. 1998

Knight, JR. **Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes.** Real Estate Economics, v. 30, n. 2, p. 213-237, 2002.

Koop, G.; Korobilis, D. **Forecasting inflation using dynamic model averaging.** International Economic Review, 53(3), 867-886. 2012

Krainer, J. **A Theory of Liquidity in Residential Real Estate Markets.** Journal of Urban Economics, v.49, p. 32-53, 2001

Lancaster, KJ. **A New Approach to Consumer Theory.** Journal of Political Economy, Vol. 74 (2), pp. 132-157, 1966.

LeSage, J. **Theory and Practice of Spatial Econometrics.** University of Toledo, 1999. Disponível em: < <http://www.spatial-econometrics.com/html/sbook.pdf>> (acesso em 26/01/2015)

Liu, X. **Spatial and temporal dependence in house price prediction**. The Journal of Real Estate Finance and Economics, Vol. 47, pp. 341-369. 2012

Macedo, P. B. R. **Hedonic Price Models with Spatial Effects: An Application to the Housing Market of Belo Horizonte, Brazil**. Revista Brasileira de Economia, Vol 51(1), pp. 63-82. 1998

Mankiw, G. N ; Weil, D.N. **The baby boom, the baby bust, and the housing market**, Regional Science and Urban Economics, Vol. 19, pp. 235-258, 1989.

Merlo,A; Ortalo-Magné, F; Rust, J. **The home selling problem: Theory and evidence**. International Economic Review, v. 56, n. 2, p. 457-484, 2015.

Miles, W. **Boom–Bust Cycles and the Forecasting Performance of Linear and Non-Linear Models of House Prices**, The Journal of Real Estate Finance and Economics, Vol. 36, pp. 249-264. 2008

Miller, NG. **Time on the market and selling price**. Real Estate Economics, v. 6, n. 2, p. 164-174, 1978.

Muellbauer, J.; Murphy, A. **Booms and busts in the UK housing market**. The Economic Journal, Vol. 107, pp.1701-1727.1997.

Nelson, J P. **Residential choice, hedonic prices, and the demand for urban air quality**, Journal of Urban Economics, Vol 5, Issue 3, pp. 357-369, 1978.

Pace, RK.; Barry, R.; Clapp, JM.; Rodriquez, M.; **Spatiotemporal autoregressive models of neighborhood effects**. The Journal of Real Estate Finance and Economics, Vol. 17, pp. 15-33. 1998

_____; _____; Gilley, OW.; Sirmans, CF. **A method for spatial–temporal forecasting with an application to real estate prices**. International Journal of Forecasting, Vol. 16(2), pp. 229-246. 2000

Pinto, M.; Zampier, R. L. **Os Significados do Consumo de Imóveis de Luxo na Perspectiva da Consumer Culture Theory (CCT)**. Revista ADM. MADE, 17(3), 1-26. 2013

Plaigin, C. **Exploratory study on the presence of cultural and institutional growth spillovers**. III World Conference of Spatial Econometrics, Barcelona, 2009.

Pope, J C. **Buyer information and the hedonic: The impact of a seller disclosure on the implicit price for airport noise**, Journal of Urban Economics, Vol 63, Issue 2, March, pp. 498-516, 2008

Politis, DN.; Romano, JP. **The stationary bootstrap**. Journal of the American Statistical Association, v. 89, n. 428, p. 1303-1313, 1994.

Portugal, M. S. **Notas Introdutórias Sobre o Princípio de Máxima Verossimilhança: Estimação e Teste de Hipóteses**. UFRGS. 1995. Disponível em <<http://www.ufrgs.br/decon/hp/publionline/textosdidaticos/Textodid04.pdf>> (acesso em 27/01/2015)

Rondon, V. V; Andrade, M. V **Impactos da criminalidade no valor dos aluguéis em Belo Horizonte**. Textos para discussão. Belo Horizonte: CEDEPLAR/UFMG 2003.

Rosen, S. **Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition**. Journal of Political Economy, pp. 34–55, 1974.

Rozenbaum, S., Macedo-Soares, T. D. **Proposta para construção de um índice local de preços de imóveis a partir dos lançamentos imobiliários de condomínios residenciais**. Revista de Administração Pública, 41(6), pp. 1069-1094. 2007

Sant'ana, S L J. **Determinantes do Preço de Imóveis Residenciais na Cidade de São Paulo**. Dissertação de Mestrado da Escola de Administração de Empresas da Fundação Getúlio Vargas, São Paulo. 2006

Smith, PS.; Gilbler, KM.; Zahirovic-Herbert, V. **The Effect of Relisting on House Selling Price**. The Journal of Real Estate Finance and Economics, p. 1-20, 2015.

Taylor, CR. **Time-on-the-Market as a Sign of Quality**. The Review of Economic Studies, v. 66, n. 3, p. 555-578, 1999.

Stock, J. H.; Watson, M. W. **Forecasting inflation**. Journal of Monetary Economics, 44(2), 293-335. 1999

Tyszler, M. **Econometria espacial: discutindo medidas para a matriz de ponderação espacial**. Dissertação de Mestrado da Escola de Administração de Empresas da Fundação Getúlio Vargas, São Paulo. 2006

Witte, A D.; Sumka, H J.; Erekson, H. **An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets.** *Econometrica*, Vol 47, pp. 1151-1573, 1979.

White, H. **A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity.** *Econometrica: Journal of the Econometric Society*, pp. 817-838. 1980

Yavas, Abdullah; Yang, Shiawee. **The strategic role of listing price in marketing real estate: theory and evidence.** *Real Estate Economics*, v. 23, n. 3, p. 347-368, 1995.

Zap Imóveis. **Índice Fipe ZAP.** Disponível em: <<http://www.zap.com.br/imoveis/fipe-zap-b/>>. Acesso em 24 de Março de 2015.

APÊNDICE

APÊNDICE 1 – Datas de coleta dos dados – Base completa

Venda			Aluguel		
24/10/2012	08/12/2012	11/04/2013	10/11/2012	10/01/2013	29/04/2013
26/10/2012	10/12/2012	15/04/2013	11/11/2012	14/01/2013	09/05/2013
27/10/2012	11/12/2012	17/04/2013	12/11/2012	15/01/2013	14/05/2013
28/10/2012	12/12/2012	18/04/2013	13/11/2012	16/01/2013	16/05/2013
29/10/2012	13/12/2012	24/04/2013	14/11/2012	23/01/2013	17/05/2013
30/10/2012	14/12/2012	26/04/2013	15/11/2012	24/01/2013	20/05/2013
31/10/2012	17/12/2012	29/04/2013	18/11/2012	29/01/2013	29/05/2013
01/11/2012	18/12/2012	09/05/2013	19/11/2012	07/02/2013	04/06/2013
04/11/2012	20/12/2012	14/05/2013	20/11/2012	14/02/2013	10/06/2013
05/11/2012	25/12/2012	16/05/2013	21/11/2012	15/02/2013	13/06/2013
06/11/2012	08/01/2013	17/05/2013	22/11/2012	18/02/2013	17/06/2013
07/11/2012	10/01/2013	20/05/2013	23/11/2012	20/02/2013	20/06/2013
08/11/2012	14/01/2013	27/05/2013	24/11/2012	21/02/2013	03/07/2013
11/11/2012	15/01/2013	29/05/2013	25/11/2012	25/02/2013	12/07/2013
13/11/2012	16/01/2013	04/06/2013	26/11/2012	28/02/2013	15/07/2013
14/11/2012	23/01/2013	10/06/2013	27/11/2012	01/03/2013	25/07/2013
15/11/2012	24/01/2013	13/06/2013	28/11/2012	04/03/2013	29/07/2013
18/11/2012	29/01/2013	17/06/2013	29/11/2012	05/03/2013	12/08/2013
19/11/2012	07/02/2013	20/06/2013	30/11/2012	07/03/2013	15/08/2013
20/11/2012	14/02/2013	03/07/2013	02/12/2012	11/03/2013	19/08/2013
21/11/2012	15/02/2013	12/07/2013	03/12/2012	14/03/2013	27/08/2013
22/11/2012	18/02/2013	15/07/2013	04/12/2012	16/03/2013	02/09/2013
23/11/2012	20/02/2013	23/07/2013	05/12/2012	18/03/2013	19/10/2013
24/11/2012	21/02/2013	25/07/2013	06/12/2012	21/03/2013	25/10/2013
25/11/2012	25/02/2013	29/07/2013	07/12/2012	25/03/2013	29/10/2013
26/11/2012	28/02/2013	12/08/2013	08/12/2012	26/03/2013	11/11/2013
27/11/2012	01/03/2013	15/08/2013	10/12/2012	28/03/2013	22/11/2013
28/11/2012	04/03/2013	19/08/2013	11/12/2012	01/04/2013	
29/11/2012	05/03/2013	27/08/2013	12/12/2012	03/04/2013	
30/11/2012	07/03/2013	02/09/2013	13/12/2012	05/04/2013	
02/12/2012	08/03/2013	19/10/2013	14/12/2012	09/04/2013	
03/12/2012	11/03/2013	29/10/2013	17/12/2012	11/04/2013	
04/12/2012	01/04/2013	11/11/2013	18/12/2012	15/04/2013	
05/12/2012	03/04/2013	22/11/2013	20/12/2012	18/04/2013	
06/12/2012	05/04/2013		25/12/2012	24/04/2013	
07/12/2012	09/04/2013		08/01/2013	26/04/2013	

APÊNDICE 2 – Datas de coleta dos dados – Base restrita

Venda	Aluguel
24/10/2012	11/11/2012
11/11/2012	23/11/2012
23/11/2012	10/12/2012
10/12/2012	25/12/2012
25/12/2012	14/01/2013
14/01/2013	29/01/2013
29/01/2013	14/02/2013
14/02/2013	28/02/2013
28/02/2013	11/03/2013
11/03/2013	01/04/2013
01/04/2013	24/04/2013
17/04/2013	09/05/2013
24/04/2013	20/05/2013
09/05/2013	10/06/2013
20/05/2013	20/06/2013
10/06/2013	12/07/2013
20/06/2013	25/07/2013
12/07/2013	12/08/2013
25/07/2013	27/08/2013
12/08/2013	02/09/2013
27/08/2013	19/10/2013
02/09/2013	29/10/2013
19/10/2013	11/11/2013
29/10/2013	22/11/2013
11/11/2013	
22/11/2013	