Albuquerque, Pedro H. M.; Furtado, Bernardo Alves

Working Paper

Índices de preço e o mercado imobiliário: Literatura e ilustração - o caso de São Paulo

Texto para Discussão, No. 2200

Provided in Cooperation with:
Institute of Applied Economic Research (IPEA), Brasília

2200

TEXTO PARA DISCUSSÃO

ÍNDICES DE PREÇO E O MERCADO IMOBILIÁRIO: LITERATURA E ILUSTRAÇÃO – O CASO DE SÃO PAULO

Pedro H. M. Albuquerque
Bernardo Alves Furtado

ipea

2200
TEXTO PARA DISCUSSÃO

Brasília, maio de 2016

ÍNDICES DE PREÇO E O MERCADO IMOBILIÁRIO: LITERATURA E ILUSTRAÇÃO – O CASO DE SÃO PAULO[1]

Pedro H. M. Albuquerque[2]
Bernardo Alves Furtado[3]

1. Os autores agradecem pelas contribuições de Vanessa Nadalin e Vicente Lima Neto, do Ipea, e pelos comentários de Lucas Mation e Rodrigo Orair.

2. Professor da Universidade de Brasília (UnB) e bolsista do Subprograma de Pesquisa para o Desenvolvimento Nacional (PNPD) do Ipea.
3. Coordenador e técnico de planejamento e pesquisa do Ipea e bolsista de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Texto para Discussão

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

Texto para discussão / Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1.Brasil. 2.Aspectos Econômicos. 3.Aspectos Sociais. I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

JEL: C43, R32, E31

SUMÁRIO

SINOPSE

Este texto faz uma revisão teórica de índices de preço ao consumidor, com ênfase nas especificidades do mercado imobiliário, apontando vantagens e limitações de cada um. Discute, assim, os índices de médias de preço, quais sejam: Laspeyres, Paasche, Fisher e Jevons. Adicionalmente, apresenta as duas variações de construção por métodos hedônicos: com inserção de variáveis *dummies* e por imputação. As especificidades típicas do mercado imobiliário, os atributos do imóvel, sua longa temporalidade e inserção urbana sugerem maiores cuidados para a construção de índices para o mercado imobiliário. O texto apresenta alguns avanços teóricos recentes para o mercado imobiliário e conclui com uma aplicação possível a partir de dados disponíveis para São Paulo.

Palavras-chave: índices de preço; mercado imobiliário; São Paulo.

ABSTRACT

This paper is a theoretical review of consumer price indices, with an emphasis on the specifics of the real estate market, highlighting advantages and limitations. It discusses the following indices of average prices: Laspeyres, Paasche, Fisher and Jevons. In addition, it presents two variations of construction by hedonic methods: with insertion of dummy variables and imputation. Typical specificities of the real estate market, such as property attributes, its long temporality and urban insertion, recommend greater care when building indices for the housing market. The paper presents some recent theoretical advances for the real estate market and concludes with a brief application for Sao Paulo.

Keywords: prices indices; real estate market; Sao Paulo.

1 INTRODUÇÃO

Índices de preço ao consumidor medem a evolução do nível de preços de cestas de bens e serviços adquiridos pelas famílias no mercado. De forma mais genérica, índices são estimativas estatísticas baseadas em preços amostrais de itens representativos coletados periodicamente. Subíndices e subsubíndices são calculados para diferentes categorias e subcategorias de bens e serviços, sendo combinados com o intuito de produzir índice geral composto (Aizcorbe, 2014). Os pesos do índice refletem seus impactos no total das despesas de consumo em questão.

De maneira geral, são necessários dois tipos básicos de dados para construir um índice de preços: dados dos preços e informações acerca das regras de ponderação de cada produto na composição da cesta. Os dados sobre os preços são coletados para uma amostra de produtos e serviços a partir de uma amostra de pontos de oferta em locais selecionados em tempos especificados. Os dados utilizados na ponderação do impacto de diferentes tipos de despesa no total da despesa coberta pelo índice são obtidos por meio de um processo de estimação estatística.

A periodicidade de um índice de preços também pode variar, e geralmente é calculado mensalmente ou trimestralmente em alguns países, como a média ponderada dos subíndices para diferentes componentes da despesa do consumidor, tais como alimentação, habitação e vestuário; cada um dos quais, por sua vez, é uma média ponderada dos subsubíndices.

A partir da construção do índice, é possível comparar os preços de cada mês com os preços no mês de referência. Nesse sentido, as ponderações utilizadas para combiná-los em agregados de nível superior e, em seguida, a formulação do índice geral referem-se aos gastos estimados durante todo um ano passado para os consumidores abrangidos pelo índice sobre os produtos dentro de seu escopo na região investigada. Assim, o índice é considerado um índice de peso fixo, mas raramente pode ser considerado um verdadeiro índice de Laspeyres (veja a definição na seção seguinte), uma vez que o período do peso de referência de um ano e o período preço-referência, normalmente um mês mais recente, não coincidem, pois é preciso tempo para reunir e processar as informações utilizadas para ponderação, que, além de pesquisas de despesas familiares, podem incluir dados comerciais e fiscais.

Especificamente, segundo Nadalin e Furtado (2011), a construção de índices de preço para o mercado imobiliário apresenta algumas particularidades, haja vista que, em especial, a habitação é heterogênea no que se refere a seus atributos e a sua localização. Em outras palavras, os preços refletem tanto os custos de construção das estruturas físicas quanto a valorização das amenidades urbanas, sendo, portanto, dependente de atributos estruturais e espaciais.

Ainda segundo Nadalin e Furtado (2011), as habitações, por serem muito distintas umas das outras em suas características, apresentam um nível de qualidade variável no tempo e impõem dificuldades à construção de índices de preço. Objetivamente, o cerne da dificuldade está em acompanhar no tempo as mudanças no preço de um mesmo produto idêntico. Ademais, o bem habitação é indivisível, heterogêneo, durável e negociado em um mercado fino, características essas que, associadas com a baixa frequência de negociação do bem, dificultam a operacionalização da construção de índices de preço por meio das abordagens tradicionais. O principal desafio, então, é controlar as diferenças e alterações da qualidade das unidades efetivamente transacionadas em cada período de aferição do índice.

Este texto se insere no projeto mais amplo de "índices de preço de imóveis", que tem como objetivo a discussão acerca das metodologias de construção de índices de preço para o Brasil. As metodologias possíveis estão restritas à existência de base de dados, sua disponibilidade, regularidade e confiabilidade. Como resultados preliminares, a equipe produziu dois textos iniciais: este – que versa sobre a metodologia de índices centrais e índices hedônicos, em especial – e índice de preços de vendas repetidas (em finalização). O *tradeoff* ponderado pela equipe do projeto é tripartite: *i)* o acesso e a possibilidade da base de dados; *ii)* a simplicidade e a efetividade do índice; e *iii)* a adequação entre metodologia, *i)* e *ii)*.

Em relação à base de dados, ambas as propostas (a deste texto de índice de preços hedônicos e a do outro texto, de índice de vendas repetidas) podem vir a ser testadas caso se obtenha acesso à base de dados descrita na Portaria nº 4.088 do Banco Central do Brasil (BCB), regulada pela Circular nº 3.747, que prevê série completa de variáveis sobre transações imobiliárias que será devida pelos bancos ao BC a partir de fevereiro de 2016.

Nesse contexto, este texto tem por objetivo apresentar a revisão teórica na construção de índices de preço ao consumidor na sua forma tradicional (seção 2), por meio de regressão hedônica (seção 3), e uma revisão das propostas para índices de preço no mercado imobiliário (seção 4). Adicionalmente, apresenta uma aplicação da construção do índice de preços de imóveis para o caso de São Paulo (seção 5). As considerações finais são apresentadas na seção 6.[1]

2 ÍNDICES DE PREÇO

Um índice de preços fornece uma medida agregada da variação do preço para um particular produto ou segmento, indústria ou até mesmo para toda a economia. A maioria das fórmulas para a construção de números-índice se baseia no monitoramento das variações dos preços de uma particular cesta de produtos ao longo do tempo. Essas técnicas são também aplicadas para mensurar a diferença de preços entre regiões geográficas, no tempo ou até mesmo na dimensão tempo e espaço. Numericamente, as propostas matemáticas mais utilizadas para a construção de índices baseiam-se no cômputo de uma média ponderada da variação dos preços para determinados itens.

Inicialmente, esta seção tem por objetivo apresentar as propostas típicas na construção de índices de preço, quais sejam: índice de Laspeyres, índice de Paasche, índice ideal de Fisher, índice de Törnqvist e índice de Jevons. A formulação teórica e a justificativa ampla para a utilização de cada um desses índices estão bem relatadas no texto de Diewert (2008).

Uma das primeiras propostas para a formulação de índices de preço é o índice de Laspeyres, o qual é frequentemente escrito como:

$$I_{0,1}^{L} = \frac{\sum_{m=1}^{M} (P_{m,1}\, Q_{m,0})}{\sum_{m=1}^{M} (P_{m,0}\, Q_{m,0})}, \tag{1}$$

em que P e Q denotam os preços e as quantidades de bens e os valores 0 e 1 referem-se a dois momentos no tempo, cujo tempo-base é definido como sendo o período 0 e o período atual (contemporâneo) é definido pelo período 1. Nessa proposta, supõe-se que M bens foram vendidos em ambos os períodos considerados, bens cuja indexação é dada pelo índice m.

1. Este texto se insere no projeto Índices de preços de imóveis, que prevê ainda texto sobre índices de vendas repetidas e aplicação ao caso do Distrito Federal, dentre outros produtos, visando ao melhor entendimento sobre as possibilidades teóricas e empíricas de construção de índices de preço de imóveis no Brasil, até o momento inexistentes ou muito restritos.

A noção por trás da proposta de Laspeyres é comparar o custo de aquisição dos bens adquiridos no período-base (isto é, $t = 0$), dado pelo numerador da equação (1), em relação ao custo hipotético de compra dos mesmos bens no período 1, representado numericamente pelo denominador da equação (1). Note que o índice de Laspeyres pode ser escrito como uma média aritmética ponderada da variação dos M preços dos produtos considerados na forma:

$$I_{0,1}^{L} = \sum_{m=1}^{M} w_{m,0} \frac{P_{m,1}}{P_{m,0}}, \tag{2}$$

em que os pesos $w_{m,0} = \frac{P_{m,0}\, Q_{m,0}}{\sum_{m=1}^{M} (P_{m,0}\, Q_{m,0})}$ representam a razão entre a despesa com o m-ésimo bem em relação à despesa total no período de referência, isto é, o período-base $t=0$; esses pesos também são denominados de *importância relativa*, já a razão $\frac{Pm,1}{Pm,0}$ mensura a variação dos preços para cada item individual.

Cabe ressaltar aqui que, na proposta de Laspeyres, os bens considerados no período de referência também devem ter sido observados no período contemporâneo, pois somente assim a variação nos preços pode ser computada; ademais, o índice de Laspeyres mantém fixa a importância relativa dos itens, o que pode ser inconveniente caso a estrutura de consumo mude frequentemente entre as unidades consumidoras. Apesar dessa desvantagem, o índice de Laspeyres é fácil de ser construído, pois necessita somente das quantias e preços exercidos nos períodos de interesse, o que facilita o trabalho do analista na construção de uma proposta, mesmo que ingênua, e na estimação do índice de preços de consumo.

O índice de Paasche é similar ao índice de Laspeyres no sentido de também manter fixa a cesta de produtos considerada no cômputo do índice, mas usa uma cesta de mercado diferente para mensurar a variação dos preços, pois o índice compara o custo atual de aquisição no período contemporâneo ($t=1$), dado por $\sum_{m=1}^{M} (P_{m,1}\, Q_{m,1})$, em relação ao custo total das mesmas quantidades adquiridas no tempo contemporâneo ao preço no tempo inicial, isto é $\sum_{m=1}^{M} (P_{m,0}\, Q_{m,1})$. Matematicamente, o índice de Paasche é dado por:

$$I_{0,1}^{P} = \frac{\sum_{m=1}^{M} (P_{m,1}\, Q_{m,1})}{\sum_{m=1}^{M} (P_{m,0}\, Q_{m,1})}. \tag{3}$$

Assim como o índice de Laspeyres, o índice de Paasche também pode ser escrito por meio de uma média aritmética ponderada. Particularmente, o inverso do índice de

Paasche representa a média ponderada do inverso dos preços relativos; assim, o índice de Paasche é uma média harmônica ponderada nesta forma:

$$I_{0,1}^{P} = \frac{1}{\sum_{m=1}^{M} w_{m,1}\left(\frac{P_{m,0}}{P_{m,1}}\right)}, \tag{4}$$

em que os pesos $w_{m,1} = \frac{P_{m,1}\,Q_{m,1}}{\sum_{m=1}^{M} P_{m,1}\,Q_{m,1}}$ representam a razão do gasto com o *m*-ésimo bem no período contemporâneo em relação à despesa total com todos os bens no período corrente, isto é, t=1. Como destacado anteriormente, a variação dos preços, representada por $\frac{P_{m,0}}{P_{m,1}}$, mensura a variação dos preços entre dois momentos distintos para o mesmo bem considerado, qual seja, o *m*-ésimo bem.

Outra similaridade compartilhada entre os índices de Laspeyres e Paasche é o fato de ambos considerarem a mesma cesta de produtos em ambos os períodos e a exigência do consumo dos mesmos bens no período de referência e o período de interesse, usualmente definido como o período contemporâneo; já a diferença existente entre essas duas propostas está no fato de a ponderação (isto é, importância relativa do bem) ser realizada de maneira distinta.

O índice ideal de Fisher é a união das propostas de Laspeyres e Paasche por meio do cômputo de uma média geométrica entre esses dois índices:

$$I_{0,1}^{F} = \sqrt{I_{0,1}^{L}\, I_{0,1}^{P}} = \left\{\left[\frac{\sum_{m=1}^{M}(P_{m,1}\,Q_{m,0})}{\sum_{m=1}^{M}(P_{m,0}\,Q_{m,0})}\right]\left[\frac{\sum_{m=1}^{M}(P_{m,1}\,Q_{m,1})}{\sum_{m=1}^{M}(P_{m,0}\,Q_{m,1})}\right]\right\}^{1/2}. \tag{5}$$

Por meio de operações algébricas, pode-se mostrar que o índice ideal de Fisher é escrito como função de média ponderada, isto é:

$$I_{0,1}^{F} = \left\{\frac{\sum_{m=1}^{M} w_{m,0}\,\frac{P_{m,1}}{P_{m,0}}}{\sum_{m=1}^{M} w_{m,1}\,\frac{P_{m,0}}{P_{m,1}}}\right\}^{1/2}, \tag{6}$$

com o índice de Laspeyres no numerador e o inverso do índice de Paasche no denominador. Cabe ainda notar que, diferentemente do índice de Laspeyres, o índice ótimo de Fisher utiliza as parcelas das despesas de ambos os períodos por meio dos pesos $w_{m,0}$ e $w_{m,1}$, possibilitando assim uma mudança, ao longo do tempo, das contribuições dos bens no cômputo do índice. Segundo Aizcorbe (2014), uma razão intuitiva para a superioridade do índice de Fisher está no fato de este refletir a composição dos bens atualmente adquiridos. O índice de Laspeyres, por sua vez, precificará, para a variação do preço entre 2000 a 2010, os bens adquiridos em 2000 e ignorará qualquer alteração na composição da cesta efetiva dos consumidores em 2010 ao longo desses dez anos. O índice de Paasche, no entanto, possui a mesma estrutura fixa, supondo a composição de 2010 da distribuição dos bens na cesta de consumo em relação ao período-base exemplificado, qual seja, o ano 2000.

O índice de Fisher apresenta ainda outras propriedades úteis. Diewert (1997) demonstrou que, de todas as possíveis médias tomadas entre os índices de Laspeyres e Paasche, o índice ótimo de Fisher é o único que apresenta duas propriedades desejáveis, quais sejam: homogeneidade e simetria. Além de satisfazer o teste de reversão do tempo, o qual determina que a variação dos preços em relação ao período-base e contemporâneo deveria ser o inverso da variação do preço entre o período atual e o período de referência (contemporâneo e base respectivamente), isso significa que, não importa qual período seja escolhido como período-base, os resultados devem se manter equivalentes.

O último índice revisado nesta seção é o índice de Jevons, dado por:

$$\log I_{0,1}^{J} = \frac{1}{M} \sum_{m=1}^{M} \log \frac{P_{m,1}}{P_{m,0}} . \qquad (7)$$

O índice de Jevons utiliza o logaritmo natural da variação dos preços na construção dos índices, por meio de uma média aritmética simples, substituindo os pesos $\frac{(w_{m,0} + w_{m,1})}{2}$ por $1/M$, gerando assim uma estrutura com pesos constantes. Observe que esse índice é adequado quando as quantias consumidas nos períodos não são conhecidas, o que facilita o cômputo pelos analistas na ausência de dados suficientes.

3 MÉTODOS HEDÔNICOS

Regressões hedônicas foram incialmente desenvolvidas por Court (1939) e posteriormente refinadas por Griliches (1961). Especificamente, uma regressão hedônica relaciona a variação nos preços entre bens e períodos simultaneamente para diferenciar adequadamente as características de cada bem considerado. Ressalte-se, por exemplo, que apartamentos menores apresentam valores para o metro quadrado maiores que apartamentos maiores e que a idade do imóvel afeta de modo diferente casas e apartamentos. Essas diferenças derivadas das características dos bens e que se refletem nos preços devem ser consideradas.

As regressões hedônicas relacionam as características observadas dos bens que compõem a cesta de produtos e relacionam essas características com a variação observada no preço ao longo do tempo com o objetivo de controlar a variação existente entre as características dos bens observados.

De maneira prática, as regressões hedônicas são implementadas de duas formas: por meio da inserção de uma variável *dummy* para considerar os períodos observados ou por meio do método de imputação.

3.1 Inserção de variáveis *dummies*

O índice de preços hedônicos tradicional é também denominado de índice de preços por meio de variáveis *dummies* e é uma abordagem que utiliza o método de regressão para explicar os preços de bens individuais como função das características desses mesmos bens. Os dados para os bens são unidos, e o índice de preços é estimado utilizando-se uma variável *dummy* para cada período considerado.

O índice de preços por meio de variáveis *dummies* é então calculado por meio dos coeficientes estimados das variáveis *dummies* inseridas no modelo de regressão. Pakes (2003) afirma que uma das vantagens dessa abordagem é que o efeito das características dos bens se mantém constante ao longo do tempo, dado o efeito fixo gerado pela inserção das variáveis dicotômicas para cada período considerado. Os outros parâmetros do modelo são permitidos variar.

Considere um conjunto de dados na forma de painel (*panel data*) para T períodos e M produtos, os quais foram vendidos ao preço $P_{m,t}$ no tempo t=1,...T para o produto m, o qual pode variar dentre os M produtos que compõem a cesta, isto é, m=1,..., M. Já as características determinantes do preço de cada bem são representadas por $X_{k,m,t}$, no qual há k=1,…,K características disponíveis. O método ainda considera a inserção das variáveis *dummies* para cada período temporal ($D_{m,t}$, para t=1,…,T). Tradicionalmente, com o objetivo de aproximar a distribuição dos preços observados a uma distribuição normal, utiliza-se como variável dependente o logaritmo natural do preço desta forma:

$$\log P_{m,t} = \alpha + \sum_{k=1}^{K} \beta_k X_{k,m,t} + \sum_{t=1}^{T} \delta_t D_{m,t} + \epsilon_{m,t}, \tag{8}$$

em que α , β_k e δ_t são os parâmetros do modelo a serem estimados e $\epsilon_{m,t}$ é o termo de ruído aleatório usual nos modelos de regressão lineares multivariados. Por meio do modelo apresentado em (8), cada produto possui K possíveis características que podem influenciar o valor do bem. O termo $\sum_{k=1}^{K} \beta_k X_{k,m,t}$ é responsável por controlar a diferença entre os atributos dos produtos e todas as demais influências associadas à variação do preço são computadas pelos termos de efeitos fixos, representados pelos parâmetros δ_t, isto é, qualquer variação no preço de um determinado bem (consideradas as suas características fixas) é devido ao tempo, o qual é modelado por meio das variáveis binárias.

Os coeficientes δ_t capturam, portanto, o valor médio das influências de cada período de tempo e formam a base do índice de preços por meio de variáveis *dummies*. Cabe ressaltar ainda que, como em qualquer outro modelo de regressão, é necessário considerar as possíveis patologias existentes na estimação dos parâmetros, quais sejam: heterocedasticidade, multicolinearidade, correlação serial e desvios na normalidade. Outro ponto importante no processo de estimação dos parâmetros do modelo descrito pela equação (8) é a possibilidade da existência de variáveis omitidas, principalmente pela dificuldade prática de listar *todas* as características que afetam o preço dos produtos considerados. Nesse sentido, Bajari e Benkard (2005) e Erickson e Pakes (2011) propõem abordagens próprias para a estimação de regressões hedônicas quando variáveis omitidas poderiam existir. Dentre outras soluções, destaca-se a proposta de Triplett (2004), que aconselha a utilização de preços de transações efetivamente realizadas em vez de uma lista simples de preços.

Nesse sentido, no modelo de regressão com variáveis binárias (*dummies*) o intercepto representado pelo parâmetro α fornece o preço médio para algum grupo de referência. Por exemplo, se todas as características consideradas no modelo forem de mensuração contínua e as únicas variáveis *dummies* presentes estiverem associadas com a dimensão temporal, então, ao se escolher um determinado ano como referência (por exemplo, 2010) e os demais anos forem relacionados com as variáveis *dummies* (por exemplo, 2011), o intercepto representará o preço médio para o ano 2010 e a estimativa $\hat{\delta}_{2011}$ representará o preço médio em 2011 relativo a 2010, tal que $\hat{\delta}_{2011} = \log I_{2010,2011}^{DV}$. Cabe ressaltar que, por se tratar de uma transformação monótona, a estimativa para $\hat{\delta}_{2011}$ pode conter algum viés, segundo desigualdade de Jensen, entretanto, Gujarati (2012) sugere que as estimativas são, em geral, muito próximas.

De maneira mais geral, os dados podem envolver outros períodos de tempo com variáveis *dummies* para todos os períodos de interesse, exceto para o período de referência, seguindo a mesma interpretação apresentada anteriormente. Ademais, diferenças entre os coeficientes das variáveis *dummies* mensuram a variação no preço entre dois períodos – por exemplo, se o modelo de regressão possuir dados para os anos 2010, 2011 e 2012 e considerando 2010 como o ano de referência, a variação entre os anos 2011 e 2012 pode ser medida usando os coeficientes estimados para os respectivos anos, isto é, $\log I^{DV}_{2011,2012} = \hat{\delta}_{2012} - \hat{\delta}_{2011}$. Note que $\log I^{DV}_{2010,2012} = \log I^{DV}_{2010,2011} + \log I^{DV}_{2011,2012} = \hat{\delta}_{2011} + (\hat{\delta}_{2012} - \hat{\delta}_{2011}) = \hat{\delta}_{2012}$.

Cabe notar aqui que a conversão dessa media, em termos do logaritmo natural, para o índice de preços hedônico envolve um viés de ajuste, pois, mesmo que os coeficientes sejam estimados sem viés, ao aplicar a função exponencial não se obtém uma estimativa não enviesada. Um procedimento popular que se pressupõe normalidade dos resíduos é usar o ajuste na forma $\exp[0{,}5 \operatorname{Var}(\epsilon_{m,t})]$, ou seja, para o índice de preços $\log I^{DV}_{2010,2012} = \exp \hat{\delta}_{2012} \times \exp[0{,}5 \operatorname{Var}(\epsilon_{m,t})]$, em que ϵ representam os resíduos estimados. Woolridge (2012) fornece uma alternativa quando a normalidade não é considerada; nesse caso, $\log I^{DV}_{2010,2012} = \exp \hat{\delta}_{2012} \times \frac{1}{n} \sum \epsilon_{m,t}$.

Ao fim e ao cabo, dadas as especificidades do bem habitação, o controle das características do imóvel, por meio do processo de estimação hedônico, pode garantir a comparabilidade das *cestas* e, portanto, a construção de índices mais fidedignos.

3.2 Método de imputação

O método de imputação é uma alternativa ao método de inserção de variáveis *dummies*, o qual combina elementos das técnicas tradicionais para estimação de números-índice, e pela abordagem hedônica, particularmente, escolhe-se inicialmente alguma das propostas para a construção de índices, como o índice de Fisher, e então utiliza-se a regressão hedônica para estimar o valor predito dos preços *missings*.

Diferentemente do método das variáveis *dummies*, essa abordagem permite uma maior flexibilidade ao escolher a forma funcional do índice que se deseja utilizar em vez de restringir a estimação dos índices de preço a uma única forma. Especificamente, a regressão hedônica utilizada no método de imputação toma a mesma forma do método com as variáveis *dummies*, exceto pelo fato de que regressões separadas são realizadas para cada período de tempo, o que permite maior flexibilidade ao modelo, pois os parâmetros não estarão fixos no tempo. Considere *r* períodos de tempo. Nesse caso, a regressão hedônica é dada por:

$$\log P_m^r = \alpha^r + \sum_k \beta_k^r X_{k,m}^r + \epsilon_m = \log\left[\hat{P}_m^r(X_m^r)\right] + e_i^r, \tag{9}$$

em que o subscrito *r* denota quais períodos de tempo são utilizados para se obterem as estimativas dos coeficientes. As variáveis dependentes e independentes desse modelo são as mesmas do modelo com inserção de *dummies*, exceto pelos efeitos fixos temporais, os quais inexistem nessa proposta.

O método de imputação utiliza os coeficientes estimados de regressões na forma da equação (9) para predizer o preço nos *r* períodos considerados para algum espaço de características considerado. Esses valores preditos são representados por $\log\left[\hat{P}_m^r(X_m^s)\right]$, o qual especifica que esse é um valor predito para o logaritmo natural do preço para o modelo *m* considerado no período *s*. Usando os coeficientes da regressão estimados para o período *r*, essa abordagem fornece o quanto o vetor de características X_m^s deveria custar no período *r*.

Segundo Aizcorbe (2014), os coeficientes estimados da regressão hedônica (9) podem ser utilizados de três formas diferentes para gerar os índices de preço; nesses três casos, uma fórmula para os índices de preço precisa ser definida; em seguida, uma combinação dos preços observados e estimados são utilizados para se calcular o índice de preços de interesse.

Sem perda de generalidade, considere o índice de Laspeyres. Em termos de notação, o índice possui forma funcional equivalente à apresentada em (2). Nessa proposta, alguns itens podem ter sido computados no período *t*=0, mas não no período *t*=1; nesse caso, esses itens não são considerados no cômputo da primeira etapa por meio da fórmula (2), mas serão considerados na etapa de imputação. Pode-se expandir a expressão (2) para considerar os itens não frequentes em ambos os períodos considerados na forma:

$$I_{0,1}^L = \sum_{m=1}^M w_{m,0} \frac{P_{m,1}}{P_{m,0}} + \sum_{d=1}^D w_{d,0} \frac{P_{d,1}^u}{P_{d,0}} \tag{10}$$

O primeiro termo da expressão (10) inclui os itens que são observados em ambos os períodos *t*=0,1. Já o segundo termo mensura a variação no preço para os itens não frequentes indexados por *d*, em que essa variação é calculada por meio do preço no período observado ($P_{d,0}$) e o preço no período não observado ($P^{u}_{d,1}$). Nesse sentido, todas as três abordagens consideradas por Aizcorbe (2014) substituem os preços não observados em (10) por uma estimativa do preço gerada por meio da regressão hedônica; há ainda algumas abordagens que substituem inclusive os preços observados.

Em todos os casos, os pesos (importâncias relativas dos bens) são calculados por meio dos preços observados e incluem todos os bens negociados no período-base. Considere o preço estimado para o período *r* e modelo *m* negociado no período *s*, dado por:

$$\hat{P}^{r}_{m} = \exp\left[\log\left(\hat{P}^{r}_{m}(X^{s}_{m})\right)\right][\exp\left(0{,}5 \text{ Var}(\epsilon^{s}_{m})\right)], \tag{11}$$

em que o subscrito *r* denota o período associado à regressão hedônica e *s* denota o período de tempo correspondente aos dados. O primeiro termo é uma estimativa baseada na regressão hedônica (9) e o segundo termo é um fator de ajuste.

No caso mais simples do método de imputação, qualquer dado relativo a preços observados é utilizado na construção do índice e a regressão hedônica é utilizada somente para se imputar os preços não observados nos dados. Em termos gerais, considerando a expressão (10), o índice de Laspeyres por meio da imputação simples (IS) é dado por:

$$I^{L,IS}_{0,1} = \sum\nolimits_{m=1}^{M} w_{m,0} \frac{p_{m,1}}{p_{m,0}} + \sum\nolimits_{d=1}^{D} w_{d,0} \frac{\hat{p}^{1}_{d}(x^{0}_{d})}{p_{d,0}}, \tag{12}$$

em que a imputação ocorre no termo $\hat{P}^{u}_{d,1} = \hat{P}^{1}_{d}(X^{0}_{d})$. Essa proposta foi incialmente desenvolvida por Pakes (2003), Triplett (2004) e Silver e Heravi (2007), os quais afirmam que, por meio dessa abordagem, a variabilidade nos preços é introduzida, fornecendo assim melhores índices de preço na presença de dados faltantes (*missing data*).

No caso de variáveis omitidas, Diewert, Heravi e Silver (2009) e Haan (2009) consideram ser melhor a utilização do preço imputado em ambos os termos da razão entre os preços relativos. Essa abordagem é denominada de imputação dupla (ID). Particularmente, todos os preços relativos observados são utilizados na construção do índice de preços, mas a regressão utilizada tem por objetivo estimar a razão entre os preços, e não somente o preço *missing*. Matematicamente, a representação para o índice de Laspeyres é dada por:

$$I_{0,1}^{L,IS} = \sum_{m=1}^{M} w_{m,0} \frac{P_{m,1}}{P_{m,0}} + \sum_{d=1}^{D} w_{d,0} \frac{\hat{P}_d^1(x_d^0)}{\hat{P}_d^0(x_d^0)}, \tag{13}$$

em que a imputação ocorre para o termo $\frac{P_{d,1}^u}{P_{d,0}} = \frac{\hat{P}_d^1(x_d^0)}{\hat{P}_d^1(x_d^0)}$.

Aizcorbe (2014) propõe uma abordagem de imputação completa (IC), para a qual todos os preços são imputados pelos valores preditos da regressão estimada. Especificamente para o índice de Laspeyres, a imputação completa apresenta a seguinte forma:

$$I_{0,1}^{L,IC} = \sum_{m=1}^{M} w_{m,0} \frac{\hat{P}_m^1(x_m^0)}{\hat{P}_m^0(x_d^0)} + \sum_{d=1}^{D} w_{d,0} \frac{\hat{P}_d^1(x_d^0)}{\hat{P}_d^0(x_d^0)} \tag{14}$$

É possível notar, pela expressão (14), que todos os preços são substituídos pelos seus valores preditos, obtidos por meio da regressão hedônica. Aizcorbe (2014) destaca que, devido à quantidade de valores estimados na composição do índice, esse pode não ser um método adequado, pois pode apresentar "ruído" além do desejado, por isso é frequentemente desencorajado o uso dessa proposta.

4 AVANÇOS EM ÍNDICES DE PREÇO PARA O MERCADO IMOBILIÁRIO

Uma das primeiras propostas para a construção de índices de preço para o mercado imobiliário refere-se ao estudo de Bailey, Muth e Nourse (1963). Os autores discutem o desafio da construção de índices de preço para o bem habitação, em especial a dificuldade de controlar a estimação do índice de preços para itens homogêneos pouco transacionados. Segundo Bailey, Muth e Nourse (1963), o maior problema é devido a grande variação na qualidade das habitações consideradas. Dessa forma, os índices de preço com base nos preços médios de venda de todas as propriedades de algum tipo de produto negociado, em um determinado período, estão suscetíveis a duas deficiências: primeiro, variação na qualidade dos imóveis vendidos a partir de um período a outro fará com que o índice varie mais amplamente do que o valor de qualquer propriedade; segundo, se há uma alteração progressiva na qualidade das propriedades vendidas em diferentes períodos, o índice apresentará um viés em sua construção.

Para contornar as dificuldades apresentadas, Bailey, Muth e Nourse (1963) estabeleceram o método que consiste em uma regressão do logaritmo da mudança nos preços contra variáveis categóricas (*dummies*) do período da venda, denominado *análise de regressão dos preços por meio de vendas repetidas*.

Nadalin e Furtado (2011) afirmam que o problema dessa proposta é a ausência de uma estratégia sobre como lidar com períodos de tempo diferentes entre as unidades até a segunda venda, além de não controlar se a unidade vendida pela segunda vez teve alguma reforma, ou se houve depreciação. Ademais, caso as unidades com maior liquidez no mercado não sejam representativas do mercado como um todo, o índice poderá será enviesado.

Com o objetivo de controlar a heterogeneidade frequentemente presente nos dados do mercado imobiliário, Goodman (1978) construiu uma regressão hedônica com ajuste de Box e Cox (1964) para a forma funcional do modelo de regressão. Apesar de não ser o primeiro texto a utilizar regressão hedônica como ferramenta de modelagem de dados em habitação, Goodman (1978) afirma que raros foram os estudos até a data de publicação do seu texto que utilizaram essa abordagem para o mercado imobiliário.

Can e Megbolugbe (1997), utilizando a estrutura de dependência espacial frequentemente presente em dados do mercado imobiliário, afirmam que a construção de índices de preço para imóveis deve considerar essa estrutura espacial em seus modelos, pois a omissão dessa dependência espacial pode afetar a precisão das estimativas de preço. Segundo Can e Megbolugbe (1997), há consenso na literatura sobre estimação de índices de preço de que a abordagem de regressão hedônica é a mais adequada para a construção de índices de preço de qualidade em dados transversais.

Segundo Can e Megbolugbe (1997), o primeiro passo na construção de índices de preço para imóveis envolve a estimação de uma função de preços hedônicos, a qual segue a forma geral:

$$P = f\,(S\beta, N\gamma) + \epsilon, \qquad (15)$$

em que P é um vetor de preços observados para os imóveis considerados; S é uma matriz de atributos estruturais que contém variáveis como a idade e o tamanho do imóvel; N é uma matriz de características da vizinhança, incluindo medidas socioeconômicas para a região de interesse, como características ambientais da região e serviços públicos disponíveis; β

e γ são vetores de parâmetros correspondentes às matrizes S e N; finalmente ϵ é o termo aleatório do modelo (15).

Can e Megbolugbe (1997) afirmam que, devido à natureza dos dados disponíveis para a construção de índices de preço para imóveis, a suposição de independência, bem como a identidade distribucional do termo aleatório, não é válida, e, portanto, há uma covariância não nula entre termos estocásticos distintos no modelo (15).

Nesse sentido, é importante considerar a dependência espacial, pois esta afetará a validade estatística dos resultados, uma vez que a sua presença leva a estimativas tendenciosas da variância residual, além de ineficiência dos coeficientes do modelo, quando a estimação por mínimos quadrados ordinária é utilizada. Outras estatísticas prejudicadas quando a dependência espacial não é corretamente considerada são os testes-padrão, como Teste F e Teste T, para o modelo e parâmetros respectivamente.

Assim, para considerar a frequente dependência espacial nos preços dos imóveis, Can e Megbolugbe (1997) propuseram adicionar à estrutura funcional (15) um termo autorregressivo espacial para o preço dos imóveis, seguindo a estrutura *Spatial Autoregressive Model* (SAR), desenvolvida por Anselin (1988). Essa proposta é capaz de considerar dois efeitos existentes para dados de habitação (Can, 1992), quais sejam: efeito de adjacência, o qual é uma externalidade associada à localização absoluta do imóvel e estrutura; e efeito de vizinhança, o qual é o conjunto de características espaciais, tais como vizinhança, acessibilidade e previsão de serviços públicos.

Uma fragilidade da proposta de Can e Megbolugbe (1997) é a mesma de todo índice de preços construído por meio de regressão hedônica: os índices precisam ser atualizados à medida que novas informações estão disponíveis. Por vezes, o analista deseja que os índices de preço passados sejam fixos. Nesse sentido, Bourassa, Hoesli e Sun (2006) propuseram uma abordagem para o cômputo de índices de preço em mercados imobiliários que mantém fixos os valores passados.

Os autores denominaram de *Sale Price Appraisal Ratio* (Spar) a abordagem proposta, a qual combina as informações de avaliações e os preços de venda para a construção do índice de preços de imóveis. Em contraste com o método de vendas repetidas, o método Spar utiliza todas as transações que ocorreram em um determinado mercado imobiliário; portanto, é menos propenso a viés de seleção da amostra. Além disso, o índice construído não necessita de atualização quando novos dados estão disponíveis para novos períodos de tempo, o que não é o caso para o método de vendas repetidas e o método da regressão hedônica.

O índice da relação entre avaliação e preço de venda (Spar) é um índice de repetição aritmética. A primeira medida de cada par de repetições é a da avaliação do governo da propriedade, enquanto a segunda medida é o preço efetivo da transação. Segundo Bourassa, Hoesli e Sun (2006), uma vantagem de usar a avaliação oficial como medida para o primeiro par de repetições é que todas as avaliações de uma área geográfica são tomadas tipicamente a partir de uma determinada data, o que significa que os imóveis vendidos em um determinado período normalmente possuem avaliações em um período de base única para fins de comparação. Isto simplifica o cálculo do índice, uma vez que não há necessidade de utilizar uma técnica de estimação. Matematicamente, o índice é dado por:

$$I_{0,1}^{SPAR} = \left\{ \left[\sum_{j=1}^{n_1} \left(\frac{Pj1}{Aj0} \right) \div n_1 \right] \left[\sum_{j=1}^{n_0} \left(\frac{Pj0}{Aj0} \right) \div n_0 \right] \right\} \tag{16}$$

em que $t = 0{,}1$ são dois períodos de tempo considerados; P_{jt} e A_{jt} são os valores para o preço e a avaliação do j-ésimo imóvel no período t; e n_t é o total de observações disponíveis no período t. Por meio da relação entre valor de avaliação e preço negociado, a heterogeneidade entre os bens é tratada, uma vez que uma base de comparação homogênea é criada ao longo do tempo.

Mais recentemente, Dorsey *et al.* (2010) propuseram uma regressão hedônica para a construção de índices de preço de mercados imobiliários que considerassem também a dependência espacial. Empiricamente, os resultados encontrados por meio da proposta desenvolvida se aproximaram das estimativas obtidas pelo método de vendas repetidas, fornecendo indícios da estabilidade do índice de preços estimado por Dorsey *et al.* (2010).

5 BREVÍSSIMA ILUSTRAÇÃO: ÍNDICES HEDÔNICOS DE PREÇO DE IMÓVEIS PARA SÃO PAULO

A título de ilustração, apresentam-se a seguir os passos metodológicos e os resultados para a construção de um índice de preços hedônicos com *dummies* para o caso de São Paulo, a partir de dados disponibilizados pelo Centro de Estudos da Metrópole (CEM). Ademais, apresentam-se os resultados em conjunto com o índice de Laspeyres, o índice geral de preços do mercado (IGP-M) e o índice calculado pelo Banco Central, Índice de

Valores de Garantia de Imóveis Residenciais Financiados (IVG-R),[2] no intuito de facilitar a comparação entre os índices.

5.1 Base de dados

A análise utilizou a Base de Lançamentos Imobiliários Residenciais da Região Metropolitana de São Paulo.[3] A base contém os dados de lançamento comercializados na cidade de São Paulo. Como os dados estão disponíveis em formato *shapefile*, a primeira etapa foi realizar a leitura dos arquivos de interesse para os anos posteriores a 2000. A base residencial contém 10.585 observações, com a seguinte tabela de frequências por ano (tabela 1):

TABELA 1
Frequência de lançamentos imobiliários por ano constantes na base utilizada (2000-2013)

Ano	Número de lançamentos
2000	558
2001	521
2002	608
2003	684
2004	659
2005	617
2006	656
2007	867
2008	874
2009	765
2010	982
2011	992
2012	758
2013	1.044

Elaboração dos autores.

Note que há consistência do número médio de lançamentos ao ano, com ligeiro aumento a partir de 2007. É prudente remover os *outliers* da base, por isso removeram-se o quinto e o 95º centil empírico do valor do preço das habitações registradas, conforme código constante do apêndice, obtendo-se, com isso, uma base final com 10.038 observações.

2. Ver descrição básica do IVG-R em: <http://goo.gl/vNohBw>.
3. Disponível em: <http://goo.gl/IaAICl>.

Em seguida, construíram-se as variáveis *dummies* para cada período de tempo, exceto para a primeira data, a qual será a base do índice de preços. As variáveis selecionadas como candidatas ao modelo de preços hedônicos foram:

- DORM_UNID: quantidade de dormitórios por unidade.
- BANH_UNID: quantidade de banheiros por unidade.
- GAR_UNID: quantidade de vagas por unidade.
- ELEV: quantidade de elevadores por lançamento.
- COB: quantidade de unidades na cobertura.
- BLOCOS: quantidade de blocos no empreendimento.
- UNIDAND: quantidade de unidades por andar.
- ANDARES: quantidade de andares do empreendimento.
- AR_UT_UNID: área útil da unidade em m².
- PC_TT_UN: variável dependente –preço de venda da unidade.

As estatísticas descritivas das variáveis utilizadas (tabela 2) indicam apartamentos em média com três dormitórios, dois banheiros e duas vagas de garagem, com elevador e em bloco de apartamento único, com quatro unidades por andar e 14 andares. O preço médio foi de cerca de R$ 440 mil, e a mediana dos preços foi de pouco mais de R$ 260 mil. Os apartamentos são pequenos, com média de pouco mais de 100 m^2 e mediana de 76.

TABELA 2
Estatísticas descritivas da base

Variável	Observações	Média	Desvio-padrão	Mediana	Mínimo	Máximo
DORM_UNID	10.038	2,71	0,88	3	1	6
BANH_UNID	10.038	2,02	0,88	2	1	6
GAR_UNID	10.038	1,93	1,16	2	0	12
ELEV	10.038	1,85	1,94	2	0	26
COB	10.038	1,1	3,17	0	0	92
BLOCOS	10.038	1,05	1,29	1	0	30
UNIDAND	10.038	4,2	3,67	4	0	60
ANDARES	10.038	13,33	8,14	14	0	41
AR_UT_UNI	10.038	105,91	86,66	76	12,72	1.975
PC_TT_UN	10.038	441.621	653.589	262.650	66.150	13.929.850

Elaboração dos autores.

5.2 Resultados

O próximo passo é a estimação do modelo via mínimos quadrados ordinários (MQOs).

TABELA 3
Resultados da estimação MQO

Variável	Estimativa	Erro-padrão	*t* valor	Pr(>|t|)
(Intercepto)	10,4996206	0,0804569	130,499972	0,0000000
DORM_UNID	-0,0422304	0,0067031	-6,300147	0,0000000
BANH_UNID	0,2752171	0,0081864	33,618835	0,0000000
GAR_UNID	0,2193424	0,0064675	33,914461	0,0000000
ELEV	0,0089332	0,0022302	4,005620	0,0000623
COB	0,0061409	0,0011325	5,422390	0,0000001
BLOCOS	-0,0626714	0,0030629	-20,461347	0,0000000
UNIDAND	0,0044436	0,0012224	3,635236	0,0002791
ANDARES	0,0154606	0,0005339	28,959739	0,0000000
AR_UT_UNID	0,0035208	0,0000784	44,899609	0,0000000

Elaboração dos autores.

O modelo apresenta ajustado igual a 0,8439428. Em seguida, obtêm-se as estimativas para os índices de preço aplicando a função exponencial para cada parâmetro estimado das variáveis *dummies*.

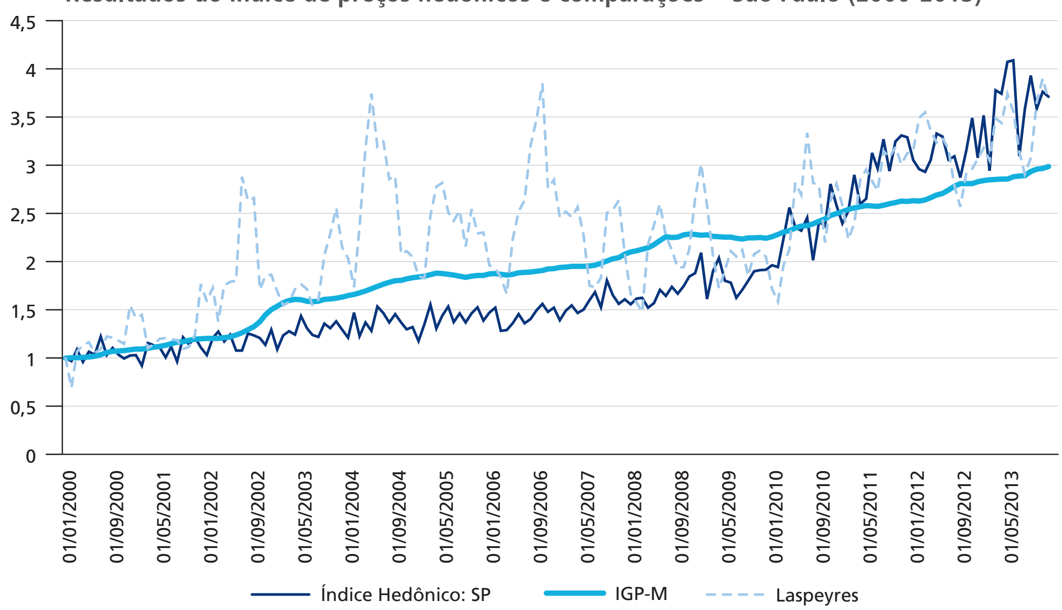
Por se tratar de uma regressão hedônica para estimação dos índices de preço no mercado imobiliário residencial em São Paulo, o interesse está na estimação dos coeficientes de regressão para os períodos de tempo (variáveis binárias), pois esses referem-se ao logaritmo natural do índice de preços no período correspondente. Assim, a estimativa adequada para índices de preço é obtida aplicando-se a função exponencial para cada parâmetro relativo ao tempo nos valores estimados da regressão. As demais variáveis, como número de dormitórios, banheiros etc., são apenas controles para garantir que os valores do logaritmo natural do preço de venda sejam relacionados somente a imóveis com características "similares", isto é, condicionados nas demais características estruturais do bem, e, portanto, não são de interesse direto dos pesquisadores interessados na construção dos índices de preço.

O modelo apresenta bom ajuste, e os resultados indicam que o mercado de lançamentos em São Paulo apresenta crescimento contínuo no período de 2000 a 2008, com alteração relevante após 2008 – qual seja, aumento na variabilidade e nos preços médios para o período posterior, conforme observado no gráfico 1.

O gráfico 1 também apresenta outros resultados interessantes. Em primeiro lugar, demonstra como a variabilidade de um índice de tendência central, tal como o índice de Laspeyres, possivelmente não é o mais adequado para a análise de imóveis, dada a sua grande variabilidade. O índice de preços hedônicos também apresenta variabilidade grande, mas de menor magnitude, se comparado ao índice de Laspeyres.

A variabilidade no índice hedônico, ainda que controlada pelas características do imóvel, provavelmente deriva exatamente da dificuldade de se comparar a "cesta" de bens imóveis (Sweeney, 1974; Arnott, 1987; Dubin, 1992). Ainda que todas as características do imóvel sejam similares, sua localização é sempre única. Portanto, suas relações de proximidade e interação com o restante do ambiente urbano também é singular (Whitehead, 1999). Ainda assim, pode-se considerar a possibilidade de que, de fato, os preços no mercado variaram de forma bastante intensa no período.

GRÁFICO 1
Resultados do índice de preços hedônicos e comparações – São Paulo (2000-2013)



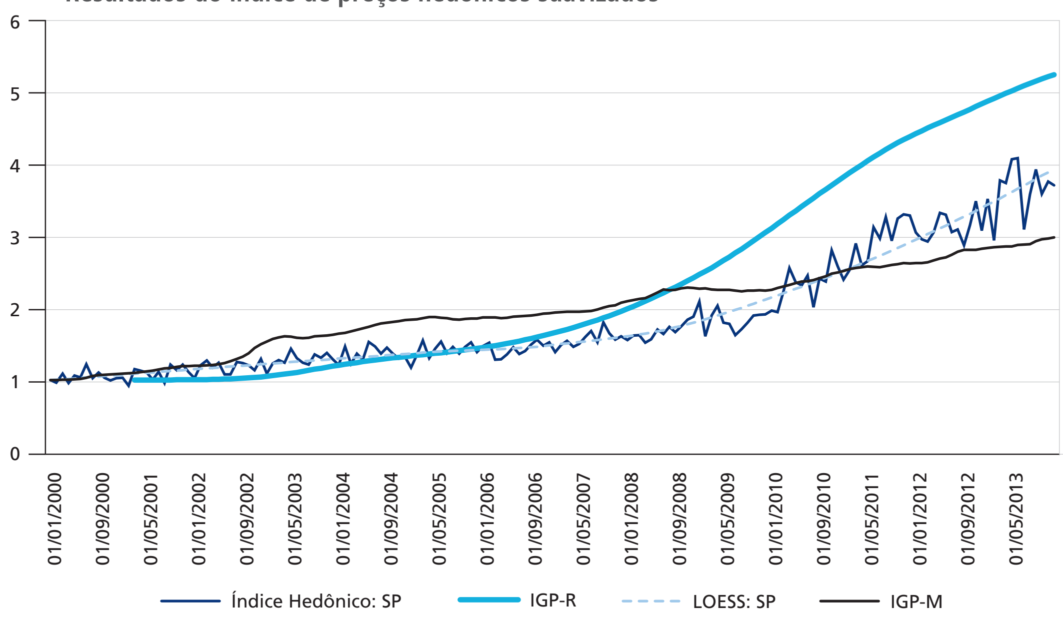
Elaboração dos autores.

No sentido de reduzir parte da variabilidade e garantir maior estabilidade aos usuários do índice, com variações menos bruscas de período a período, pode-se considerar a adoção de suavização, como por meio da abordagem de Loess (gráfico 2).

O gráfico 2 indica que o comportamento geral e ao final do período do índice de "aluguéis", o IGP-M e o IVG-R apresentam comportamentos bastante distantes um do outro.[4] Enquanto o IVG-R cresce quase 120% de dezembro de 2008 a dezembro de 2013, o IGP-M cresce apenas cerca de 30% no mesmo período. Isso indica, no mínimo, particularidades do mercado imobiliário que precisam ser tratadas por índice de preços específico.

O índice hedônico suavizado apresenta comportamento mais próximo do IGP-M ao longo de toda a série, o primeiro com variação de cerca de 200% e o segundo, 300%. O IVG-R, por sua vez, avançou 425% no período. No entanto, a variação do IVG-R e do índice suavizado Loess, a partir de 2008, foram similares com 117% e 119% respectivamente.

GRÁFICO 2
Resultados do índice de preços hedônicos suavizados



Elaboração dos autores.

A base de dados disponível para o estudo de caso (como, ademais, para o restante do país) não permite, por exemplo, a aplicação do método de Spar, sugerido por Bourassa,

4. O índice FipeZap não foi incluído, dado que sua disponibilidade não compreende o período todo e sua construção se baseia na oferta de imóveis apenas.

Hoesli e Sun (2006). Ao mesmo tempo, o método de vendas repetidas (discutido em outro texto deste mesmo projeto) é de difícil aplicação para o país, dada a indisponibilidade de dados. Com isso, a ilustração acima foi feita de acordo com as possibilidades da base. Dado que a premissa de um índice de preços de imóveis nacional deveria ser de fácil entendimento e aplicação, não aplicamos a regressão espacial.

6 CONSIDERAÇÕES FINAIS

Este texto apresenta uma revisão dos índices de preço mais comuns na literatura, detalhando as peculiaridades de índices para imóveis e ilustrando com o caso de São Paulo. O texto diferencia os índices de Laspeyres e Paasche a partir de seus distintos métodos de ponderação, porém, mantendo a cesta de consumo fixa para os períodos analisados. O índice de Fisher é uma proposta de agregação dos índices de Laspeyres e Paasche, oferecendo a vantagem da homogeneidade e simetria.

Os índices hedônicos – nos quais há ênfase nas características constituintes do bem – são adequados para imóveis, cujos atributos influenciam diferentemente o preço final único do bem indivisível. O método de variáveis *dummies* permite o controle dessas características típicas de cada imóvel. Já o método da imputação traz mais flexibilidade na escolha da forma funcional.

Recentemente, a dependência espacial e a utilização de valores atribuídos por avaliadores foram consideradas na literatura de índices de preço de imóveis. Todavia, sua aplicação é restrita à disponibilidade de dados. Finalmente, ilustrou-se a aplicação da construção de um índice de preços hedônicos para o mercado de lançamentos de imóveis em São Paulo. O índice gerado indica aumento nos preços e na variabilidade a partir de 2008.

REFERÊNCIAS

AIZCORBE, A. M. **A practical guide to price index and hedonic techniques**. Oxford: Oxford University Press, 2014.

ANSELIN, L. **Spatial econometrics**: methods and models. Springer: Springer Netherlands, 1988.

ARNOTT, R. Economic theory and housing. *In*: MILLS, E. S. (Ed.). **Handbook of Regional and Urban Economics**. Amsterdam: Elsevier Science Publishers, 1987.

BAILEY, M. J.; MUTH, R. F.; NOURSE, H. O. A regression method for real estate price index construction. **Journal of the American Statistical Association**, v. 58, n. 304, p. 933-942, 1963.

BOURASSA, S. C.; HOESLI, M.; SUN, J. A simple alternative house price index method. **Journal of Housing Economics**, v. 15, n. 1, p. 80-97, 2006.

BOX, G. E.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**. Series B (Methodological), p. 211-252, 1964.

CAN, A. Specification and estimation of hedonic housing price models. **Regional Science and Urban Economics**, v. 22, n. 3. p. 453-474, 1992.

CAN, A.; MEGBOLUGBE, I. Spatial dependence and house price index construction. **The Journal of Real Estate Finance and Economics**, v. 14, n. 1-2, p. 203-222, 1997.

CASE, K. E.; SHILLER, R. J. Prices of single-family homes since 1970: new indexes for four cities. **New England Economic Review**, p. 45-56, Sept. 1987.

COURT, A. Hedonic price indexes with automotive examples. **The Dynamics of Automobile Demands**. General Motors, p. 99-119, 1939.

DIEWERT, W. E. Commentary. **Federal Reserve Bank of St. Louis Review**, May-Jun. 1997.

DIEWERT, W. E. Index numbers. *In*: DURLAUF, S. N.; BLUME, L. E. (Eds.). **The new Palgrave Dictionary of Economics**, 2. ed. Basingstoke: Palgrave Macmillan. 2008.

DIEWERT, W. E.; HERAVI, S.; SILVER, M. Hedonic imputation versus time dummy hedonic indexes. *In*: **Price index concepts and measurement**. Chicago: University of Chicago Press, 2009.

DORSEY, R. E. *et al*. Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle. **Journal of Housing Economics**, v. 19, n. 2, p. 75-93, Jun. 2010.

DUBIN, R. A. Spatial autocorrelation and neighborhood quality. **Regional Science and Urban Economics**, v. 22, p. 433-452, 1992.

GOODMAN, A. C. Hedonic prices, price indices and housing markets. **Journal of Urban Economics**, v. 5, n. 4, p. 471-484, 1978.

GRILICHES, Z. Hedonic price indexes for automobiles: An econometric of quality change. **In The Price Statistics of the Federal Goverment**, p. 173-196, 1961, New York: National Bureau of Economic Research.

GUJARATI, D. N. **Basic econometrics**. India: Tata McGraw-Hill Education, 2012.

HAAN, J. Hedonic imputation *versus* time dummy hedonic indexes. *In*: DIEWERT, E.;

GREENLEES, J.; HULTEN, C. R. **Price index concepts and measurement**. Chicago: University of Chicago Press, 2009.

NADALIN, V. G.; FURTADO, B. A. Índices de preços para imóveis: uma revisão. **Boletim Regional, Urbano e Ambiental**, Brasília, n. 6, dez. 2011. Disponível em: <http://goo.gl/sV7YwU>.

PAKES, A. A Reconsideration of hedonic price indices with an application to PC's. **The American Economic Review**, v. 93, n. 5, p. 1578-1596, 2003.

SILVER, M.; HERAVI, S. The difference between hedonic imputation indexes and time dummy hedonic indexes. **Journal of Business & Economic Statistics**, v. 25, n. 2, p. 239-246, 2007.

SWEENEY, J. A commodity hierarchy model of rental housing market. **Journal of Urban Economics**, v. 1, p. 288-323, 1974.

TRIPLETT, J. Handbook on hedonic indexes and quality adjustments in price indexes. **Directorate for Science, Technology and Industry**, Paris, 2004.

WHITEHEAD, C. M. E. Urban housing markets: theory and policy. *In*: CHESHIRE, P.; MILLS, E. S. (Ed.). **Handbook of Regional and Urban Economics**. Amsterdam, North-Holland: Elsevier, 19999. v. 3, p. 1559-1594.

WOOLDRIDGE, J. **Introductory econometrics**: a modern approach. Stamford: Cengage Learning, 2012.

APÊNDICE

CÓDIGO DE REGRESSÃO HEDÔNICA (R)

```
#Faz a leitura dos shapes
require(maptools)
shape <- as.data.frame(readShapePoints("LanRes_85_13_RMSP_CEM.shp"))
#Remove os valores NA
shape<-shape[!is.na(shape$PC_TT_UN),]
#Mantem somente dados posteriores a 2000
shape<-shape[which(shape$ANO_LAN>1999),]
#Tabela de frequencias
tt<-as.data.frame(table(shape$ANO_LAN))
colnames(tt)<-c("Ano","NÃºmero de lancamentos")
library(knitr)
kable(tt)
#Encontra os centis
centis<-quantile(shape$PC_TT_UN, probs=c(0.05,0.95))
shape<-subset(shape,PC_TT_UN>centis[1],PC_TT_UN<centis[2])
#Aplica o logaritmo natural no preco nominal de venda da unidade
lnPreco <- log(shape$PC_TT_UN)
#Define a base com as informacoes suficientes
dados<-data.frame(lnPreco)
#Cria as variaveis dummies
library(dummies)
dummy<-dummy(shape$MES_LAN)
#Remove o periodo base
ibase<-which(colnames(dummy)=="MES_LAN15-JAN-2000")
dummy<-dummy[,-ibase]
#Mantem somente as variaveis de interesse
```

```
vars<-c("DORM_UNID","BANH_UNID","GAR_UNID","ELEV",
    "COB","BLOCOS","UNIDAND","ANDARES",
    "AR_UT_UNID")
#Define a base com as informacoes suficientes
shape<-shape[,vars]
#Une com as variaveis dummies e lnPreco
shape<-cbind(lnPreco,shape,dummy)
#Converet AR_UT_UNID para numeric
shape$AR_UT_UNID<-as.numeric(gsub(",",".",as.character(shape$AR_UT_UNID)))
#Remove as observacoes com preco zero
shape<-shape[shape$lnPreco>-Inf,]
#Estima modelo
mod<-lm(lnPreco~.,data=shape)
sum<-summary(mod)
tab<-sum$coef[1:10,]
library(knitr)
kable(tab)
#Linhas
sum<-as.data.frame(coef(sum))
sum<-sum[11:nrow(sum),]
linhas<-sum[,"Estimate"]
#Obtem as estimativas para os indices de precos
coef<-data.frame(exp(linhas))
#Cria o Mes e Ano
coef$Text<-rownames(sum)
coef$Text<-gsub("MES_LAN15-","",coef$Text)
coef$Text<-gsub("`","",coef$Text)
Text<-as.data.frame(do.call("rbind",strsplit(coef$Text,"-")))
Text$V1<-as.character(Text$V1)
Text[which(Text$V1=="JAN"),"V1"]<-'1'
```

```
Text[which(Text$V1=="FEB"),"V1"]<-'2'
Text[which(Text$V1=="MAR"),"V1"]<-'3'
Text[which(Text$V1=="APR"),"V1"]<-'4'
Text[which(Text$V1=="MAY"),"V1"]<-'5'
Text[which(Text$V1=="JUN"),"V1"]<-'6'
Text[which(Text$V1=="JUL"),"V1"]<-'7'
Text[which(Text$V1=="AUG"),"V1"]<-'8'
Text[which(Text$V1=="SEP"),"V1"]<-'9'
Text[which(Text$V1=="OCT"),"V1"]<-'10'
Text[which(Text$V1=="NOV"),"V1"]<-'11'
Text[which(Text$V1=="DEC"),"V1"]<-'12'
coef<-cbind(exp(sum[,1]),Text)
colnames(coef)<-c("Indice Hedonico","Mes","Ano")
coef$Mes<-as.numeric(as.character(coef$Mes))
#Ordena a base
coef<-coef[ order(coef$Ano,coef$Mes), ]
#Adicona a base
base<-data.frame(`Indice Hedonico`=1.0,Mes=1,Ano="2000")
colnames(base)<-colnames(coef)
coef<-rbind(base,coef)
coef$Tipo<-"Indice Hedonico"
colnames(coef)[1]<-"Indice"
#Cria a data coef$Data<-as.Date(paste0("15/",coef$Mes,"/",coef$Ano), "%d/%m/%Y")
#Faz o Grafico:
library(ggplot2)
library(ggthemes)
ggplot(coef,aes(x=Data,y=`Indice`,colour=Tipo,group=Tipo)) + geom_line(lwd=0.5)+
  xlab("Data") +  ylab("Indice de precos.") + ggtitle("Indice de precos para habitacao.\n
Sao Paulo.")+
  theme_economist()+scale_colour_economist(name="")
```

```
  coef$Time<-seq(1,nrow(coef))
lo2 <- loess(coef[which(coef$Tipo==”Indice Hedonico”),1] ~ coef[which(coef$Tipo==”Indice Hedonico”),”Time”])
coef[which(coef$Tipo==”Indice Hedonico”),1]<-lo2$fitted
#Faz o Grafico:
library(ggplot2)
library(ggthemes)
ggplot(coef,aes(x=Data,y=`Indice`,colour=Tipo,group=Tipo)) + geom_line(lwd=0.5)+
  xlab(“Data”) +  ylab(“Indice de precos.”) + ggtitle(“Indice de precos para habitacao.\n Sao Paulo.”)+
  theme_economist()+scale_colour_economist(name=””)
```

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Coordenação
Cláudio Passos de Oliveira

Supervisão
Everson da Silva Moura
Reginaldo da Silva Domingos

Revisão
Clícia Silveira Rodrigues
Idalina Barbara de Castro
Leonardo Moreira Vallejo
Marcelo Araujo de Sales Aguiar
Marco Aurélio Dias Pires
Olavo Mesquita de Carvalho
Regina Marta de Aguiar
Alessandra Farias da Silva (estagiária)
Paulo Ubiratan Araujo Sobrinho (estagiário)
Pedro Henrique Ximendes Aragão (estagiário)
Thayles Moura dos Santos (estagiária)

Editoração
Bernar José Vieira
Cristiano Ferreira de Araújo
Daniella Silva Nogueira
Danilo Leite de Macedo Tavares
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa
Raul Vinicius Fernandes Gonçalves (estagiário)

Capa
Luís Cláudio Cardoso da Silva

Projeto Gráfico
Renato Rodrigues Bueno

The manuscripts in languages other than Portuguese published herein have not been proofread.

Livraria Ipea

SBS – Quadra 1 - Bloco J - Ed. BNDES, Térreo.
70076-900 – Brasília – DF
Fone: (61) 2026-5336

Correio eletrônico: livraria@ipea.gov.br

Composto em adobe garamond pro 12/16 (texto)
Frutiger 67 bold condensed (títulos, gráficos e tabelas)
Brasília-DF

Missão do Ipea

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ISSN 1415-4765

ipea Instituto de Pesquisa Econômica Aplicada

Ministério do Planejamento

GOVERNO FEDERAL BRASIL PÁTRIA EDUCADORA