

# Construção de índices de preços de imóveis para o Distrito Federal por meio de vendas repetidas e GWR

*House Price Index for Distrito Federal using Repeat Sales Model and GWR*

*Pedro Henrique Melo Albuquerque* <sup>(1)</sup>

*Vanessa Gapriotti Nadalin* <sup>(2)</sup>

*Vicente Correia Lima Neto* <sup>(3)</sup>

*Mariana Rosa Montenegro* <sup>(4)</sup>

<sup>(1)</sup> Universidade de Brasília

<sup>(2)</sup> Instituto de Pesquisa Econômica Aplicada

<sup>(3)</sup> Instituto de Pesquisa Econômica Aplicada

<sup>(4)</sup> Universidade de Brasília

## Abstract

The need for monitoring the real estate market to evaluate the effectiveness of urban planning policies in the territory is crucial step towards good governance, in this sense, this article aims to build the price index for properties transacted in Distrito Federal between 2003 and 2012 through the repeat sales method. The proposed index allows the evaluation of public policies and their territorial impact on the price of real estate, through the assessment of the evolution of real estate prices in the Distrito Federal. This continuous monitoring constitutes a tool of urban policy in the federal capital, which will serve not only for evaluation of public policies, as well as implementation of the same, from the integration with urban instruments, such as the onerous grant.

## Keywords

price index; real estate; repeated sales; GWR.

**JEL Codes** C43; C21; C31.

## Resumo

*A necessidade de monitoramento do mercado imobiliário como instrumento de avaliação da efetividade das políticas de planejamento urbano no território é etapa crucial para a boa gestão pública. Nesse sentido, o presente texto tem por objetivo construir o índice de preços para os imóveis transacionados no Distrito Federal entre 2003 e 2012 por meio do método de vendas repetidas. O índice proposto permite a avaliação de políticas públicas e seus impactos territoriais no preço dos bens imóveis, por meio do monitoramento contínuo da evolução dos preços de imóveis no Distrito Federal. Esse acompanhamento contínuo constitui-se em instrumento da política urbana no âmbito da capital federal, o qual servirá não somente para avaliação de políticas públicas, como também de operacionalização das mesmas, a partir da integração com instrumentos urbanísticos, tais como a outorga onerosa.*

## Palavras-chave

*índice de preços; mercado imobiliário; vendas repetidas; GWR.*

**Códigos JEL** C43; C21; C31.

## 1 Introdução

A necessidade de monitoramento do mercado imobiliário como instrumento de avaliação da efetividade das políticas de planejamento urbano no território é etapa crucial para a boa gestão pública (Kapp, 2012). Nesse sentido, há lacunas quanto à construção de índices que possam acompanhar a evolução dos preços de imóveis no país, seja devido à escassez de dados, seja pelas dificuldades na elaboração de um índice de preços que incorpore adequadamente as características do bem habitação, que se referem tanto aos seus atributos físicos quanto à sua localização. Em outras palavras, para esse tipo de bem, os preços deveriam refletir os custos de construção das estruturas físicas, bem como a valorização das amenidades urbanas, sendo, portanto, dependente de atributos estruturais e espaciais.

Vale ressaltar que o mercado imobiliário também sofre influência de um contingente de variáveis, em especial as macroeconômicas. Além do enfoque urbanístico, a taxa de juros, o crédito, a renda e a confiança do consumidor são variáveis relevantes para esse mercado.

Segundo Nadalin e Furtado (2010), por serem muito distintas umas das outras em suas características, além de apresentarem um nível de qualidade variável no tempo, as habitações apresentam dificuldades para construção de índices de preços. O cerne de tal dificuldade consiste em acompanhar no tempo as mudanças ocorridas no preço de um produto idêntico. Quando se trata de índice de preços ao consumidor, é acompanhada uma cesta de produtos que não varia no tempo, facilitando, desse modo, a construção de índices de preços nessa situação.

Ademais, o bem habitação é indivisível, heterogêneo e durável, características essas que associadas com a baixa frequência de negociação deste dificultam a operacionalização da construção de índices de preços por meio das abordagens tradicionais, como os métodos de Laspeyres e Paasche (Sadoulet; Janvry, 1995). O principal desafio, então, é controlar as diferenças e alterações da qualidade das unidades efetivamente transacionadas em cada período de aferição dos preços observados.

Entre as metas que obtiveram maior sucesso para a transferência da capital da república para o Distrito Federal destacam-se a ocupação do interior do Brasil e o desenvolvimento econômico das regiões mais longínquas do país, no entanto, nos últimos anos, a cidade se desenvolveu de tal

forma que exacerbou a desigualdade entre pobres e ricos, conforme dados do Instituto de Pesquisa Econômica Aplicada (IPEA):

Tabela 1 Índice de GINI por estado em 2013

<b>Estado</b>	<b>Gini 2013</b>
Distrito Federal	0.578
Sergipe	0.560
Maranhão	0.560
Bahia	0.558
Amazonas	0.543
Rio Grande do Norte	0.541
Rio de Janeiro	0.532
Roraima	0.531
Paraíba	0.525
Alagoas	0.525
Acre	0.525
Amapá	0.522
Tocantins	0.519
Piauí	0.515
Ceará	0.514
Mato Grosso	0.505
Pernambuco	0.502
Pará	0.502
Mato Grosso do Sul	0.497
Espírito Santo	0.494
São Paulo	0.494
Minas Gerais	0.489
Goiás	0.484
Rio Grande do Sul	0.478
Rondônia	0.476
Paraná	0.469
Santa Catarina	0.435

Fonte: Instituto de Pesquisa Econômica Aplicada – Ipeadata.

Outro fator que afeta a distribuição espacial da renda no Distrito Federal é o número de servidores públicos residentes na cidade, os quais representam boa parte das pessoas economicamente ativas no DF.

Tais características fazem com que os imóveis no Distrito Federal apresentem padrão específico. Em Brasília, por exemplo, boa parte dos órgãos públicos federais e distritais encontram-se no centro da cidade e, portanto, os imóveis mais caros estão próximos da região central, aqueles trabalhadores que podem dispender de maiores valores adquirem imóveis próximos dos seus trabalhos, formando assim regiões no espaço onde os valores dos imóveis são mais caros devido, especialmente, a sua localização e não a sua estrutura como metragem ou número de quartos, por exemplo.

Dessa forma, o presente texto tem por objetivo apresentar a construção do índice de preços para o mercado imobiliário no Distrito Federal por meio da abordagem de vendas repetidas (Bailey; Muth; Nourse, 1963; Case; Shiller, 1987), metodologia que defendemos ser a que melhor neutraliza no índice as mudanças nas características das unidades vendidas. Um dos usos deste índice, e objetivo de futura pesquisa, é a avaliação de políticas públicas e seus impactos territoriais no preço dos bens imóveis e em tendências de flutuações do mercado imobiliário, por meio do monitoramento contínuo da evolução dos preços de imóveis no Distrito Federal. Esse acompanhamento contínuo constitui-se em instrumento da política urbana no âmbito da capital federal, o qual servirá não somente para avaliação de políticas públicas, como também para operacionalização das mesmas, a partir da integração com instrumentos urbanísticos, tais como a outorga onerosa.

O trabalho está dividido da seguinte maneira: a seção 2 apresenta a revisão teórica sobre a construção de índices de preços, a seção 3 apresenta os dados obtidos por meio do Acordo de Cooperação Técnica entre o Instituto de Pesquisa Econômica Aplicada – IPEA e a Secretaria de Estado de Gestão do Território e Habitação do Governo do Distrito Federal – SEDHAB/GDF – utilizados para a construção do índice, na seção 4, o método de vendas repetidas é apresentado e aplicado juntamente com a proposta de Regressão Espacialmente Ponderada. Finalmente, na última seção, os resultados e a conclusão do trabalho são apresentados, além das considerações futuras.

## 2 Revisão teórica

Um Índice de Preços ao Consumidor mede a evolução do nível de preços de uma cesta de bens e serviços adquiridos pelas famílias durante o consumo no mercado. Esse índice é uma estimativa estatística construída basea-

da nos valores de uma amostra de itens representativos cujos preços são coletados periodicamente. Subíndices e sub-subíndices são calculados para diferentes categorias e subcategorias de bens e serviços, sendo combinadas com o intuito de produzir o índice geral, cujos pesos reflitam seus impactos no total das despesas de consumo abrangidos pelo índice em questão. De maneira geral, são necessários dois tipos básicos de dados para construir um índice de preços: dados dos preços e dados para a ponderação. 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, já os dados utilizados na ponderação do impacto de diferentes tipos de despesas no total da despesa coberta pelo índice é obtido por meio de um processo de estimação estatística.

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

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 um índice de peso fixo, mas raramente pode ser considerado um verdadeiro índice de Laspeyres, 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 a ponderação que, além de pesquisas de despesas familiares, podem incluir dados comerciais e fiscais.

Um índice de preços fornece, portanto, 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 e índices 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 mais simplistas na construção de índices de preços, quais sejam: Índice de Laspeyres, Paasche, Índice ideal de Fisher, Índice de Törnqvist e Índice de Jevons. A formulação teórica e justificativa ampla para a utilização de cada um desses índices está bem relatada no texto de Diewert (2008).

## 2.1 Índices de preços

Dada a relevância do tema, serão abordados modelos relativos à formulação de índice de preços. Uma das primeiras propostas é o Índice de Laspeyres, o qual é frequentemente escrito como:

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

onde os  $P$  e  $Q$  denotam os preços e 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 cuja indexação dos bens é dado pelo índice  $m$ .

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 preços dos  $M$  produtos considerados na forma:

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

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 despe-

sa com o  $m$ -ésimo bem em relação a despesa total no período de referência, isto é, o período base  $t = 0$ , esses pesos também são denominados de *importância relativa*, já a razão  $P_{m,1}/P_{m,0}$  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, 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 mesurar a variação dos preços: 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:

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

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 na forma:

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

na qual 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  $P_{m,1}/P_{m,0}$  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 os dois considerarem a mesma cesta de produtos em ambos os períodos, e a exigência do consumo dos mesmos bens no período de referência: usualmente definido como o período contemporâneo. Já a diferença existente entre essas duas propostas está no fato da ponderação (por exemplo: 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 deste refletir a composição dos bens atualmente adquiridos, por exemplo, o Índice de Laspeyres para a variação do preço entre 2000 até 2010 precificará os bens adquiridos em 2000 e ignorará qualquer alteração na composição da cesta efetiva dos consumidores em 2010 ao longo desses 10 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 e 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), o que significa que não importa qual período seja escolhido como período base, os resultados devem se manter equivalentes.

O próximo índice considerado nesta revisão é o índice de Törnqvist, o qual apresenta similaridades com o Índice ótimo de Fisher no sentido de considerar simultaneamente ambos os pesos  $w_{m,0}$  e  $w_{m,1}$ , mas que difere substancialmente desse mesmo índice quando se considera a forma de agregação entre os componentes de Laspeyres e Paasche: enquanto no Índice ótimo de Fisher essa agregação é realizada por meio do cômputo de uma média geométrica, no Índice de Törnqvist a união é realizada utilizando-se a log variação dos preços, isso é:

$$\log I_{0,1}^T = \sum_{m=1}^M \frac{(w_{m,0} + w_{m,1})}{2} \frac{1}{\left[ \log \left( \frac{P_{m,1}}{P_{m,0}} \right) \right]} \quad (7)$$

em que novamente, os pesos dos índices de Laspeyres e Paasche são utilizados simultaneamente. O último índice revisado nessa seção é o Índice de Jevons dado por:

$$\log I_{0,1}^J = \frac{1}{M} \sum_{m=1}^M \log \frac{P_{m,1}}{P_{m,0}} \quad (8)$$

Existe certa similaridade entre o Índice de Törnqvist e o Índice de Jevons no sentido de que ambos utilizam o logaritmo natural da variação dos preços na construção dos índices, entretanto, o Índice de Jevons realiza uma média aritmética simples, substituindo os pesos  $(w_{m,0} + w_{m,1})/2$  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.

Considerando as questões desenvolvidas na revisão acima, o presente estudo propõe a construção de um método novo para a formulação de índice de preços, mediante variação do método de vendas repetidas com a regressão espacialmente ponderada.

## 2.2 Índices de preços para o mercado imobiliário

Nesta seção são abordados os modelos para a construção de índices de preços, a saber: venda repetida, regressão hedônica, dependência espacial e modelos econométricos para modelagem de dados espaciais, com enfoque em regressões com dados *cross-section*.

No mercado imobiliário, uma das primeiras propostas para a construção de índices de preços foi o estudo de Bailey, Muth e Nourse (1963). Em tal estudo, 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. Bailey, Muth e Nourse (1963) identificam como maior problema 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, à variação na qualidade dos imóveis vendidos de um período a outro fará com que o índice varie mais amplamente do que o valor de qualquer propriedade, em segundo lugar, 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 (2010) afirmam que o problema de tal 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 ser enviesado.

Como mencionado anteriormente, o índice de preços imobiliários possui natureza heterogênea. Os principais modeladores econométricos desse índice são: o modelo de vendas repetidas (MVR) e o modelo de preços hedônicos (MPH). Um exemplo de aplicação do MVR é o índice de Case/Shiller, atualmente utilizado pela Standard & Poors no mercado norte-americano.

Relevante referência acerca do MPH, o trabalho de Rosen (1974) aponta que a presença de vários atributos caracteriza um bem heterogêneo. A diferença na composição dos atributos faz com que ocorra diferenciação no preço de cada unidade ou modelo. Rosen (1974) explicou que para cada atributo do bem há um mercado específico em que o preço é estipulado por meio das forças de demanda e oferta.<sup>1</sup>

Com o objetivo de controlar a heterogeneidade frequentemente presente nos dados do mercado imobiliário, Goodman (1978) construiu uma regressão hedônica, com ajustes 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 fizeram uso dessa abordagem para o mercado imobiliário.

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

Segundo os autores, o primeiro passo na construção de índices de preços para imóveis envolve a estimação de uma função de preços hedônicos a qual segue a forma geral:

.....  
 1 São diferentes as maneiras de utilização do MPH para construir índices de preços imobiliários. Cabe citar a proposta de Triplett (2004) de uma taxionomia acerca dos diversos métodos hedônicos, que foi adaptada ao mercado imobiliário por Hill (2013). São diferentes as maneiras de utilização do MPH para construir índices de preços imobiliários. Cabe citar a proposta de Triplett (2004) de uma taxionomia acerca dos diversos métodos hedônicos, que foi adaptada ao mercado imobiliário por Hill (2013).

$$P = f(\mathbf{S}\beta, \mathbf{N}\gamma) + \epsilon \quad (9)$$

onde  $P$  é um vetor de preços observados para os imóveis considerados,  $\mathbf{S}$  é uma matriz de atributos estruturais contendo variáveis como a idade e tamanho do imóvel,  $\mathbf{N}$  é uma matriz de características da vizinhança, incluindo medidas socioeconômicas para a região de interesse, como características ambientais da região e serviços públicos disponíveis, e  $\beta$  e  $\gamma$  são vetores de parâmetros correspondentes as matrizes  $\mathbf{S}$  e  $\mathbf{N}$ , finalmente  $\epsilon$  é o termo aleatório do modelo (9).

Se Can e Megbolugbe (1997) afirmam que devido à natureza dos dados disponíveis para a construção de índices de preços 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 (9). Nesse sentido, é importante considerar a dependência espacial, pois esta afetará a validade estatística dos resultados, uma vez que a sua presença levará 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 padrões como Teste F e Teste T, para o modelo e parâmetros, respectivamente.

Assim, para considerar a frequente dependência espacial nos preços dos imóveis, Se Can e Megbolugbe (1997) propuseram adicionar à estrutura funcional (9) um termo autoregressivo espacial para o preço dos imóveis, seguindo a estrutura SAR (*Spatial Autoregressive Model*) desenvolvida por Anselin (1988). Essa proposta é capaz de considerar dois efeitos existentes para dados de habitações (Se Can, 1992), quais sejam: efeito de adjacência, o qual é uma externalidade associada à localização absoluta do imóvel e a estrutura e o efeito de vizinhança, os quais são um conjunto de características espaciais, tais como acessibilidade, previsão de serviços públicos, etc.

Uma fragilidade da proposta de Se 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 preços passados sejam fixos, assim, Bourassa, Hoesli e Sun (2006) propuseram uma abordagem para o cômputo de índices de preços em mercados imobiliários que mantém fixos os valores passados.

Os autores denominaram de *Sale Price Appraisal Ratio* (SPAR) a abordagem proposta, a qual combina informações de avaliações e 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, e, portanto, é menos propenso ao viés de seleção da amostra. Além disso, o índice construído não necessita de ser atualizado quando novos dados estão disponíveis para novos períodos de tempo, o que não é o caso para o método de vendas repetidas e de 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 preço efetivo da transação. Segundo os autores, 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. Isso simplifica o cálculo do índice, uma vez não há necessidade de utilizar uma técnica de estimação. Matematicamente, o índice é dado por:

$$I_{0,1}^{SPAR} = \left\{ \left[ \sum_{j=1}^{n_1} \left( \frac{P_{j1}}{A_{j0}} \right) \div n_1 \right] \left[ \sum_{j=1}^{n_0} \left( \frac{P_{j0}}{A_{j0}} \right) \div n_0 \right] \right\} \quad (10)$$

em que  $t = 0,1$  são dois períodos de tempo considerados e  $P_{jt}$ ,  $A_{jt}$  são os valores para o preço e avaliação do  $j$ -ésimo imóvel no período  $t$ ,  $n_t$  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 preços de mercados imobiliários que consideram 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).

Quanto aos modelos econométricos comumente utilizados para modelagem de dados espaciais, com enfoque em regressões com dados *cross-section*, pode-se dizer, de forma sucinta, que são os seguintes: SEM, SAR e SARMA e GWR.

O modelo SAR, ou autorregressivo espacial (*spatial autorregressive model*), é frequentemente utilizado para modelagem de correlação espacial. Ele incorpora um termo de lag entre os regressores da equação, seguindo, assim, a proposta dos modelos AR (autorregressivos) em séries temporais. O SAR, de forma simples, pode ser expresso da seguinte maneira:

$$y = \rho W y + \epsilon \quad (11)$$

em que  $y$  é um vetor coluna composto por  $n$  observações da amostra para  $y_i$  (variável resposta). O parâmetro autorregressivo é o coeficiente escalar  $\rho$ , que pode ser observado como o efeito médio da variável dependente relativo à vizinhança espacial na região em questão.  $\epsilon$  refere-se ao vetor coluna, que contém os resíduos da equação. São atribuídas independência e distribuição idêntica aos resíduos nesse momento, além de distribuição normal, com média zero e variância homogênea. A matriz  $W$ , constituída por diversos modelos espaciais e conhecida como matriz de vizinhança, possui diversas definições, ocasionando críticas aos modelos espaciais que a utilizam (Ywata; Albuquerque, 2011).

Já os modelos de erros espaciais (*spatial error models*), conhecidos como SEM, partem da especificação de modelos MA (médias móveis) para observações no tempo. Eles apresentam a seguinte estrutura:

$$y = X\beta + u \quad (12)$$

Os resíduos da equação 12 têm a seguinte estrutura autorregressiva:

$$u = \lambda W u + \epsilon \quad (13)$$

em que  $\epsilon$  é um vetor de resíduos com distribuição normal multivariada, com média nula e matriz de covariância  $\sigma^2 I$ . A intensidade da autocorrelação espacial entre os resíduos da equação 13 é indicada pelo coeficiente escalar  $\lambda$ . Esse parâmetro mensura com base no resíduo da região em questão o efeito médio dos erros dos vizinhos. Os modelos SEM não indicam que a variável resposta é função direta dos seus *lags* espaciais, análise

oposta à realizada pelos modelos SAR. Já a autocorrelação espacial dos modelos SEM é encontrada nos termos de erro.

Quanto ao modelo SARMA (*spatial autorregressive and moving average*), é obtido por meio da combinação dos modelos SEM e SAR em uma especificação mais geral e tem por base os modelos ARMA (*autorregressive and moving average*) para séries temporais. Eles observam o seguinte padrão:

$$y = \rho W_1 y + X\beta + u \quad (14)$$

no qual a fórmula a seguir apresenta a estrutura autorregressiva dos resíduos da equação 14.

$$u = \lambda W_2 \epsilon + \epsilon \quad (15)$$

$W_1$  e  $W_2$  são matrizes de contiguidade, não obrigatoriamente idênticas. O modelo caracteriza-se como não identificado quando  $W_1 = W_2$  e as estimativas para os coeficientes podem ser muito instáveis (Anselin, 1988; Lesage; Pace, 2009). Essa instabilidade não ocorrerá se a matriz de delineamento contiver ao menos uma variável exógena, além do intercepto. Devido à necessidade de especificação de duas matrizes de contiguidade diferentes em alguns casos, a utilização do modelo SARMA é criticada. Geralmente é realizada de forma arbitrária a escolha de uma matriz de contiguidade. Assim, o grau de arbitrariedade na escolha de duas matrizes diferentes é maior.

Até o momento, foi assumido que modelos SAR, SEM e SARMA possuíam estrutura constante na área de estudo (não havia variações locais nas estimativas dos parâmetros).

Já no caso em que ocorre heterogeneidade espacial nas estimativas dos parâmetros, é utilizado o modelo GWR (*Geographically Weighted Regression*) (Fotheringham *et al.*, 2003), que possibilita grande variação local das estimativas dos parâmetros, de forma similar ao desvio de um parâmetro (*parameter drift*) em um modelo de série temporal. Tal modelo reescreve o modelo linear de forma um pouco diferente:

$$y_i = X\beta_i + \epsilon \quad (16)$$

em que  $i$  é a localização estimada do parâmetro local. Esse modelo será tratado, de forma mais aprofundada, na seção 4.

### 3 Dados disponíveis

A primeira atividade necessária para o monitoramento do mercado imobiliário é a construção de uma base de dados capaz de integrar os dados que não estejam consolidados além da base dos imóveis construídos. Assim, identificou-se no âmbito institucional do Distrito Federal – DF uma base de dados de imóveis que poderia fornecer os dados necessários para geração das informações para o desenvolvimento de um observatório da evolução dos índices de preços em Brasília. Tem-se a base proveniente da Secretaria de Estado da Fazenda do Distrito Federal (SEFAZ-DF) com os imóveis alienados que pagam o Imposto de Transmissão de Bens Intervivos (ITBI) para a construção do índice de preços por meio do método de vendas repetidas.

Em relação à base da SEFAZ-DF, a escolha de uma fonte de dados oficiais do valor transacionado teve como motivação eliminar as variações dos valores de oferta dos bens imóveis no mercado, já que outra possível fonte de dados seria os valores de oferta de imóveis disponíveis em classificados de imóveis. É sabido que o valor pedido raramente é o transacionado, tendo em vista a margem existente em qualquer processo de negociação. Especificamente, a Secretaria de Estado de Fazenda do Distrito Federal, por ser o órgão distrital responsável pela cobrança do Imposto sobre a Transmissão Inter Vivos de Bens Imóveis por Natureza ou Acesso Física e de Direitos Reais sobre Imóveis – ITBI, é o órgão mais adequado para fornecer as informações necessárias para a estimação dos índices de preços do mercado imobiliário no Distrito Federal. Já o imposto, conforme estabelece o Art. 1º da Lei Nº 3.830, de 14 de março de 2006 e o Decreto Nº 27.576, de 28 de dezembro de 2006 (Distrito Federal, 2006a e 2006b) apresenta:

- i)* a transmissão “inter vivos”, a qualquer título, por ato oneroso, da propriedade ou domínio útil de bens imóveis por natureza ou acesso física;
- ii)* a transmissão “inter vivos”, a qualquer título, por ato oneroso, de direitos reais sobre imóveis, exceto os de garantia;
- iii)* a cessão de direitos à sua aquisição, por ato oneroso, relativo às transmissões referidas nos incisos anteriores.

A base de cálculo do imposto no Distrito Federal é definida em lei, sendo o valor devido decorrente do valor venal dos bens ou direitos transmiti-

dos ou cedidos (Art 5º, Distrito Federal, 2006). O valor venal, conforme Art 6º, é definido pela administração através de avaliação, que considerará elementos do imóvel como forma, dimensão, uso, localização, estado de conservação, além da declaração do contribuinte do imposto. Nesse caso, vale ressaltar que o imposto devido incide sobre o maior valor, seja o decorrente da avaliação ou do valor da transação declarado no instrumento de compra e venda, sendo a alíquota definida de 2%.

As informações colhidas mediante o pagamento do imposto e do preenchimento do Documento de Arrecadação – DAR são os elementos para a construção da base de imóveis do ITBI da SEFAZ-DF utilizada neste trabalho. No DAR devem constar, conforme previsto Art.11, § 3º, do Decreto Nº 27.576/2006:

- I - nome, domicílio fiscal e número de inscrição, no CPF ou no CNPJ, do adquirente e do transmitente;*
- II - natureza da transmissão;*
- III - identificação e valor do bem, sua localização, dimensões, e informação sobre a existência de edificação ou benfeitoria;*
- IV - fração ideal, área útil e área total construída, no caso de imóvel em condomínio;*
- V - preço pelo qual se realiza a transmissão;*
- VI - número de inscrição do imóvel no cadastro imobiliário do Distrito Federal.*

Para fins desta pesquisa, foi solicitada à SEFAZ-DF a exclusão da identificação do contribuinte, em razão do sigilo, restando as informações dos imóveis, os campos presentes na base são:

- 1) *Inscrição do Imóvel*: número exclusivo de cada imóvel na base da fazenda.
- 2) *Natureza do Imóvel*: especifica a categoria de uso em que se enquadra o imóvel, de um total de 119 categorias existentes.
- 3) *Cidade*: cidade onde se localiza o imóvel, de um total de 18 cadastradas.
- 4) *Área fisco*: área utilizada pela SEFAZ-DF para fins tributários.
- 5) *Área habite-se*: área original presente no habite-se da edificação.
- 6) *Área declarada*: área declarada no instrumento tributário.
- 7) *Data de Lançamento do Imposto*: data do pagamento do imposto.
- 8) *Valor da Transação*: valor de referência utilizado para o cálculo do imposto devido.
- 9) *Percentual transacionado*: percentual do imóvel transacionado.

A base da SEFAZ-DF apresenta um total de imóveis transacionados entre os anos de 2003 a 2012. O universo cadastral presente na base fornecida corresponde a 214.807 imóveis, com registro de 295.780 transações, para todos os usos presentes na base, em 23 Regiões Administra-

tivas presentes na base de dados. Considerando a série histórica de 2003 a 2012 (10 anos) tem-se 214.352 imóveis, correspondendo a 294.943 transações de alienação. Do total de alienações, observa-se que 58.940 imóveis foram transacionados mais de uma única vez, em um total de 139.531 transações.

Com relação ao percentual transacionado, ou seja, o quanto de um imóvel foi posto à venda, 210.113 imóveis foram 100% transacionados. Considerando este recorte, o universo se reduz para, 279.193 transações, sendo que 127.122 dessas transações correspondem à venda de um mesmo imóvel em períodos distintos da série.

Outros campos foram adicionados à base original. O primeiro campo refere-se às Regiões Administrativas (Campo A) que existem no Distrito Federal. No caso da base original, por exemplo, não existia a separação de endereços que, mesmo estando na mesma cidade, estão localizados em regiões administrativas distintas, como é o caso de Águas Claras, que estava classificado como Taguatinga. Tal ajuste foi realizado, sendo possível identificar 23 regiões administrativas entre as 18 cidades originalmente listadas.

Além do mais, a variável Área Útil dos imóveis cadastrados (Campo B) foi também adicionada e consiste na área de real usufruto do imóvel, descontadas áreas como circulação, garagem, fração ideal da área comum etc. Foi aplicado um algoritmo para identificar, entre os campos de áreas existentes (área fisco, área habite-se e área declarada), o que apresentasse maior área. A regra para escolha da área foi aplicada para cada imóvel presente na base. A necessidade desse campo é decorrente de a área útil ser utilizada como referência na definição do valor do metro quadrado de venda – indicador empregado no mercado imobiliário. Vale ressaltar que a obtenção direta é válida apenas para determinados tipos de imóveis, como os de habitação unifamiliar.

O terceiro campo adicionado consiste no Valor do Metro Quadrado (Campo C), necessário para relativizar o impacto do porte para fins de comparação entre os imóveis constantes na base. Este campo é obtido pela relação direta entre o Valor de Transação e a Área Útil do Imóvel. Com base na estrutura apresentada, procedeu-se com o método de vendas repetidas de Bailey, Muth e Nourse (1963) estendido por Case e Shiller (1987), abordagem descrita na seção seguinte.

## 4 Método de vendas repetidas

Case e Shiller (1987) estendendo a proposta de Bailey, Muth e Nourse (1963), sugerem um método de vendas repetidas, o qual é uma extensão do método de regressão hedônica para construção de índices de preços. 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 binária (*dummy*) que indica se a transação ocorreu em cada um dos períodos considerados.

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 o fato do efeito das características dos bens se manterem constantes ao longo do tempo por meio do efeito fixo gerado pela inserção das variáveis dicotômicas para cada período considerado, enquanto os outros parâmetros do modelo são permitidos variar. Na abordagem de regressão hedônica, um modelo de regressão na forma seguinte é estimado:

$$y_{it} = \alpha + \beta_i' x_i + \delta_2 D_{2,it} + \dots + \delta_T D_{T,it} + u_{it} \quad (17)$$

em que  $y_{it}$  é o logaritmo natural do preço para a  $i$ -ésima habitação no tempo  $t$ ,  $x_i$  é um vetor de características estruturais e espaciais associadas à habitação, as variáveis  $D_{j,it}$  com  $j = 2, \dots, T$  são os períodos considerados para a estimação do índice de preços e  $u_{it}$  é o termo de erro aleatório do modelo de regressão.

Entretanto, um problema frequente na estimação dos modelos de regressão hedônicas é a existência de variáveis omitidas relevantes ao processo gerador dos dados para o preço da habitação. Essas variáveis omitidas, como por exemplo variáveis estruturais e de localização, bem como a qualidade da vizinhança, podem induzir a um elevado viés na estimação dos parâmetros para o índice de preços, em especial se essas variáveis estiverem correlacionadas com as *dummies* temporais.

Por exemplo, suponha que a variável omitida seja uma medida do padrão de acabamento da casa. Se as casas vendidas em datas posteriores tendem a ser de melhor padrão de acabamento que as anteriormente ven-

didadas, o parâmetro  $\delta$  de períodos posteriores apresentará um viés positivo, sobrevalorizando o preço de uma habitação padrão. Para tratar essa possibilidade, Case e Shiller (1987) propuseram uma abordagem para tratar a omissão de variáveis relevantes por meio do acompanhamento das transações de habitações ao longo do tempo.

Para um subconjunto de habitações que foram transacionadas pelo menos duas vezes, pode-se calcular a diferença entre os logaritmos naturais dos preços de venda para dois períodos subsequentes de tempo  $s < t$ , cuja representação matemática é dada por:

$$y_{it} - y_{is} = \delta_2 (D_{2,it} - D_{2,is}) + \dots \delta_T (D_{T,it} - D_{T,is}) + (u_{it} - u_{is}) \tag{18}$$

O vetor de características da habitação,  $x_i$ , não aparece nessa equação porque se assume que as características e os parâmetros  $\beta'_i$  não mudam ao longo do tempo, caso essas suposições estejam corretas, o estimador de vendas repetidas fornece estimativas não-viesadas sobre o índice de preços sem a necessidade de dados sobre todas as características da habitação. Assim, uma variável omitida, como o padrão de acabamento da habitação, não influenciará as estimativas, a menos que ela se modifique ao longo do tempo ou o seu coeficiente. Mas caso o coeficiente ou as variáveis estruturais e espaciais mudem ao longo do tempo, o método de vendas repetidas pode ser adaptado a essas suposições, nesse caso, adiciona-se uma nova variável  $z$  que muda ao longo do tempo, na forma:

$$y_{it} = \alpha + \beta'_i x_i + \delta_2 D_{2,it} + \dots \delta_T D_{T,it} + z_{it} + u_{it} \tag{19}$$

Note que a fonte da variação é irrelevante, pois o modelo (19) pode ser reescrito como o produto da variável *dummy* temporal e um coeficiente de variação no tempo, na forma:

$$y_{it} = \alpha + \beta'_i x_i + \delta_2 D_{2,it} + \dots \delta_T D_{T,it} + [z_i + \lambda_2 z_i D_{2,it} + \dots \lambda_T z_i D_{T,it}] + u_{it} \tag{20}$$

O qual apresenta a seguinte estrutura para o modelo de vendas repetidas:

$$y_{it} - y_{is} = \delta_2 (D_{2,it} - D_{2,is}) + \dots \delta_T (D_{T,it} - D_{T,is}) + [z_i + \lambda_2 z_i (D_{2,it} - D_{2,is}) + \dots \lambda_T z_i (D_{T,it} - D_{T,is})] + (u_{it} - u_{is}) \tag{21}$$

Captando assim a variação no tempo das variáveis estruturais ou dos coeficientes. Após a estimação dos parâmetros  $\delta$ , o índice de preços para o período  $t$  é dado por  $\exp \hat{\delta}_t$ , no caso de período base, tem-se  $\exp 0 = 1$ .

Case e Shiller (1987) alertam para a possibilidade de heteroscedasticidade nos dados associados a preços de habitação. Para tratar essa heterogeneidade na variância das observações, os autores sugerem um método em três etapas denominado de método de vendas repetidas ponderado. Especificamente, enquanto Bailey, Muth e Nourse (1963) consideram erros independentes e identicamente distribuídos, Case e Shiller (1987) pressupõem que os erros assumem uma estrutura heteroscedástica.

Os passos considerados por Case e Shiller (1987) são estruturados em três estágios. Na primeira etapa, a diferença entre os logaritmos dos preços do segundo registro de venda e do primeiro registro de venda é regredido contra um conjunto de variáveis binárias, uma para cada período de tempo na amostra (exceto para o primeiro período considerado). Geram-se variáveis *dummies*, as quais assumem zero em cada trimestre, exceto para os trimestres em que as duas vendas ocorreram. Para a primeira venda, a variável *dummy* assume valor igual  $a - 1$ , já para a segunda venda a variável *dummy* assume valor igual  $a + 1$ , esse modelo é representado pela equação (18).

Após obtidas as estimativas dos parâmetros, o vetor de resíduos é calculado e então, na segunda fase, uma regressão ponderada dos quadrados desses resíduos é realizada considerando um termo constante e o tempo entre as vendas com variáveis independentes. Nesse modelo, segundo Case e Shiller (1987), o intercepto é uma estimativa para  $2\sigma^2$ , isto é, duas vezes a variância do erro estocástico associado à habitação, já o coeficiente angular é uma estimativa da variância da variação trimestral.

Na terceira e última fase, uma regressão generalizada de mínimos quadrados é executada usando os mesmos insumos da primeira fase, exceto que nessa etapa cada observação é dividida pela raiz quadrada do valor ajustado da segunda etapa com o objetivo de corrigir possível heteroscedasticidade.

A abordagem de Case e Shiller (1987) apresentam algumas limitações, quais sejam: é necessário uma base de dados “suficientemente” grande para que possa-se considerar a representatividade do conjunto de habitações vendidas mais de uma vez; período de tempo considerado também “suficientemente” grande de maneira a considerar no cálculo do índice tanto unidades que são vendidas mais rapidamente quanto as que demo-

ram mais para serem vendidas; e constância no tempo da qualidade das habitações consideradas. Sem esses pressupostos, o índice construído com base na proposta dos autores pode apresentar viés e, conseqüentemente, prejudicar a qualidade do índice de preços proposto.

**Tabela 2 Número de vendas repetidas, por RA, tempo médio de revenda, porcentagem e valor médio da transação**

<b>RA</b>	<b>Frequência</b>	<b>Porcentagem</b>	<b>Valor médio da transação</b>
Águas Claras	10978	15.86%	R\$ 226,353.00
Brasília	14040	20.28%	R\$ 278,684.00
Brazilândia	133	0.19%	R\$ 55,483.00
Candangolândia	302	0.44%	R\$ 108,992.00
Ceilândia	5393	7.79%	R\$ 67,915.00
Cruzeiro	7373	10.65%	R\$ 243,390.00
Gama	2030	2.93%	R\$ 110,540.00
Guará	5237	7.57%	R\$ 219,235.00
Lago Norte	2139	3.09%	R\$ 256,090.00
Lago Sul	746	1.08%	R\$ 810,119.00
Núcleo Bandeirante	948	1.37%	R\$ 125,425.00
Paranoá	0	0.00%	R\$ 542,205.00
Park Way	866	1.25%	R\$ 347,173.00
Planaltina	734	1.06%	R\$ 60,911.00
Recanto das Emas	793	1.15%	R\$ 55,501.00
Riacho Fundo	1288	1.86%	R\$ 95,608.00
Riacho Fundo II	0	0.00%	R\$ 15,000.00
Samambaia	3524	5.09%	R\$ 117,124.00
Santa Maria	1373	1.98%	R\$ 84,909.00
São Sebastião	202	0.29%	R\$ 226,913.00
Sia	260	0.38%	R\$ 473,955.00
Sobradinho	2777	4.01%	R\$ 149,918.00
Taguatinga	8020	11.59%	R\$ 132,445.00

*Fonte: Elaboração dos autores.*

Utilizando os dados da SEFAZ-DF para os anos de 2003 a 2012 com mensuração mensal, construiu-se o conjunto de dados para todos os pares de vendas repetidas registrados para o imóvel segundo seu código de inscri-

ção. É importante ressaltar que a base de dados utilizada cumpre os requisitos expostos por Case e Shiller (1987), uma vez que é composta de 55.456 observações e 120 períodos temporais, além de se admitir constância na qualidade das habitações, pois, como a cidade foi reconhecida como patrimônio cultural da humanidade pela UNESCO em 1987 e tombada como patrimônio histórico federal em 1990, qualquer nova construção deve atender aos padrões estabelecidos, mantendo-se todos os imóveis aproximadamente homogêneos quanto aos quesitos estruturais.

A tabela 2 apresenta uma estatística descritiva da base de dados. Ela mostra o número total de vendas repetidas, por RA, o tempo médio de revenda, a porcentagem de cada RA e, por fim, o valor médio da transação por RA, com fundamento em vendas repetidas. Cada frequência refere-se à quantidade de imóveis vendidos mais de uma vez. A partir da quantidade de transações dos imóveis, faz-se uso ou não do método de vendas repetidas.

**Tabela 3 Preço total da base e estatísticas relevantes**

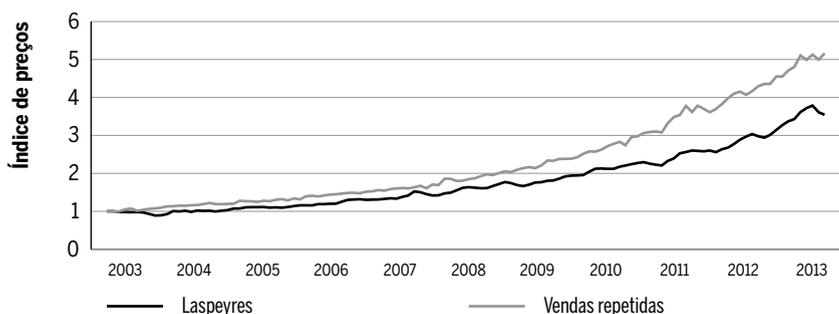
	<b>Valor da transação (Base completa)</b>	<b>Valor da transação 1 (Base vendas repetidas)</b>	<b>Valor da transação 2 (Base vendas repetidas)</b>
<b>Mínimo</b>	R\$ 1,926.10	R\$ 2,712.95	R\$ 2,712.95
<b>Máximo</b>	R\$ 287,000,000.00	R\$ 26,850,850.00	R\$ 47,176,624.01
<b>Intervalo</b>	R\$ 286,998,073.90	R\$ 26,848,137.05	R\$ 47,173,911.06
<b>Soma</b>	57361714361.23	9647060702	16109922666
<b>Mediana</b>	R\$ 111,104.74	R\$ 82,467.00	R\$ 138,000.00
<b>Média</b>	R\$ 205,105.73	R\$ 139,341.94	R\$ 232,691.39
<b>SE. médio</b>	1796.55	1495.324	2430.409
<b>CI.médio.0.95</b>	3521.19	2930.832	4763.597
<b>Variância</b>	902656843659.44	1.54805E+11	4.08952E+11
<b>Desvio padrão</b>	950,082.55	393,452.11	639,493.22
<b>Coefficiente de variação</b>	4.63	2.824	2.748

*Fonte: Elaboração dos autores.*

A tabela 3 mostra o preço total da base, completa e de vendas repetidas (no período  $t - 1$  de venda do imóvel e no período  $t$  de venda do imóvel), além de estatísticas relevantes na análise como: soma, mediana, média, desvio padrão e coeficiente de variação.

A partir dos dados, foi obtida a Figura 1, que apresenta a série da evolução do índice de preços por meio da metodologia de vendas repetidas e a de Laspeyres utilizando mediana:

Figura 1 **Evolução do índice de preços da habitação para o Distrito Federal**



Conforme apresentado na Figura 1, há um deslocamento entre as estimativas para o índice de preços utilizando medida de tendência central (Laspeyres) e o método de vendas repetidas de Case e Shiller (1987). Essa divergência poderia ser explicada, entre outros fatores, pela existência de variáveis omitidas não modeladas por meio do método de Laspeyres e, portanto, subestimando a evolução dos preços dos imóveis no Distrito Federal.

É importante ressaltar que o método de Laspeyres é a abordagem mais comumente utilizada para a construção de índices de preços conforme destacado por Diewert (2008), consequentemente, caso as variáveis omitidas relevantes para a explicação da evolução dos preços dos imóveis não sejam consideradas, um viés é observado nas estimativas dos índices de preços, afetando assim a formulação e avaliação da efetividade das políticas de planejamento urbano no território, pois poder-se-ia, erroneamente, considerar um cenário de menor evolução dos preços do que o que foi realmente observado.

Fica evidente, portanto, pela Figura 1, que a evolução dos índices de preços para os imóveis no Distrito Federal foi maior do que o medido pela abordagem de Laspeyres, essa situação afeta diretamente a população de menor renda, pois, no período, os rendimentos observados cresceram a uma taxa menor do que a taxa observada para os preços dos imóveis, de fato, a inflação segundo Índice de Preços ao Consumidor Amplo (IPCA) registra-

da entre janeiro de 2003 a dezembro de 2012 foi de 76,61%. Essa situação evidencia a dificuldade de suprir o déficit habitacional no Distrito Federal que, segundo Furtado, Lima Neto e Krause (2013), é de 116.601 domicílios.

O período em análise também é reconhecido como sendo de alta elevação dos preços no país, principalmente a partir de 2008, quando houve maior acesso ao crédito. Os dados utilizados foram fornecidos pela Caixa Econômica Federal, empresa responsável por quase 90% dos financiamentos imobiliários. Com a maior oferta de crédito, a demanda por apartamentos aumentou bastante, e, em consequência, o preço de venda dos referidos imóveis.

No entanto, há uma questão que surge ao observar esses resultados: a evolução dos preços dos imóveis ocorreu de maneira semelhante em todo o Distrito Federal? Ou há regiões onde essa evolução foi maior? Essa questão é crucial para o desenvolvimento de políticas públicas pró-habitação, em especial se é possível observar que, para as regiões com menor renda, houve uma maior evolução dos valores dos imóveis do que para as demais regiões do Distrito Federal. Assim, utilizando as recomendações de Se Can e Megbolugbe (1997) e Dorsey *et al.* (2010) procedeu-se com a abordagem de vendas repetidas em um contexto de regressão espacial, isto é, modelos de regressão que incorporam a dependência espacial em suas estimativas.

Devido ao esforço computacional envolvido na realização da estimação dos parâmetros de modelos espaciais, decidiu-se trabalhar como uma amostra probabilística representativa da população observada para os dados da SEFAZ-DF para os anos de 2003 a 2012. Usando a abordagem de Cohen (2013), calculou-se ser necessário uma amostra 4080 observações para alcançar um poder no teste de pelo menos 99% na detecção de um coeficiente de determinação de pelo menos 0,05.

#### 4.1 Modelos de vendas repetidas espaciais

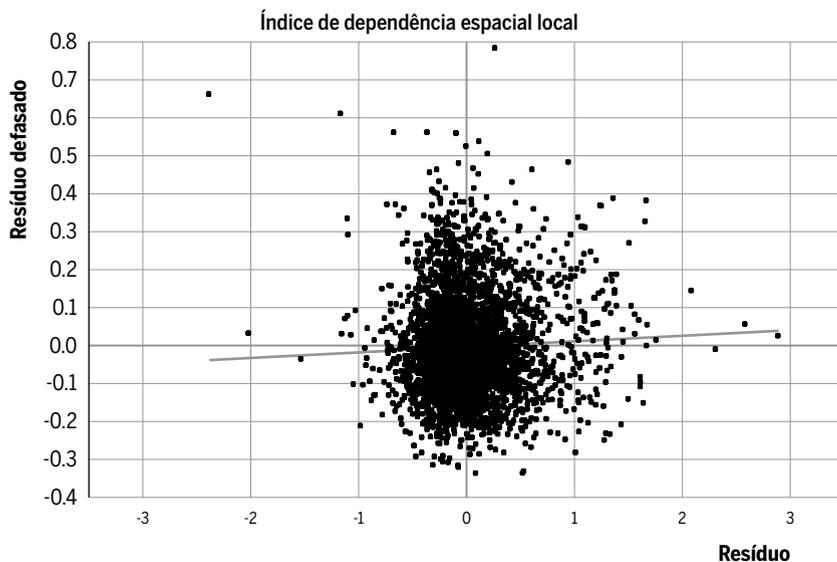
Há então duas possíveis propostas de modelos de vendas repetidas espaciais: coeficientes de autocorreção fixos no espaço ou coeficientes variáveis no espaço. O primeiro modelo é mais adequado quando a variável dependente apresenta certa estacionariedade espacial, enquanto o segundo modelo, de coeficientes variáveis no espaço, é mais adequado quando da existência de não-estacionariedade espacial, isto é, pode haver no espaço

para um mesmo período de tempo, regiões cujos índices de preço variaram entre si de maneira extrema.

Avaliou-se, portanto, o grau de dependência espacial dos resíduos obtidos por meio da regressão (5), os quais sobre os pressupostos de Case e Shiller (1987) deveriam ser independentes e identicamente distribuídos. Seguindo Anselin (1988), calculou-se o Índice de Moran para os resíduos por meio de matriz de contiguidade espacial relativa aos 10 vizinhos mais próximos, cujo valor obtido foi igual a 0,0085 com nível crítico de 9%. Há evidências da independência espacial, e portanto inadequabilidade dos modelos de coeficientes de autocorreção fixos no espaço, em outras palavras, observou-se baixa dependência global aliada a um alto p-valor, sugerindo assim serem inadequados modelos de dependência espacial globais, tais como SAR, SEM e SARMA (Ywata; Albuquerque, 2011).

Para avaliar a estacionariedade dos dados (pressuposto de identidade distribucional), a dependência espacial local estimada por meio do LISA (*Local Indicators of Spatial Association*) foi calculada:

Figura 2 **Evolução do índice de preços da habitação por região utilizando box-plot e GWR**



Pode-se observar na Figura 2 pontos extremos em todos os quatro quadrantes, sugerindo assim a possibilidade de uma heterogeneidade espacial

quanto à direção da dependência espacial local, implicando em uma possível não-estacionariedade espacial, uma vez que os pontos se desviam mutuamente de um padrão linear estável representado pela reta em azul, cujo coeficiente angular é equivalente ao Índice de Moran calculado.

Nesse sentido, a abordagem de Regressão Espacialmente Ponderada (*Geographically weighted regression* – GWR) para a estimação dos índices de preços no Distrito Federal é adequada por ser indicada, principalmente, quando há evidências de heterogeneidade espacial, como visualizado na Figura 2 (Cahill; Mulligan, 2007). A vantagem dessa abordagem é que a heterogeneidade presente nas vendas registradas para o mesmo imóvel na região considerada é levada em conta, pois, para cada observação pontual no espaço, o método GWR fornece uma estimativa pontual para o índice de preços naquele ponto, ponderado pelos registros dos imóveis vizinhos. O resultado é um “contínuo” no espaço para os índices de preço e, portanto, é possível avaliar a evolução da mudança do preço no espaço e no tempo simultaneamente. A presença de submercados com dinâmica imobiliária própria dentro de grandes mercados – que são as Regiões Metropolitanas – é reconhecida pelos profissionais da área de imóveis. A possibilidade de visualizar as diferenças espaciais da evolução dos preços ajuda a caracterizar esses submercados e consequentemente propor políticas públicas específicas para essas regiões.

O método GWR proposto por Fotheringham, Charlton e Brunsdon (1998) pode ser utilizado para examinar a variabilidade espacial dos resultados da regressão ao longo de uma região e assim informar sobre a presença de padrões não-estacionários espaciais. A presença de uma não-estacionariedade espacial induziria a índices de preços diferentes dependendo da região avaliada. No caso da estacionariedade espacial, os resultados para os parâmetros estimados de cada região convergiriam para o valor global, obtido por meio do método usual de vendas repetidas. Ao invés de aceitar um conjunto de resultados globais para a regressão (vendas repetidas), a regressão espacialmente ponderada permite produzir resultados da regressão localmente para todos os pontos amostrados na região de interesse, fornecendo, assim, estimativas locais para o índice de preços do mercado imobiliário.

Especificamente, considere o seguinte modelo de regressão local:

$$y_i = \beta_0(i) + \beta_1(i)X_{1i} + \beta_2(i)X_{2i} + \dots + \beta_p(i)X_{pi} + u_i \quad (22)$$

em que  $i = 1, \dots, n$  são as observações amostradas no espaço e  $u_i$  é o termo estocástico associado àquela região. Na forma matricial, tem-se:

$$y_i = \beta(i)' \mathbf{X}_i + u_i \tag{23}$$

Com  $\beta(i) = (\beta_0(i), \beta_1(i), \beta_2(i), \dots, \beta_p(i))$  e  $\mathbf{X}_i = (1, X_{1i}, X_{2i}, \dots, X_{pi})'$ , o estimador para o vetor de parâmetros locais, ponderado pela distância entre as observações, é dado por (Mennis, 2008, p. 80):

$$\beta(i) = (\mathbf{X}'\mathbf{W}(i)\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}(i)\mathbf{y} \tag{24}$$

tal que  $\mathbf{X}$  é a matriz de dimensão  $n \times (p + 1)$  de covariáveis observadas, as quais neste texto são as *dummies* estipuladas no método de vendas repetidas sem a presença de intercepto,  $\mathbf{W}(i)$  é a matriz de pesos calculada para cada observação amostrada e  $\mathbf{y}$  é o vetor de dimensão  $n \times 1$  com a variável resposta para todas as observações. A matriz  $\mathbf{W}(i)$  é diagonal na forma:

$$\mathbf{W}(i) = \begin{pmatrix} w_{i1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_{in} \end{pmatrix} \tag{25}$$

em que  $w_{ij} = \exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right\}$  é o peso entre as observações  $i$  e  $j$ ,  $d_{ij}$  é a

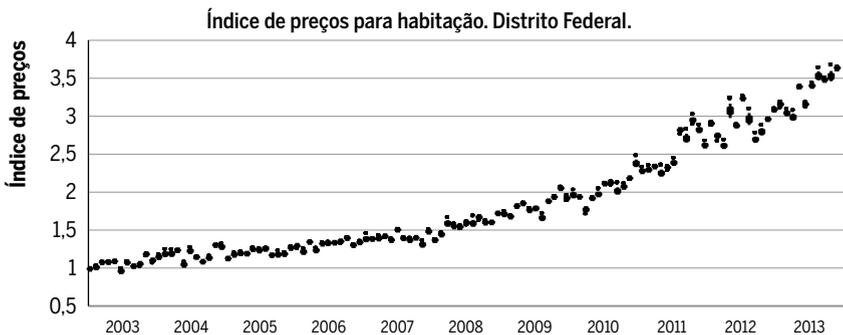
distância entre as observações  $i$  e  $j$ , e  $h$  é a janela (*bandwidth*) de ajuste. Há aqui um *trade-off* para a escolha de  $h$  (*trade-off* viés-variância), pois quanto maior for  $h$  mais pontos são considerados, aumentando assim o viés das estimativas dos parâmetros, caso  $h$  seja pequeno demais, menos pontos são considerados e, portanto, a variância dos parâmetros é aumentada e menos incerteza temos acerca do parâmetro populacional.

Ao calcular as estimativas dos índices de preço para o Distrito Federal entre 2003 e 2012 usando o método de Regressão Espacialmente, obtém-se para cada localidade amostrada uma estimativa de seu índice de preços, o qual é ponderado em função de seus vizinhos espaciais com o objetivo de modelar adequadamente a heterogeneidade espacial local observada por meio do gráfico do LISA (Figura 2).

Ademais, Fotheringham, Charlton e Brunsdon (1998) sugerem que a escolha da janela  $h$  deva ser realizada por meio de um processo iterativo, obje-

ativando minimizar o Critério de Informação de Akaike (AIC) (Akaike, 1998), dessa forma, as estimativas locais foram calculadas usando-se o menor AIC observado por meio de busca em grelha (Cortez, 2014). Gráficamente, ao invés de uma única série temporal como visualizado na Figura 1, obtêm-se diversos possíveis índices para cada período de tempo. Uma maneira de visualizar o nível de heterogeneidade espacial ao longo do tempo é utilizando o box-plot para cada período de tempo, como apresentado a seguir:

Figura 3 Evolução do índice de preços da habitação por região utilizando box-plot e GWR



Observa-se por meio da Figura 3 que à medida que o tempo evolui, maior é a variação dos índices de preços locais, além do mais, o padrão observado é muito similar ao padrão da série temporal construída com base nos dados populacionais para o método de vendas repetidas de Case e Shiller (1987) apresentado na Figura 1.

Essa evolução temporal da heterogeneidade espacial observada para os índices de preços no Distrito Federal pode ser em função das diferentes políticas públicas implementadas no Distrito Federal após 2011, gerando índices de preços muito distintos no espaço para um mesmo período. A evolução das benfeitorias em algumas regiões, mantendo outros espaços absolutamente sem alterações na qualidade de vida, poderia justificar esse avanço na desigualdade espacial dos índices de preços após o ano de 2011.

## 5 Conclusão

O desenvolvimento de um índice de preços para o mercado imobiliário

é tarefa desafiadora, uma vez que os imóveis têm características muito peculiares, tais como: indivisibilidade, heterogeneidade, e durabilidade. Especificamente, quanto à heterogeneidade, os resultados apresentados na seção 4 confirmam a existência de uma heterogeneidade espacial local, com aumento ao longo do tempo, em especial nos períodos após 2011.

Essa heterogeneidade foi modelada por meio do método de Regressão Espacialmente Ponderada, o qual fornece um índice de preços para cada período e localidade amostrada. Também foi utilizado o método de vendas repetidas Diewert (2008), por não exigir que a base de dados contenha informações sobre as diversas características do imóvel. Assim, são diminuídos os custos de construção da base de dados e o tratamento econométrico torna-se menos complexo.

Ademais, a construção do observatório, por meio da proposta aqui desenvolvida, supre uma lacuna quanto à construção de índices de preços, de modo a ser possível a partir de agora acompanhar a evolução dos preços de imóveis no Distrito Federal, pois os dados estão disponíveis e presentes na SEFAZ-DF e a metodologia proposta trata a existência de variáveis omitidas relevantes associadas à estrutura do imóvel e ao seu espaço geográfico.

O acompanhamento contínuo propiciado por meio dessa proposta constitui-se em instrumento eficaz da política urbana no âmbito da capital federal, apoiando os gestores na formulação e avaliação de políticas públicas para o sistema de habitação do Distrito Federal juntamente com a integração de instrumentos urbanísticos, tais como a outorga onerosa. Sugere-se em estudos futuros que os pesquisadores interessados estimem os índices de preços via método de vendas repetidas, utilizando para isso Regressão Quantílica como McMillen & Thorsnes (2006) e Salvati *et al.* (2012) para se corrigir quaisquer desvios devido à existência de valores extremos que poderiam afetar a equação média de regressão, por meio da estimativa da mediana.

## Referências

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. *Selected Papers of Hirotugu Akaike*, p. 199-213. New York: Springer, 1998.
- AIZCORBE, A. M. *A practical guide to price index and hedonic techniques*. Oxford University Press, 2014.

- ANSELIN, L. *Spatial econometrics: methods and models*. v. 4. Springer Science & Business Media. Santa Barbara: Springer, 1988.
- BAILEY, M. J.; MUTH, R. F.; NOURSE, H. O. A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58 (304), p. 933-942, 1963.
- BOURASSA, S. C.; HOESLI, M.; SUN, J. A simple alternative house price index method. *Journal of Housing Economics*, 15(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.
- CAHILL, M.; MULLIGAN, G. Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), p. 174-193, 2007.
- 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.
- SE CAN, A. Specification and estimation of hedonic housing price models. *Regional science and urban economics*, 22(3), p. 453-474, 1992.
- SE CAN, A.; MEGBOLUGBE, I. Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics*, 14 (1-2), p. 203-222, 1997.
- COHEN, J. *Statistical power analysis for the behavioral sciences*. Academic press. New York: Lawrence Erlbaum Associates, 2013.
- CORTEZ, P. *Modern Optimization with R*. New York: Springer, 2014.
- DIEWERT, W. E. Commentary. *Federal Reserve Bank of St. Louis Review*, May/June 1997.
- DIEWERT, W. E. Index Numbers. In: DURLAUF, S. N.; BLUME, L. E. (Eds.). *The new palgrave Dictionary of Economics*. 2nd Ed. Basingstoke, UK: Palgrave Macmillan, 2008.
- DORSEY, R. E.; HU, H.; MAYER, W. J.; WANG, H. Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle. *Journal of Housing Economics*, v. 19, Issue 2, June, p. 75-93, 2010.
- FOTHERINGHAM, A. S.; CHARLTON, M. E.; BRUNSDON, C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11), p. 1905-1927, 1998.
- FOTHERINGHAM, A. S.; BRUNSDON, C.; CHARLTON, M. Geographically weighted regression: the analysis of spatially varying relationships. *Geographical Analysis*. 35, p. 272-275. 2003.
- GOODMAN, A. C. Hedonic prices, price indices and housing markets. *Journal of Urban Economics*, 5(4), p. 471-484, 1978.
- HILL, R. Constructing Price Indexes across Space and Time: the case of the European Union. *American Economic Review*, v.04, n. 5, p. 1379-1410, 2004.
- KAPP, S. Uma tipologia de espaços cotidianos. *Risco: Revista de Pesquisa em Arquitetura e Urbanismo (Online)*, 15, p. 5-20, 2012.
- LESAGE, J.; PACE, R. K. *Introduction to spatial econometrics*. Boca Raton: CRC Press, 2009.
- MCMILLEN, D. P.; THORSNES, P. Housing Renovations and the Quantile Repeat-Sales Price Index. *Real Estate Economics*, 34(4), p. 567-584, 2006.

- MENNIS, J. Local Models for Spatial Analysis-by Christopher D. Lloyd. *Transactions in GIS*, 12 (5), p. 647-649, 2008.
- NADALIN, V.; FURTADO, B. A. Índices de preços para imóveis: uma revisão. *Boletim regional, urbano e ambiental, Instituto de Pesquisa Econômica Aplicada*, 06, Setembro, 2010.
- PAKES, A. A Reconsideration of Hedonic Price Indices with an Application to PC's. *The American Economic Review*, 93(5), p. 1578-1596, 2003.
- ROSEN, S. Hedonic price and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, v. 82, n. 1, p. 35-55, 1974.
- SALVATI, N.; TZAVIDIS, N.; PRATESI, M.; CHAMBERS, R. Small area estimation via M-quantile geographically weighted regression. *Test*, 21 (1), p. 1-28, 2012.
- SADOULET, E.; de JANVRY, A. *Quantitative development policy analysis*. Baltimore: Johns Hopkins University Press, 1995.
- TRIPLETT, J. Handbook on Hedonic and Quality Adjustments in Price Indexes: special application to information technology products. *Working Paper 2004/9*. Paris: Organization for Economic Co-operation and Development (OECD), 2004.
- YWATA, A. X. C.; de ALBUQUERQUE, P. H. Métodos e modelos em econometria espacial. Uma revisão. *Rev. Bras. Biom*, 29(2), 273-306, 2011.

### Sobre os autores

*Pedro Henrique Melo Albuquerque - pedroa@umb.br*

Universidade de Brasília, Brasília, Distrito Federal.

*Vanessa Gapriotti Nadalin - vanessa.nadalin@ipea.gov.br*

Instituto de Pesquisa Econômica Aplicada, Brasília, Distrito Federal.

*Vicente Correia Lima Neto - vicente.neto@ipea.gov.br*

Instituto de Pesquisa Econômica Aplicada, Brasília, Distrito Federal.

*Mariana Rosa Montenegro - mrosamontenegro@gmail.com*

Universidade de Brasília, Brasília, Distrito Federal.

### Sobre o artigo

Recebido em 19 de abril de 2016. Aprovado em 30 de setembro de 2016.