

**Área de Interesse:** Área 1 - Economia Pernambucana

**Índice de Preço Hedônico de Imóveis com Variáveis Estruturais: Um Análise Empírica para a Cidade do Recife (2000 -2012)**

**Álvaro Furtado Coelho Júnior**

Doutorado em Economia pelo PIMES/UFPE

Professor do DECON/UFRPE

Graduado em Economia pela Universidade Federal de Campina Grande (UFCG) (2009),

Mestre em Economia pela Universidade Federal da Paraíba (UFPB) (2011),

Doutorado em Economia pela Universidade Federal de Pernambuco (UFPE/PIMES), Brasil (2015), com período de doutorado sanduíche na *University of Illinois at Urbana-Champaign* (UIUC)/ *Regional Economics Applications Laboratory* (REAL), Estados Unidos (2014-2015). A tese de doutorado concentrou-se na área de Economia Regional e Urbana.

Endereço Postal: Universidade Federal Rural de Pernambuco, Departamento de Ciências

Econômicas. Dois Irmãos

52171900 - Recife, PE - Brasil - Caixa-postal: 50740520

alvarojuniorbr@gmail.com

(81)99912-0270

**Raul da Mota Silveira Neto**

Pós Doutorado em Economia pela UIUC

Professor do PIMES/UFPE

Possui graduação em Economia pela Universidade Federal de Pernambuco (1989), mestrado em Economia pela Universidade de São Paulo (1995) e doutorado em Economia pela Universidade de São Paulo (2000). Atualmente é professor associado do Departamento de Economia da Universidade Federal de Pernambuco. Os trabalhos e pesquisas concentram-se nas áreas de Economia Regional, Economia Urbana, Economia do Trabalho e Desenvolvimento Econômico, com interesse específico na avaliação de políticas públicas, nas políticas de desenvolvimento regional ou local, nos determinantes da desigualdade e da migração, nos condicionantes da distribuição das famílias e atividades nas cidades, nos diferenciais de qualidade de vida entre cidades e nos determinantes do crescimento das cidades.

Endereço Postal: Universidade Federal de Pernambuco, Centro de Ciências Sociais Aplicadas, Departamento de Ciências Econômicas.

Br 101 - Cidade Universitária

Cidade Universitária

50000000 - Recife, PE - Brasil

netoraul2015@gmail.com

## Área 1 - Economia Pernambucana

### **Índice de Preço Hedônico de Imóveis com Variáveis Estruturais: Um Análise Empírica para a Cidade do Recife (2000 -2012)**

#### **RESUMO:**

As cidades brasileiras verificaram, ao longo dos anos 2000, um considerável aumento de preço dos imóveis. Motivado por essa conjuntura e a partir do caso específico do Recife, este artigo objetiva analisar a dinâmica de preço dos apartamentos na referida cidade entre 2000 e 2012. Para tal o trabalho constrói um índice de preços para o valor dos apartamentos a partir do pareamento nas características de variáveis estruturais para o nível de cidade (no caso a Cidade do Recife). Essa construção é aplicada à subdivisão geográfica da cidade (as regiões políticas administrativas (RPAs)) algo, na verdade, inexistente. São apresentadas evidências empíricas a respeito de como a consideração das características estruturais influenciam na trajetória de variação de preço dos apartamentos na Cidade do Recife. Para tal, são utilizados dados do ITBI no período de 2000 a 2012. Os resultados obtidos, quanto ao índice de preços, indicam que a dinâmica de preço dos imóveis varia de acordo com a RPA analisada justamente devido ao comportamento dos estoques de variáveis estruturais disponíveis em cada RPA, indicando, assim, que inferências sobre o mercado de imóveis com base em preços médios dos imóveis (ou medidas de tendência central) podem não refletir a realidade analisada.

**PALAVRAS-CHAVE:** Índice de Preços Hedônicos, Abordagens de Emparelhamento, Características Estruturais.

#### **ABSTRACT:**

Brazilian cities have presented, over the 2000s, a considerable increase in real estate prices. Motivated by this situation and from the specific case of Recife, this paper aims to analyze the dynamics of prices of apartments in that city between 2000 and 2012. For this work we build a price index for the value of apartments from matching of characteristics of structural variables for the city level (in this case the Recife City). This construction is applied to the geographical subdivision of the city (administrative policy areas (RPAs)) something actually nonexistent. We present the empirical evidences about how the consideration of structural features influence the trajectory of price change of the apartments in Recife City. For this purpose, we used the ITBI data from 2000 to 2012. The results obtained as to the price index, show that the dynamic of price of properties varies according to the analyzed RPA, indicating that inferences about the real estate market based on average prices of dwellings (or central tendency statistics) may not reflect the analyzed reality.

**KEY-WORDS:** Hedonic Price Index, Matching Approaches, Structural Characteristics.

**Classificação JEL:** R31, R32.

## 1. Introdução

Segundo dados do Sistema Brasileiro de Poupança e Empréstimo (SBPE) do Banco Central do Brasil (BACEN, 2015) no período de 2002 a 2012 o crescimento no número de unidades habitacionais adquiridas por financiamento a partir de recursos do SBPE foi de 1532%. Em termos nominais, o valor negociado, tendo como referência o ano de 2002, expandiu-se em 4655,7%, chegando em 2012 ao montante total superior a 54 bilhões de reais. Em termos regionais esse conjunto de dados mostra que o Nordeste do Brasil foi a segunda região do país a apresentar o maior crescimento de unidades financiadas no período de 2002 a 2012 (crescimento de 5843%), atrás apenas da região Norte. O estado de Pernambuco correspondeu ao segundo estado da região Nordeste com maior número de imóveis financiados no ano de 2012 (14,75%) atrás apenas da Bahia.

Como argumentam Coelho Júnior & Silveira Neto (2015a), a partir de informações do Secovi (o sindicato da habitação no Brasil), a primeira década dos anos 2000 foi marcada por uma elevação do preço dos imóveis nas cidades brasileiras, por exemplo, a magnitude do aumento real do preço do metro quadrado na cidade de São Paulo foi cerca de 188% no período de 2002 a 2012. No que diz respeito a Cidade do Recife, a partir de informações do Imposto sobre a Transmissão de Bens Imóveis (ITBI), é possível registrar que a magnitude do aumento do valor médio dos preços dos imóveis chegou a 326,5% para o período compreendido de 2002 a 2012. Como mostrado por Coelho Junior *et al.* (2014), o mercado imobiliário é bem relevante para a atividade econômica da cidade, dado que o somatório dos preços das habitações negociadas corresponde ao percentual médio de 5,19% do PIB ao longo do período 2000-2011.

Para dimensionar o tamanho desse movimento de valorização dos imóveis nesse mesmo período de tempo (2002-2012) a renda *per capita* ao nível de domicílio no Brasil da parte da população que apresentou o maior percentual de elevação (o grupo dos 40% mais pobres) registrou um aumento de 85,9% da renda segundo dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), magnitude essa que não chega a metade das cifras registradas para o aumento no preço dos imóveis em São Paulo e em Recife. No Brasil, o gasto com habitação corresponde a um considerável montante do dispêndio da renda familiar. Segundo Almeida & Azzoni (2013), esse custo habitacional nas regiões metropolitanas do país (Recife é uma das regiões metropolitanas que foram objeto dessa citada análise) equivale a magnitude de 25% a 30% dos gastos das famílias.

Logo, o expressivo aumento do crédito imobiliário adicionado à grande elevação do nível de renda das famílias e, conseqüentemente, ao aumento do dispêndio com habitação (dado a este gasto corresponder a um percentual do nível da renda das famílias, o qual apresentou crescimento), contribuíram para o movimento ascendente dos preços das residências no Brasil. Apesar disso pouco se conhece sobre a magnitude da contribuição das características estruturais para o preço da habitação ao longo dos anos. O caminho para entender a dinâmica de preços dos imóveis pode ser feito pela construção de um índice de preços mais abrangente em sua estrutura de variáveis avaliadas (considerando covariáveis relacionadas a estrutura do imóvel). Uma vez que, o cálculo de um índice de preços, considerando variáveis estruturais em seu cálculo, permitirá a obtenção da trajetória das mudanças nos preços mais fidedigna com a dinâmica de preço dos imóveis.

A partir da revisão bibliográfica, ao melhor de nosso conhecimento, constata-se que há a lacuna na literatura de não existir um índice de preços dos imóveis ao nível de uma divisão geográfica menor que cidade (mas que abranja toda a cidade) considerando em sua estrutura de estimação características estruturais; e que use para o seu cálculo uma amostra de residências pareada por essas características, garantido a utilização de moradias semelhantes para o cálculo de um índice de variação de preços de imóveis, ou seja, permitindo que as

variações de preço constatadas sejam de fato decorrentes de mudanças na precificação do imóvel. Este artigo tem como objetivo preencher essa lacuna observada estimando, tanto ao nível do Recife, quanto ao nível de RPA, um índice de preço de apartamentos (HPI) considerando apenas suas características intrínsecas.

Para tal, o índice de preços dos apartamentos aqui proposto, usa-se a abordagem do emparelhamento (*matching approach*) para geração do banco de dados a ser utilizado, o que permite comparar residências similares ao longo do tempo, ou seja, obter um cenário semelhante ao de vendas repetidas (PAREDES, 2011; MCMILLEN, 2012; LOPEZ & AROCA, 2012; e LOPEZ & DU, 2014), estima-se esse índice pela estrutura do índice de Fisher usando regressões hedônicas e adiciona-se três contribuições para a literatura de índice de preço. A primeira no que diz respeito ao processo de emparelhar, usando-se características intrínsecas, melhorando-se a acurácia das regressões hedônicas e, portanto, a precisão dos índices de preços das habitações propostos (ao nível de cidade e ao nível de sub-regiões da cidade). Em decorrência dessa primeira contribuição tem-se a segunda, que é estimar um índice de preços de imóveis (*housing price index*, HPI), que considere em sua estrutura características intrínsecas baseado em preços de venda registrados no ITBI para um período superior a um década. A terceira contribuição é calcular o HPI ao nível de sub-regiões oficiais da cidade (mas que abrange toda a cidade), estratégia que permite ser mais preciso e capturar particularidades locais.

Para este presente trabalho tem-se como objeto de estudo a Cidade do Recife (capital do estado de Pernambuco, um dos nove estados da região do Nordeste do Brasil), no período de 2000 a 2012, com dados de todo mercado formal de imóveis negociados, obtidos junto a prefeitura municipal do Recife, a partir do banco de dados do ITBI. Além da disponibilidade de dados, ao menos duas razões fazem o caso do Recife relevante. Primeira, é a capital da maior região metropolitana do Nordeste, região que apresentou os maiores ganhos de renda entre macrorregiões brasileiras nos anos 2000. Segunda, é a sede do estado que apresentou o segundo maior crescimento em termos de financiamentos contratados de residências de 2002-2012 na região Nordeste (segunda região do país a apresentar o maior crescimento de unidades financiadas nesse período). Aguarda-se, pois, com essas duas razões, que, junto com a expansão da renda, a expansão do crédito imobiliário possa potencialmente assumir formas diversas na dinâmica de preço do mercado de imóveis.

Com a presente introdução, este artigo está composto por 4 seções. Na que segue, seção 2, Estratégia Empírica, é descrito as etapas para estimar o índice de preços dos apartamentos, detalhando-se o instrumental empregado (métodos de *matching*, estimação de preços hedônicos e o que a diferencia da tradicional, o índice de Fisher, o caminho para estimar o intervalo de confiança – *bootstrap*), bem como, apresenta-se os dados e a sua análise descritiva. Logo em seguida mostrar-se, na seção 3, os resultados dos índices de preços e finaliza-se, na seção 4, com as conclusões.

## **2. Estratégia Empírica**

### **2.1 Índice de preços dos apartamentos**

Para estimar o índice de preços dos apartamentos aqui propostos é seguido 3 etapas, a saber: a primeira, gera-se a amostra de habitações ao longo dos anos com características comparáveis (intrínsecas) pelo *matching process*. Em seguida, na segunda, é usado esses dados gerados para estimar dois índices de preços para imóveis do Recife: o índice de preços da mediana (como o mercado imobiliário faz) e o índice de preços *à la* Fischer com características intrínsecas. Na terceira, calcula-se um intervalo de confiança pela abordagem *bootstrap* para o índice de preços *à la* Fischer aqui proposto. Isso é feito, para mostrar que a

trajetória desses FHPI está dentro do intervalo de confiança com nível de significância de 5%, ou seja, há significância estatística para a informação gerada.

A desagregação espacial usada aqui também distingue este presente trabalho dos trabalhos anteriores. A desagregação foi Regiões Políticas Administrativa (RPA). O nível de RPA reúne os 94 bairros do Recife em 6 divisões oficiais da cidade. Assim, como já foi dito na introdução, torna-se possível captar com mais precisão o efeito de localização espacial sobre o comportamento do mercado imobiliário do Recife durante o período do estudo.

Respalda-se as inferências realizadas aqui em aplicações internacionais para índices de preços como Lopez & Du (2014), Lopez & Aroca (2012), McMillen (2012), e Paredes (2011). Eles realizaram estimativas de índices de preços das habitações usando apenas características intrínsecas do imóvel. Também se usa a orientação de Lonford (2009) para aplicar *propensity score matching*<sup>1</sup> para fazer frente às limitações da abordagem de preços hedônicos no contexto de imóveis. Ao melhor do conhecimento, este é o primeiro estudo a propor um índice de preços para setor de habitação que, simultaneamente, busca a comparação de apartamentos comparáveis desde a concepção da sua amostra até a estrutura da estimativa do índice, incluindo-se características intrínsecas a fim de fazer uma comparação entre moradias realmente comparáveis.

Isso permite que a variação de preço capturada no índice hedônico *à la* Fisher aqui estimado esteja refletindo apenas mudanças de preços e não mudanças em atributos do(s) imóvel(éis), dado que, são considerados imóveis semelhantes desde a amostra (que foi pareada). Na situação de utilização de imóveis não semelhantes para aferição de índice de preço, pode ser verificado uma subvariação ou supervariação do índice de preço, dependendo de para qual sentido a(as) característica(as) que distingue(em) o(s) imóvel(eis) influencia(m) na variação de preço. A título de exemplificação, ao garantir residências comparáveis ao longo da amostra (ano de referência (2000) e demais anos (2001 até 2012)), está sendo dito que um imóvel no ano de 2000 com uma lista de atributos está sendo confrontado com outras moradias semelhantes nos demais anos da amostra, ou seja, dotadas de semelhantes características. Logo, empregando-se imóveis semelhantes, está se buscando obter no FHPI aferir a variação de preço com muito mais precisão. Além disso, são seguidas três etapas para estimação do FHPI, considerando-se a heterogeneidade da região em um nível inferior a cidade, por se usar a divisão de RPA.

## 2.2 Métodos de *matching*

Utiliza-se quatro abordagens de *Matching*: (1) *One2One matching*, (2) *K-nearest neighbor matching* (correspondência pelos *k*-vizinhos mais próximos) ( $k = 5$ ), (3) *Kernel matching* e (4) *Mahalanobis matching*. A razão para fazer isso é encontrar a melhor maneira de fazer a correspondência. Escolhe-se o melhor caminho olhando para o ajustamento em termos do percentual de redução do viés/do diferencial, em cada abordagem, entre as características dos imóveis do grupo de controle e as características das residências dos outros anos que se busca parear. A métrica que se usa para tomar a decisão sobre a melhor abordagem é a métrica da diferença padronizada de Rosenbaum & Rubin (1985). De acordo com esses autores, o percentual de redução do viés será quantas variáveis em cada processo de *matching* tem menos que 10% de diferença padronizada (a métrica Rosenbaum & Rubin (1985)) entre o grupo de controle e o grupo tratado. Numericamente, a diferença padronizada em percentagem é de,  $100 (\hat{X}_1 - \hat{X}_0) / [S_1^2 + S_0^2 / 2]^{1/2}$ . Em que  $\hat{X}_1$  é a média do grupo tratado,  $\hat{X}_0$  é a média do grupo de controle,  $S_1^2$  é a variância do grupo tratado e  $S_0^2$  é a variância do grupo de controle.

---

<sup>1</sup> O método de *matching* é utilizado neste trabalho para gerar o conjunto de amostras comparáveis ao longo do período de estudo para cada região política administrativa.

Seguindo Paredes (2011) e Lopez & Du (2014), espera-se que quando sejam comparados os métodos de correspondência, ver-se-á que a maior redução de viés será na abordagem que utiliza a minimização da diferença nas características da habitação (por exemplo, *Mahalanobis*) em detrimento dos métodos que usam a minimização da distância usando apenas a representação da distribuição (esse é o caso dos métodos de correspondência que usam o *propensity score* (PS) e também do *kernel matching*). Detalhando rapidamente os métodos de correspondência, o (1) *One2One matching* vai encontrar a moradia com características mais próximas (denominadas aqui de clone) no grupo controle e no grupo tratado olhando para *propensity score*, o (2) *K-nearest neighbor matching* ( $k=5$ ) vai seguir os mesmos passos da primeira correspondência, mas ele vai mudar pelo fato de que o clone será agora dentro de  $k$  vizinhos mais próximos no que diz respeito às características (a literatura frequentemente usa  $k = 5$  vizinhos, e foi a que se adotou), o (3) *Kernel matching* vai usar uma média ponderada na busca por clone, neste processo, ele vai fazer associação com todas as observações dentro da amostra, mas ele vai colocar mais peso na observação mais próxima e o (4) *Mahalanobis matching* aplica as distâncias de *Mahalanobis* para os atributos da correspondência.

Em geral os *matching* baseados no *propensity score* (PS) são feitos com base em dois estágios. O primeiro, é para estimar o *propensity score*, utilizando um modelo logit<sup>2</sup> (COX, 1970). A segunda etapa é para construir a amostra com as residências emparelhadas. Para criar essa amostra usa-se a probabilidade predita  $\hat{q}(x)$ .

Como argumentam Lopez & Du (2014), quando se fala de *Mahalanobis matching* existem duas vias para fazer o processo de correspondência: 1º) *Mahalanobis matching with propensity scores as covariates* (MPSCov): Neste caso, o primeiro passo será a estimação do *propensity score*, no entanto a habitação clone será escolhida a partir da distância de *Mahalanobis*  $d_M(x, \hat{q}(x))$ , que será usada para fazer a correspondência entre as características. A  $d_M$  inclui o PS como uma covariada de  $\hat{V}_{x,q(x)}$ .

$$d_M(x, \hat{q}(x)) = (x - \hat{q}(x))' (\hat{V}_{x,q(x)})^{-1} (x - \hat{q}(x)) \quad (1)$$

2º) *Mahalanobis matching with propensity scores as calipers* (MahalPSCal): Esta abordagem irá fazer a correspondência combinando os métodos *One2One* e *MPSCov* para fornecer uma sofisticada pré-filtragem de correspondência. Inicialmente, estima-se o *One2One* removendo as observações não emparelhadas. A partir dessa amostra de correspondentes, aplica-se a correspondência de *Mahalanobis* utilizando os escores de propensão estimados como *calipers*. No presente trabalho foi utilizada a última, a *MahalPSCal*.

O *caliper* “ $c$ ” define a “janela” para encontrar o par de residências na distância de *Mahalanobis*. Rosenbaum & Rubin (1985) mostraram que há uma regra para obter o correto *caliper*:  $c = n \sigma$ , em que  $n$  é um número que será definido a partir do valor  $\sigma$  e utilizando a tabela 2.3.1 de Cochran & Rubin (1973). O valor  $\sigma$  é:  $\sigma = [\sigma_{t_1}^2 / \sigma_{t_0}^2]^{1/2}$ , em que  $\sigma_{t_1}^2$  e  $\sigma_{t_0}^2$  são as variâncias do *propensity score* do grupo tratado ( $t_1$ ) e do grupo de controle ( $t_0$ ), respectivamente. Para cada amostra<sup>3</sup> (foram 72 amostras *matched*), foi calculado o valor  $\sigma$  para obter o valor  $n$  (parte numérica do *caliper*) seguindo Cochran & Rubin (1973). Depois, tomou-se a média dos valores de  $\sigma$ . Isso foi feito<sup>4</sup> para o modelo com características intrínsecas. A média foi 0,89. Então, usou-se a tabela 2.3.1 de Cochran & Rubin (1973) para

<sup>2</sup> O modelo de escolha discreta binário.

<sup>3</sup> Isso significa para cada amostra emparelhada entre os grupos de controle e tratado em cada RPA. Em outras palavras, tem-se 72 amostras pareadas, uma vez que, em cada RPA, a amostra foi dividida em 12 amostras emparelhadas entre os grupos controle e tratado, e há 6 RPAs. Correspondente aos 2000 a 2012.

<sup>4</sup> Então, esse procedimento foi feito 72 vezes (dado que são 72 amostras emparelhadas).

obter o valor de  $n$ . Seguindo-se a tabela 2.3.1 o valor de  $n$  foi de 0,6. Definido  $n$  igual a 0,6; utilizou-se o valor de  $n$  para calcular o *caliper* de cada amostra emparelhada.

Ao aplicar esses métodos de emparelhamentos constata-se que para o modelo com características intrínsecas o melhor processo de correspondência, utilizando a métrica de diferença padronizada de acordo com Rosenbaum & Rubin (1985), foi *Mahalanobis matching*. Portanto, a amostra emparelhada usando a abordagem de *Mahalanobis* foi a escolhida para estimar o índice de preços das habitações *à la Fisher* (FHPI) aqui propostos<sup>5</sup>.

### 2.3 Abordagem tradicional versus abordagem alternativa (preços hedônicos)

Na abordagem tradicional, de acordo com Brueckner (2011), a demanda por habitação pode ser medida unidimensionalmente através da quantidade de espaço físico de um imóvel residencial urbano:

$$q = \alpha p^\beta y^\theta \quad (2)$$

Sendo:  $q$  consumo de habitação,  $p$  preço por unidade de habitação,  $y$  renda,  $\beta$  elasticidade-preço da demanda<sup>6</sup> (ela tem um sinal negativo, variando no *range* de  $-1 < \beta < 0$ ) e  $\theta$  elasticidade renda da demanda<sup>7</sup> (ela tem um sinal positivo, variando no *range* de  $0 < \theta < 1$ )<sup>8</sup>.

A função utilidade é expressa em termos do consumo de outros bens de consumo, diferentes de habitação ( $c$ ) (por simplificação, pode ser pensado como quanto é gasto em alimentação) e  $q$ , isto é,  $u(c, q)$ . Empiricamente, a modelagem econômica urbana faz ser possível estimar a função demanda para habitação, sem especificar como o consumo é medido. Os procedimentos adotados foram os seguintes. Multiplicando ambos os lados da igualdade (2) por  $p$ , tem-se:

$$pq = \alpha p^{\beta+1} y^\theta \quad (3)$$

Em que  $pq$  indica os gastos com habitação. Considerando a perspectiva do inquilino,  $pq$  representa o aluguel total. Do ponto de vista do proprietário,  $pq$  é o preço do imóvel ( $p$  é o preço de compra<sup>9</sup> por unidade de área de habitação). Denotando  $pq$  por  $E$ , e aplicando o logaritmo natural e denotando  $\log(\alpha)$  por  $\gamma$ :

$$\log E = \gamma + (\beta + 1)p + \theta \log(y) \quad (4)$$

A equação (4) dá as despesas com habitação sem medir o consumo de espaço físico da habitação. Obtendo-se  $\beta$  e  $\theta$  por análise de regressão usando  $p$  e  $y$ .

A abordagem alternativa, chamada hedônica, admite que a escolha da habitação envolve outros atributos (cada *i-ésimo* atributo denotado por  $x_i$ ) além do espaço físico. Ou seja, ele argumenta que não é possível medir o consumo de habitação por apenas uma variável. Então, neste contexto, a função de utilidade é expressa por  $u(\alpha, x_1, x_2, x_3, \dots, x_m)$ , indicando que a preferência dos consumidores, como a propriedade<sup>10</sup>, é individual. Sendo  $\alpha$  a constante e  $m$  indica o número de atributos considerados. Uma característica especial da residência é precisamente a sua heterogeneidade, assim, em geral, a literatura econômica que procura entender a constituição de preços dos imóveis faz uso da função de preços hedônicos, a fim de compreender o preço da habitação. Ou seja, por meio da modelagem hedônica,

<sup>5</sup> O detalhamento dos resultados dessa etapa podem ser solicitado por e-mail a um dos autores. Os mesmos não foram colocados aqui por limitação de espaço.

<sup>6</sup> Em geral, pode-se considerar a demanda por habitação é inelástica a preço, indicando baixa capacidade de responder às mudanças dos preços (BRUECKNER, 2011).

<sup>7</sup> O consumo de habitação sobe menos que proporcionalmente a renda, de modo que o gasto com habitação compreenderá uma porção cada vez menor de  $y$  quando a renda aumenta.

<sup>8</sup> Ver Mayo (1981) para uma discussão sobre a magnitude do  $\beta$  e  $\theta$ .

<sup>9</sup> A *proxy* de  $p$  é geralmente um índice de preços das habitações a nível de cada cidade e a *proxy* para  $E$  é o valor da propriedade.

<sup>10</sup> O termo propriedade neste trabalho é sinônimo de moradia.

procura-se obter a relevância marginal que cada característica da propriedade tem sobre a formação do preço da moradia.

Rosen (1974, p.34) definiu preços hedônicos como preços tácitos / implícitos relativos às características que realmente importam para os indivíduos, e estes preços são revelados aos agentes econômicos a partir de: (I) preços observados de produtos diferenciados e (II) quantidades específicas de características associadas a eles. Por esta abordagem, o preço do bem será a soma dos preços tácitos de seus atributos. No mercado imobiliário, uma vez que o preço disponível é apenas o preço de negociação, as regressões hedônicas irão explicar o preço da habitação em função das características da habitação. Portanto, quando estima-se o modelo de preço hedônico (HPM) chega-se ao preço implícito para cada atributo da habitação. Quando trabalha-se com o HPM precisa-se definir duas características: A forma funcional do modelo e o período de tempo, que o HPM será estimado. Sobre o primeiro ponto, existem muitas formas funcionais, as mais comuns são *log-log* e *log-linear*. Caso use a *log-log* pode-se interpretar o coeficiente estimado das variáveis como elasticidades, mas desde que os dados de nenhuma das variáveis explicativas apresente o valor zero. Sendo assim, aplica-se, para o índice de preços, a *log-linear* :

$$\ln y_{it_0} = \alpha_{t_0} + \sum_{k=1}^K \beta_{t_0}^k X_{it_0}^k + \sum_{l=1}^L \lambda_{t_0}^l Z_{it_0}^l + \varepsilon_{it_0}$$

$$\ln y_{it_1} = \alpha_{t_1} + \sum_{k=1}^K \beta_{t_1}^k X_{it_1}^k + \sum_{l=1}^L \lambda_{t_1}^l Z_{it_1}^l + \varepsilon_{it_1}$$
(5)

Em que:  $y_i$  é o preço da habitação “ $i$ ”,  $\alpha$  é o intercepto de regressão,  $\beta^k$  representa os preços hedônicos estimado para cada características “ $k$ ” da casa,  $k = \{1, 2, \dots, K\}$  e  $X^k$  é um vetor de variáveis, um vetor para cada característica  $k$ ,  $\lambda^l$  é o coeficiente para cada característica não observada “ $l$ ”,  $l = \{1, 2, \dots, L\}$ , e  $Z^l$  é o vetor de variáveis dos atributos não observados. Assume-se que este último vetor será controlado por causa dos benefícios de processo de *matching*.

Em relação à segunda característica, estima-se o HPM para cada amostra emparelhada em cada RPA. Para identificar se a informação é do ano de controle usa-se  $t_0$  e usa-se  $t_1$  para o ano tratado. Isso é feito porque não é possível supor-se que os coeficientes característicos não mudam ao longo dos anos. Desta forma, a variação desses coeficientes vai capturar as mudanças sobre como os consumidores vão colocar preço sobre as características da habitação para cada ano avaliado. Permitindo capturar mudanças nas preferências dos consumidores, que é extremamente importante quando o objetivo é conhecer em detalhes o mercado imobiliário.

## 2.4 Índice de Fisher

Meese & Wallace (1991) começaram a aplicação do índice de Fisher (F) ao mercado de habitação (índice de preços das habitações *à la* Fisher – FHPI). Eles usaram os dados dos condados da Alameda e de São Francisco, nos EUA, para estimar um índice de Fisher usando os *outputs* da técnica de regressão ponderada localmente (LWR). Depois, Paredes & Aroca, (2008) e Paredes (2011) aplicaram o FHPI para o caso espacial usando apenas um período de tempo. Lopez & Aroca (2012) usaram também esse índice para fazer cálculos espaciais e temporais. E Lopez & Du (2014) utilizaram o Índice Fisher para pequenas áreas estatísticas metropolitanas (SMSAs) do estado de Illinois, nos EUA.

Meese & Wallace (1991) defenderam que os índices de Fisher têm vantagens quando comparado com outros índices. Diewert (1978) mostrou que os índices estimados *à la* Fisher são superiores e exatos, porque os índices de Fisher permitem comparar os índices estimados para diferentes regiões e eles são derivados a partir de uma função subjacente de utilidade ou

de produção<sup>11</sup>. Essas características dos índices de Fisher serão importante para a construção de um índice para o mercado de imóveis. Como Griliches (1971) apontou, a maneira como o índice de Fisher é construído (equações (6) e (7)) reduz o viés potencial. Ele usa a média geométrica entre dois índices: o índice de Laspeyres (L) e o índice de Paasche (P). Caso seja usado apenas o índice Laspeyres, ter-se-á uma superestimativa do índice de preços. E se escolher usar somente o índice de Paasche, obter-se-á uma subestimativa<sup>12</sup>.

$$L = \frac{p_1 q_0}{p_0 q_0}, \quad P = \frac{p_1 q_1}{p_0 q_1}, \quad F = \sqrt{L * P} \quad (6)$$

Aplicando o logaritmo natural ao longo equação 9 e executando manipulações algébricas chega-se à seguinte representação para o índice de Fisher:

$$\ln F = 0.5[\ln(p_1 q_0) - \ln(p_0 q_0)] + 0.5[\ln(p_1 q_1) - \ln(p_0 q_1)] \quad (7)$$

Em que o número 0 e 1 conecta se a informação é do grupo de controlo ( $t_0$ ) ou se é do grupo tratado ( $t_1$ ), respectivamente. Os termos na equação 10, com o logaritmo natural aplicada pode ser compreendido como:  $\ln(p_1 q_0)$ : Preço de venda da moradia baseado em características observadas no ano de controle ( $q_0$ ) a preços do ano tratado ( $p_1$ ).  $\ln(p_0 q_0)$ : Preço de venda da moradia baseado em características ( $q_0$ ) e em preços ( $p_0$ ) observados no ano de controle.  $\ln(p_1 q_1)$ : Preço de venda da moradia baseado em características observadas ( $q_1$ ) e nos preços ( $p_1$ ) do ano tratado.  $\ln(p_0 q_1)$ : Preço de venda da moradia baseado em características observadas no ano tratado ( $q_1$ ) a preços do ano de controle ( $p_0$ ).

Os termos  $\ln(p_0 q_0)$  e  $\ln(p_1 q_1)$  são, em sequência,  $\ln \bar{y}_{it_0}$  e  $\ln \bar{y}_{it_1}$ , ou seja, eles correspondem a média dos preços de venda observados da habitação  $t_0$  e  $t_1$  respectivamente. Mas os termos  $\ln(p_1 q_0)$  e  $\ln(p_0 q_1)$  devem ser calculados antes do FHPI ser estimado através da regressão hedônica na equação (5):

$$\ln(p_1 q_0) = \hat{\alpha}_{t_1} + \sum_{k=1}^K \hat{\beta}_{t_1}^k \bar{X}_{it_0}^k \quad (8)$$

$$\ln(p_0 q_1) = \hat{\alpha}_{t_0} + \sum_{k=1}^K \hat{\beta}_{t_0}^k \bar{X}_{it_1}^k$$

Substituindo esta informação na Equação (7):

$$\ln F = 0.5[(\hat{\alpha}_{t_1} + \sum_{k=1}^K \hat{\beta}_{t_1}^k \bar{X}_{it_0}^k) - \ln \bar{y}_{it_0}] + 0.5[\ln \bar{y}_{it_1} - (\hat{\alpha}_{t_0} + \sum_{k=1}^K \hat{\beta}_{t_0}^k \bar{X}_{it_1}^k)] \quad (9)$$

Aplicando-se o exponencial:

$$F = \{\exp[(\hat{\alpha}_{t_1} + \sum_{k=1}^K \hat{\beta}_{t_1}^k \bar{X}_{it_0}^k) - \ln \bar{y}_{it_0}] * \exp[\ln \bar{y}_{it_1} - (\hat{\alpha}_{t_0} + \sum_{k=1}^K \hat{\beta}_{t_0}^k \bar{X}_{it_1}^k)]\}^{1/2} \quad (10)$$

Esta última equação é a equação que é utilizada neste trabalho para fazer a estimativa do FHPI para cada RPA em cada ano do conjunto de dados.

## 2.5 Intervalo de confiança pela abordagem *bootstrap*

Uma vez calculado os FHPIs ao nível de RPAs estima-se o intervalo de confiança (IC) para o índice aferido em cada RPA utilizando-se a abordagem *bootstrap* com nível de significância de 5%, dado que não se há a certeza de se estar trabalhando com o universo de habitações da cidade, é necessário estimar o IC para dar confiança estatística (robustez) a informação gerada. A ideia é mostrar que os FHPIs estimados têm sua trajetória dentro do IC

<sup>11</sup> Assumindo-se que os consumidores e os compradores são semelhantes, pode-se mostrar que a função demanda hedônica poderia ser representada pela equação de demanda compensada (ROSEN, 1974).

<sup>12</sup> Coelho Júnior & Silveira Neto (2015b) apresentam na Figura 4 o detalhamento do viés do Índice Laspeyres e do Índice de Paasche.

ao nível de significância considerado. Escolhe-se a abordagem *bootstrap*, porque a literatura<sup>13</sup> demonstrou que a abordagem *bootstrap* é eficiente para estimar o intervalo de confiança.

*Bootstrap* é um método de reamostragem introduzido por Efron (1979) com o objetivo de substituir as técnicas complicadas e questionáveis de análise estatística com base no cálculo intensivo. Pode-se usar o *bootstrap* para abordagens paramétricas e não paramétricas. Na primeira, assume-se que a função de distribuição ( $F$ ) pode ser estimada a partir de um modelo paramétrico conhecido. Na segunda, não se conhece ( $F$ ) e assume-se que se pode estimar pela distribuição empírica ( $\hat{F}$ ).

No caso deste trabalho, utiliza-se a abordagem não paramétrica. A ideia básica de *bootstrap* não paramétrico pode ser sintetizada da seguinte forma, dada uma amostra aleatória de tamanho  $n$ ,  $X = (x_1, x_2, \dots, x_n)$ , de uma população com uma ( $F$ ) desconhecida, calcula-se o estimador  $\hat{\beta} = s(\hat{F})$  para o parâmetro de interesse  $\beta = s(F)$ . Assume-se que  $\hat{F}$  é a distribuição empírica de  $X$ , e, em seguida, é construído uma amostra *bootstrap* de tamanho  $n$ ,  $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ , através de  $n$  extrações aleatórias com reposição de  $X$ . Isso significa que  $X$  é tratado como a população. A replicação *bootstrap* do parâmetro de interesse para esta amostra *bootstrap* é denotado por  $\hat{\beta}^*$ .

Se  $B$  amostras *bootstrap* foram geradas,  $X_1^*, X_2^*, \dots, X_B^*$ , a replicação *bootstrap* do parâmetro de interesse para  $b$ -ésima amostra é  $\hat{\beta}^* = s(X_b^*)$ . Isto é, é o valor de  $\hat{\beta}$  para amostra *bootstrap*  $X_b^*$ . Para determinar os intervalos de confiança *bootstrap*, que exige o conhecimento da amostra *bootstrap* da distribuição, Efron & Tibshirani (1993) defendem, pelo menos,  $B = 1.000$  replicações *bootstrap*, sendo isso seguido neste trabalho. Usando *bootstrap*, pode-se obter o intervalo de confiança aproximado de  $100(1 - 2\alpha)\%$  para os parâmetros de interesse  $\beta$ . Para cada amostra *bootstrap* calcula-se os erros padrões e, em seguida, estima-se os intervalos de confiança. Em termos práticos, divide-se a equação (9) do índice de preços das habitações *à la Fisher* em duas partes  $F_1$  e  $F_2$ . Desta forma, o FHPI é  $F = e^{\ln F}$ . Sendo:  $\ln F = F_1 + F_2$ .

$$F_1 = 0.5 \left[ (\hat{\alpha}_{t_1} + \sum_{k=1}^K \hat{\beta}_{t_1}^k \bar{X}_{it_0}^k) - \ln \bar{y}_{it_0} \right] \quad \& \quad F_2 = 0.5 \left[ \ln \bar{y}_{it_1} - (\hat{\alpha}_{t_0} + \sum_{k=1}^K \hat{\beta}_{t_0}^k \bar{X}_{it_1}^k) \right] \quad (11)$$

Em seguida, tem-se a amostra *bootstrap* para  $F_1$  e  $F_2$ . Depois estima-se  $\ln F = F_1 + F_2$  e  $F = e^{\ln F}$ , e calcula-se o intervalo de confiança para o FHPI proposto pela abordagem *bootstrap*. Isso é feito para cada amostra emparelhada em cada RPA. Em outras palavras, isso é feito para as 72 amostras pareadas (em cada RPA, a amostra foi dividida em 12 amostra pareadas entre os grupos de controle e tratado. São 6 RPAs).

## 2.6 Dados

O conjunto de dados usados para proceder as análises deste artigo foi do Imposto de Transmissão de Bens Imóveis (ITBI<sup>14</sup>) fornecido pela Prefeitura Municipal do Recife. Pode-se destacar duas principais vantagens em utilizar esses dados: (I) a abrangência de todo o mercado formal<sup>15</sup> de residências na cidade do Recife, garantindo uma quantidade de informação superior a de outras fontes de dados, e (II) os preços registrados nesse conjunto de dados são os preços mais próximos dos preços de transação no mundo real, porque, apesar do fato de os preços do conjunto de dados sejam declarados pelos agentes econômicos, os funcionários da prefeitura auditam esses preços declarados e em situação de discrepâncias

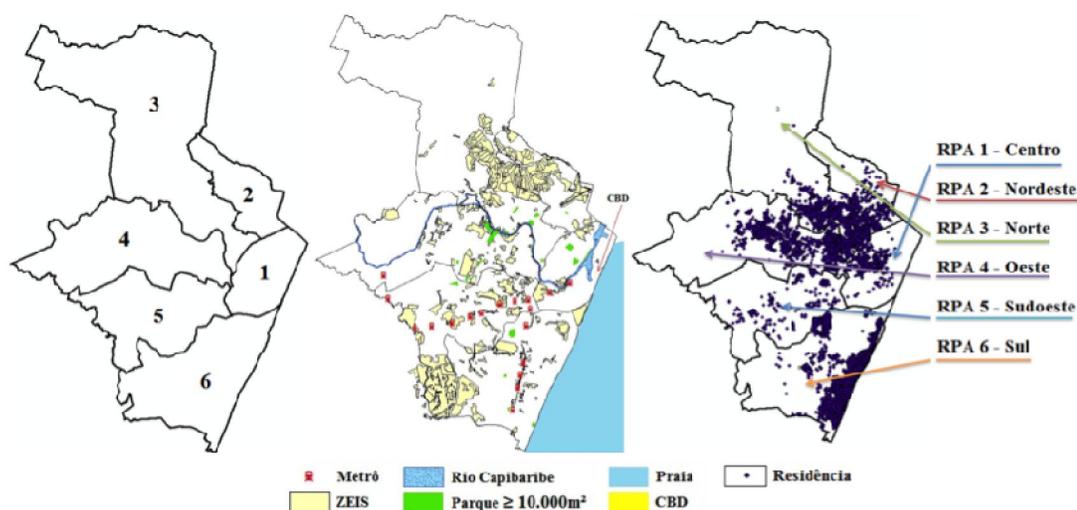
<sup>13</sup> Você pode pesquisar mais sobre nas referências: Davison & Hinkley (1998) e Mooney & Duval (1993).

<sup>14</sup> Pela Constituição Federal (art. 156, II), o ITBI corresponde ao imposto sobre transmissão, por ato oneroso, de bens imóveis, ou de direito reais sobre imóveis. .

<sup>15</sup> Por outro lado, como consequência do uso de dados do mercado formal, há a fragilidade de não abrangência da totalidade do mercado, por não estar sendo incluído os dados do mercado imobiliário informal. A consequência é que não se captura a dinâmica total do mercado imobiliário.

com os preços praticados no mercado imobiliário da cidade do Recife o imóvel tem seu preço revisado. Dessa forma, pode-se supor que os preços registrados estão muito perto do preço de equilíbrio (quando o preço do ofertante é igual ao preço que o consumidor gostaria de pagar, e que permite que a transação seja realizada). Portanto, é possível diminuir bastante o problema da superestimação e/ou da subestimação dos preços de venda da moradia<sup>16</sup>.

Nos dados ITBI há informações sobre o preço do imóvel, endereço e as características estruturais (características intrínsecas) de cada habitação vendida legalmente entre 2000 e 2012. Quanto ao preço da habitação aplicou-se o logaritmo aos preços. Para definir as variáveis intrínsecas usadas no FHPI, seguiu-se Sirmans<sup>17</sup> *et al.* (2005) e Seabra & Silveira Neto (2014) quando foi possível segui-los, dado a disponibilidade de dados. Para o FHPI utilizou-se as seguintes características intrínsecas: (1) andar onde está a habitação (*Floor*), (2) a idade da habitação (*Age*), (3) área privada construída (*Private\_bu*), (4) padrão de construção (*Standard\_construct*), assumindo valor 1 se o padrão igual a médio ou superior e 0 se igual a simples, e (5) densidade<sup>18</sup> de apartamentos por andar (*Unit\_den*).



**Figura 2. A distribuição de residências vendidas e de características de ambiente (amenidades) no Recife entre 2000 - 2012**

Fonte: Elaboração própria a partir dos dados do ITBI e *shapefiles* com as amenidades.

Nota 1: CBD = Centro da cidade.

Nota 2: ZEIS = *Proxy* de áreas pobres em Recife.

O montante total das operações formalmente registradas nesse período (2000 - 2012), o qual serviu de base para proceder o processo de emparelhamento, foi de 97.466 casos. Após esse processo só foi possível usar dados de apartamentos, porque no processo de *matching* das casas não foram encontradas casas semelhantes na RPA 5. O montante total ficou constituído

<sup>16</sup> Em um ambiente sem supervisão podem ocorrer duas situações: (1) superestimação – na ótica de um consumidor, ele poderá declarar um valor mais alto, porque quando esse consumidor for vender, no futuro, sua habitação o imposto sobre o ganho de capital (Imposto de Renda (IR)) terá uma base de cálculo menor, uma vez que o preço inicial é superestimado. Já, na ótica de um vendedor, ele poderia procurar reduzir o preço declarado para pagar menos IR. Nesta situação o mercado até poderia regular o preço. (2) subestimação – pode ocorrer quando um consumidor declara um valor mais baixo para o valor de referência ser menor no cálculo do imposto de propriedade (Imposto Predial e Territorial Urbano (IPTU)), que anualmente deve ser pago e um vendedor também declara um valor menor com o objetivo de diminuir o IR. Esta última situação, a de subestimação, é inibida pela fiscalização dos preços.

<sup>17</sup> Neste trabalho, os autores apresentam um panorama da literatura mostrando os atributos considerados frequentemente usados na literatura.

<sup>18</sup> Para obter esta variável calcula-se:  $units\_quan / floors\_q\_1$ . Em que: *units\_quant* é quantidade de unidades de apartamento que há no edifício e *floors\_q\_1* é a quantidade dos andares que o edifício tem.

por 82.501 apartamentos. Usar apenas apartamentos trouxe a vantagem, de que só foram comparadas unidades com a mesma classificação, tornando a amostra uniforme nesse aspecto, afastando o viés comparativo de se comparar algo diferente. Uma vez que, às vezes, o preço de uma casa pode refletir o poder de verticalização do terreno onde essa casa está edificada; não refletindo assim o preço associado as características consideradas no FHPI proposto. Dessa forma, isso garantiu que o FHPI proposto tivesse a mesma concepção em todas as RPAs e, conseqüentemente, fosse possível comparar as trajetórias verificadas nas RPAs.

O objetivo, ao estimar os FHPIs por RPAs, foi calculá-los para a menor unidade geográfica possível, buscando-se com isso captar as particularidades locais, mas que, ao mesmo tempo, abrangesse a totalidade da cidade. Dado que dos 94 bairros da cidade do Recife tinha-se informações no conjunto de dados do ITBI sobre o mercado imobiliário em 90 bairros, e que em anos específicos para alguns bairros não se tinha um número de observações que permitia aplicar a metodologia de emparelhamento para obter habitações semelhantes ao longo dos anos, a solução encontrada, para seguir com o objetivo de estimar o FHPI para uma divisão inferior ao nível da cidade, foi usar a divisão de regiões políticas administrativas (RPA). Há no Recife seis RPAs (Figura 2): RPA 1 (Centro), RPA 2 (Nordeste), RPA 3 (Norte), RPA 4 (Oeste), RPA 5 (Sudoeste), e RPA 6 (Sul).

Como pode ser visto na Figura 2, durante todos os anos no conjunto de dados do ITBI (2000 a 2012), a RPA 6 teve o maior número de vendas de habitações. A RPA 6 tem a particularidade de estar perto da praia. A segunda, em número de vendas de habitações, foi a RPA 3 e a terceira foi a RPA 4. Nestas três RPAs estão os bairros de classe média e alta no Recife e onde há a presença da maioria das habitações mais caras.

## 2.7 Análise descritiva

Após a definição que o método de emparelhamento a ser seguido seria o de *Mahalanobis* são apresentados, na Tabela 1, as estatísticas descritivas da amostra pareada à *la Mahalanobis*, as variáveis estruturais, ao longo dos treze anos do conjunto de dados (2000-2012). No que diz respeito as variáveis estruturais, as habitações vendidas ao longo de 2000-2012, em média, têm 115,82m<sup>2</sup> de área privativa, com uma idade média em torno de 13 anos e, em geral, elas são de padrão médio ou superior. A média de andar está entre os 7º e 8º andares e a média de apartamentos em cada andar foi de aproximadamente 4.

**Tabela 1. Estatística descritiva das variáveis dos modelos empíricos**

Variáveis	Média	Desv. Pad.	Min	Max
<b>Estrutural</b>				
<i>Floor</i>	7.50	5.96	1	41
<i>Age</i>	13.37	12.98	0	73
<i>Private_bu</i>	115.82	72.71	18	1187
<i>Standard_construct</i>	0.78	0.42	0	1
<i>Unit_den</i>	3.88	4.02	0.05	60
Total de observações	81150.00			

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota 1: O separador decimal usado foi o “.”.

Quando é feito o recorte dos dados por RPA (Tabela 02), para as variáveis estruturais, tem-se que, as habitações vendidas ao longo de 2000-2012, em média, têm maiores áreas privativas na RPA 3 (143,73m<sup>2</sup>) e RPA 6 (125,60m<sup>2</sup>), já as menores áreas privadas, na média, estão na RPA 1 (84,88m<sup>2</sup>) e RPA 4 (70,74m<sup>2</sup>). Quanto à idade, a RPA 1 é a que apresenta a maior idade média dos imóveis, que foi de 31 anos, já na RPA 5 é onde estão as habitações que apresentam, na média, os imóveis mais novos, em torno de 10 e 11 anos de idade. Quanto ao andar, os maiores valores, na média, ou seja, os andares mais elevados, na média, estão na RPA 3 (9º andar) e na RPA 6 (8º andar). Já os menores valores

**Tabela 2. Estatística descritiva das variáveis dos modelos empíricos por RPA**

Variáveis	RPA 1				RPA 2				RPA 3			
	Média	Desv. Pad.	Min	Max	Média	Desv. Pad.	Min	Max	Média	Desv. Pad.	Min	Max
<b>Estrutural</b>												
<i>Floor</i>	6.50	4.90	1	39	7.39	6.73	1	37	9.03	6.10	1	41
<i>Age</i>	31.26	13.77	0	73	12.57	12.98	0	69	10.98	11.80	0	63
<i>private_bu</i>	84.88	44.99	20	488	91.71	37.32	20.94	386.34	143.73	82.95	18	929.14
<i>standard_construct</i>	0.28	0.45	0	1	0.70	0.46	0	1	0.86	0.34	0	1
<i>unit_den</i>	5.79	4.31	0.06	20	3.42	1.86	0.17	17.5	2.75	1.60	0.06	20
Total de observações	4512.00				5642.00				18173.00			
Variáveis	RPA 4				RPA 5				RPA 6			
	Média	Desv. Pad.	Min	Max	Média	Desv. Pad.	Min	Max	Média	Desv. Pad.	Min	Max
<b>Estrutural</b>												
<i>Floor</i>	5.75	5.19	1	38	3.83	2.59	1	16	8.20	6.13	1	37
<i>Age</i>	11.06	10.70	0	65	10.61	11.44	0	63	13.90	12.72	0	68
<i>private_bu</i>	94.66	44.78	18.97	664.87	70.74	17.26	26.1	604.5	125.60	81.11	18	1187
<i>standard_construct</i>	0.73	0.45	0	1	0.63	0.48	0	1	0.85	0.35	0	1
<i>unit_den</i>	3.91	3.86	0.06	60	8.14	8.77	0.67	28	3.71	3.67	0.05	30.43
Total de observações	15858.00				4695.00				32270.00			

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota: O separador decimal usado foi o “.”.

**Tabela 3. Matriz de correlação para variáveis estruturais e de amenidades**

Variáveis	<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
<i>floor</i>	1				
<i>age</i>	<b>-0.34</b>	1			
<i>private_bu</i>	0.23	0	1		
<i>standard_construct</i>	0.32	-0.61	<b>0.23</b>	1	
<i>unit_den</i>	-0.12	0.20	<b>-0.31</b>	<b>-0.30</b>	1

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota: O separador decimal usado foi o “.”.

médios quanto ao andar estão na RPA 4 (entre o 5º e 6º andares) e na RPA 5 (entre o 3º e 4º andares). Quanto à média de unidades habitacionais em cada andar (*unit\_den*), o maior valor registrado foi na RPA 5, com 8 unidades por andar, já a região que apresenta a menor média de unidades por andar é a RPA 3, com aproximadamente 3 unidades por andar.

Em seguida, na Tabela 3, é apresentado a matriz de correlação para variáveis estruturais. Verifica-se uma correlação negativa (-0,34) entre idade (*age*) e andar (*floor*), indicando que imóveis mais velhos tende a ter menos andares. Há uma correlação positiva (0,23) entre área privada (*private\_bu*) e padrão da construção (*standard\_construct*), indicando que quando a área privada dos imóveis aumenta, tende-se a melhorar o padrão do imóvel. Para área privada e densidade de moradias por andar (*unit\_den*) tem-se uma correlação negativa (-0,31), logo maior área privada implica em menor número de unidades por andar. Quanto ao padrão da construção (*standard\_construct*) e densidade de moradias por andar (*unit\_den*) a correlação é negativa (-0,30), sendo assim, a melhora do padrão dos imóveis está associado á queda no número de unidades por andar.

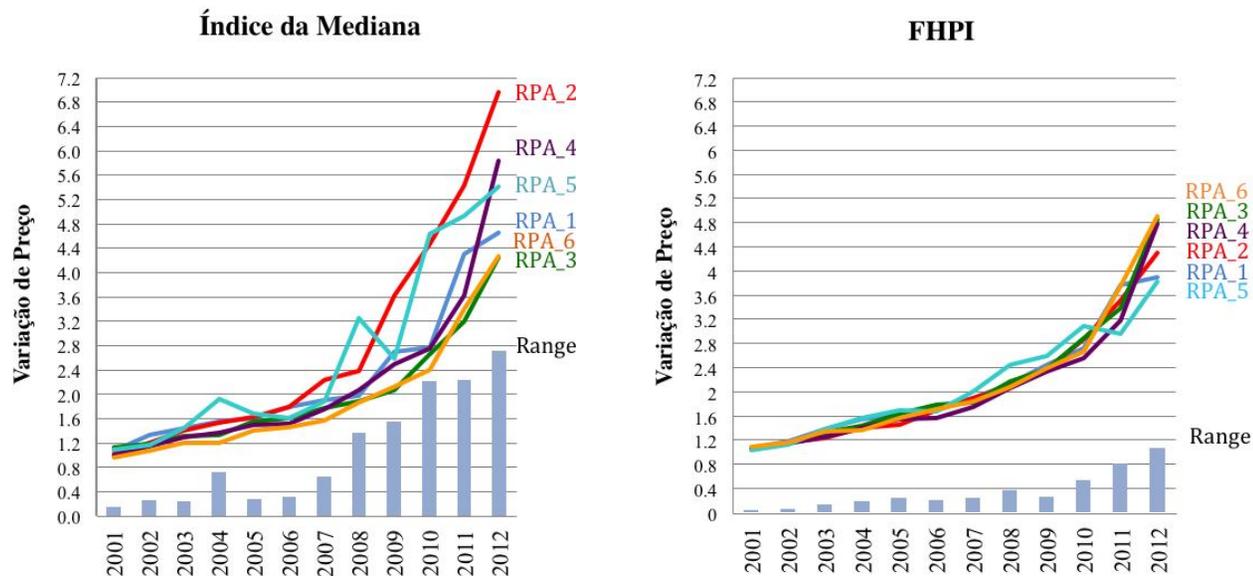
### 3. Resultados dos índices de preços

Neste ponto, apresenta-se os resultados do índice hedônico *à la* Fisher proposto para o mercado de habitação, juntamente com os resultados obtidos ao se estimar o índice da Mediana (*Median Index*) (abordagem regular feita no mercado imobiliário). Os resultados são apresentados na Figura 4 e Figura 5.

A partir das Figuras 4 e 5, pode-se destacar que o uso do FHPI, ao invés do Índice da Mediana, aponta, em geral, para menores variações de preços; muito provavelmente por se estar comparando residências mais comparáveis, dado que é feito o controle pelas características estruturais. Por exemplo, o Índice da Mediana chega a magnitude de 3,20 no ano de 2008. Já o índice FHPI chega a esse montante apenas em 2011. Este é um bom resultado para o quê este trabalho se propõe (estimar um índice de preços mais preciso no sentido de captar variações de preço e não modificações de atributos), uma vez que esse resultado indica que o índice baseado na mediana superestima as variações de preços, o que ocorre porque este índice não compara residências semelhantes, como o índice FHPI o faz. Ou seja, o índice de preços baseado na mediana mede tanto variações de preços, como variações na composição dos imóveis (variação nas características), o que leva a uma ampliação na variação de preço captada; tal evidência de captar mudança na composição dos imóveis é também consistente com o fato de que os construtores tendem a ofertar residências com características mais valorizadas com o passar dos anos, a fim de agregar maior valor aos imóveis.

O uso do FHPI, ao invés do Índice da Mediana, aponta diferenças entre o *ranking* das RPAs com maiores variações: enquanto no segundo, o índice que utiliza a mediana, a RPA 2 é aquela com maior elevação de preços; para o índice FHPI tal posto é ocupado pela RPA 6. Ou seja, quando se compara imóveis com as mesmas características, obtém-se significativas diferenças, sendo as RPAs 6, 3 e 4 as que apresentaram maiores elevações de preços. Como tais RPAs incluem as áreas de maior valorização da cidade (mais bem servidas de infraestrutura domiciliar, por exemplo), este resultado é consistente com o movimento de adensamento dos bairros destas RPAs entre 2000 e 2010 apontado por Barbosa & Silveira Neto (2015).

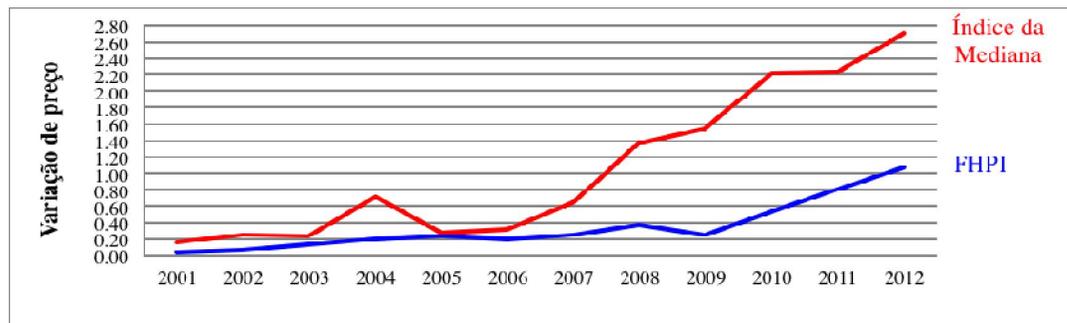
O índice FHPI também tende a indicar menor variabilidade de preços entre as RPAs. Este também é um resultado a ser destacado. Como, em geral, as vizinhanças apresentam imóveis mais semelhantes, o *matching* feito para obter o banco de dados usado para estimar o FHPI tendem a eliminar diferenças de variações de preços entre vizinhanças associadas a variações das características dos imóveis ao longo do tempo. Assim, as variações de preços



**Figura 4. A trajetória dos índices**

Fonte: Elaboração própria a partir de dados do ITBI.

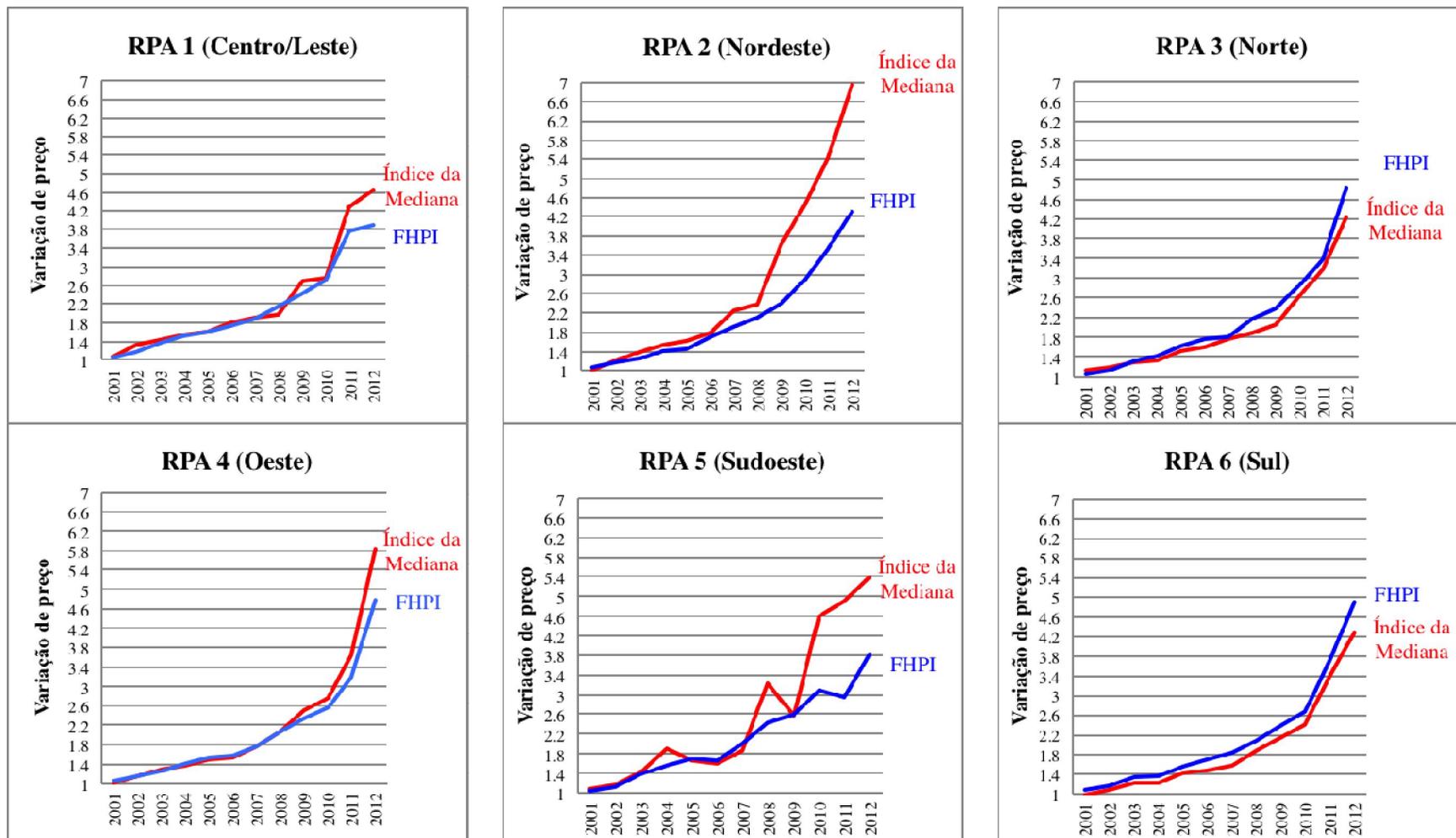
Nota: *Range* é a diferença entre os índices das RPAs a partir do maior valor e o menor valor registrado no ano. E o separador decimal usado foi o “.”.



**Figura 5. Comparativo dos ranges**

Fonte: Elaboração própria a partir de dados do ITBI.

Nota: O separador decimal usado foi o “.”.



**Figura 6. Trajetória dos índices de preços em cada RPA**

Fonte: Elaboração própria a partir de dados do ITBI.

Nota: O separador decimal usado foi o “.”.

das diferentes RPAs obtidas a partir deste índice tende a refletir com mais intensidade movimentos gerais dos preços de imóveis presentes na cidade I.

Independentemente do índice utilizado, há tendência de aumento da variância dos índices ao longo do tempo. Como a referência é a data de 2000, isto, em parte, pode ser um efeito estatístico: há possibilidade de maior variância entre RPAs quando o período de tempo é maior. De toda a forma, impressiona a forte tendência de variação entre as RPAs no período de 2008 e 2012, período de forte expansão do crédito imobiliário, que pode ser parte da explicação, dado que há a possibilidade de maior pressão de demanda e de preços em RPAs com características mais preferidas pelos consumidores, quando se aumenta o volume de crédito disponível.

A partir da Figura 6, são apresentadas as trajetórias dos índices em cada RPA. É possível identificar três grupos de RPAs com comportamentos distintos: RPAs 1 e 4, índices com comportamentos similares e com índice da mediana mais alto, RPAs 2 e 5, com índice da mediana também mais alto, mas agora com trajetória bem mais distinta daquela verificada para o FHPI, e RPAs 3 e 6, com índice da mediana mais baixo que aqueles dos FHPI. Este resultado ressalta mais uma vez a importância do índice FHPI estimado.

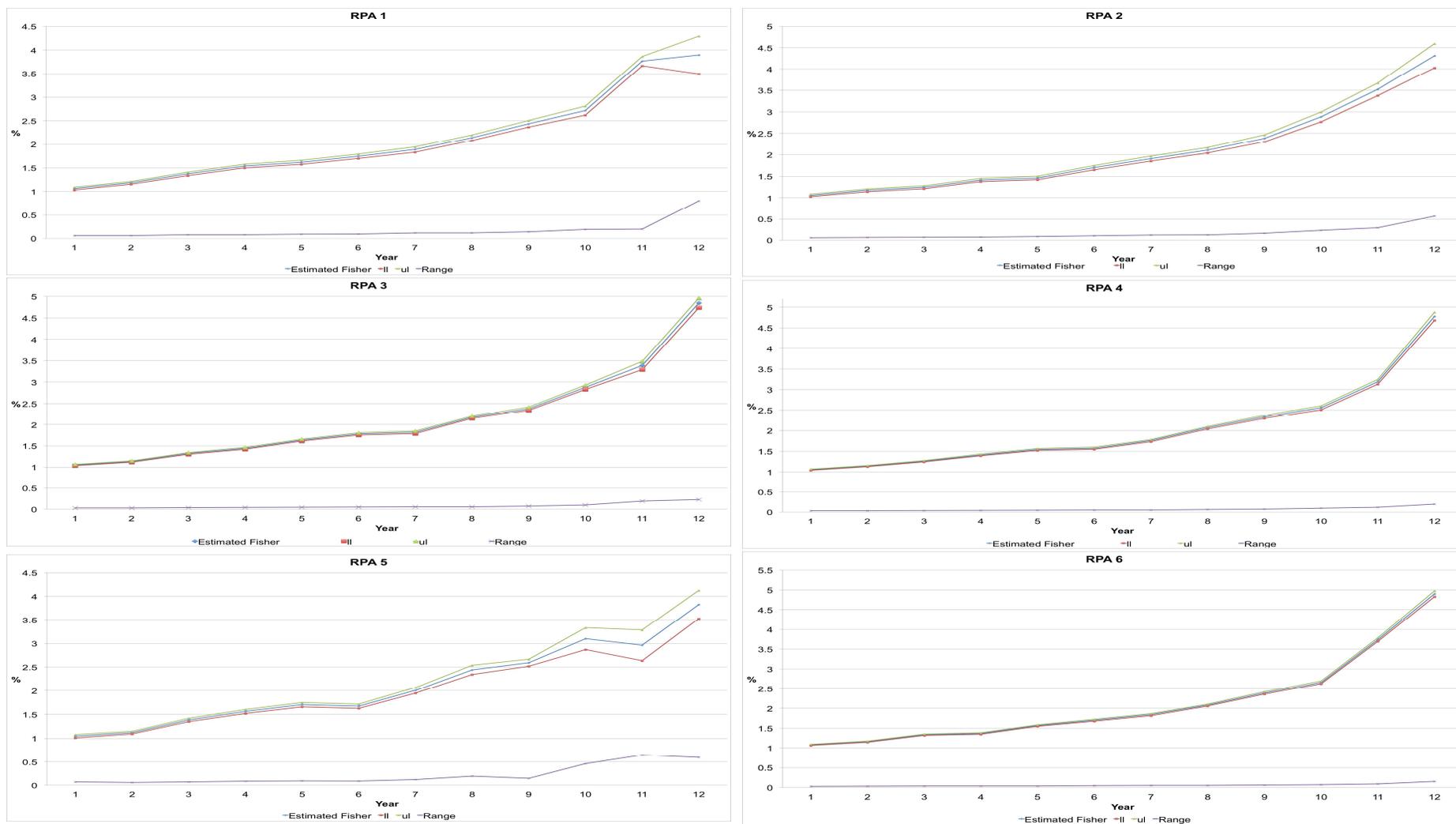
Tais evidências indicam que o controle para as características dos imóveis nas RPAs 1 (Centro) e 4 (Oeste) é muito menos importante que para as demais RPAs, o que é consistente com o fato de que os imóveis negociados na região central apresentam menor variação em suas características ao longo dos anos avaliados. Isso pode ser verificado nas Tabelas 4 e 7, no anexo, respectivamente, as quais apresentam os comportamentos médios das variáveis nas RPAs para cada ano do período estudado (2000 - 2012).

Para as RPAs 2 (Nordeste) e 5 (Sudoeste), a consideração das características dos imóveis é fundamental para a mensuração precisa das variações de preços dos imóveis. Os resultados indicam que, para estas duas RPAs, há superestimação da variação dos preços quando tais características não são consideradas. Tal fato pode estar associado a características melhores dos imóveis recentemente construídos e negociados nestas RPAs, onde habitam, em geral, indivíduos de rendas abaixo da média da cidade. Esse comportamento médio das variáveis nas RPAs 2 e 5 é verificado nas Tabelas 5 e 8, no anexo.

Para as RPAs 3 (Norte) e 6 (Sul), a consideração das características intrínsecas faz com que as variações de preços registradas pelo índice FHPI sejam superiores às obtidas com o índice da mediana. Os maiores valores do índice FHPI em relação ao índice da mediana sugerem que os imóveis negociados no período mais recente foram imóveis relativamente mais simples que aqueles dos períodos passados (ver Tabelas 6 e 9, no anexo), uma explicação que é consistente com ocupação das localidades melhor localizadas, por exemplo, mais próximas à praia (de mais alto padrão) em períodos passados.

Os resultados do conjunto dos dados do Recife mostraram que o método tradicional (através, por exemplo, da mediana, ou de um HPI sem *matching*) para HPI fornece uma imagem que pode não refletir a realidade analisada do caminho dos preços da habitação (no caso analisado apartamentos), enfatizando que quando se busca estimar um HPI deve-se incluir características intrínsecas e amenidades como co-variáveis para que não se registre a alteração de características que interferem no preço como modificação de preço. Também é possível destacar que cada região no Recife tem uma dinâmica específica.

É importante apontar que o índice hedônico estimado, além de relevante do ponto de vista analítico, é também estatisticamente significativo. De fato, a obtenção de intervalos de confiança através de *bootstrap* indica que o FHPI estimado para cada RPA tem.



**Figura 7. Trajetória e intervalo de confiança do FHPI**

Fonte: Elaboração própria a partir de dados do ITBI.

Nota 1: O separador decimal usado foi o “.”.

Nota 2: Linha contínua azul = trajetória do FHPI básico, ll = *low limit* = limite inferior (linha tracejada vermelha) e ul = *up limit* (linha tracejada verde).

Nota 3: *Range* é a diferença entre o limite inferior e o limite superior.

Nota 4: No Eixo horizontal o número 1 corresponde ao ano de 2001, e assim por diante, até 12 que correspondendo a 2012.

significância estatística para os valores obtidos, dado que os índices estimado estão dentro dos seus respectivos intervalos de confiança. Conforme pode ser visto na Figura 7.

Tais resultados corroboram com a ideia de que um índice FHPI com mais covariáveis parece ser importante quando se busca obter a trajetória do preço mais próximo com o mundo real, no sentido de que reflita mudança em precificação e não mudança de atributos. Ou seja, as características intrínsecas trabalham no sentido de que se possa ter um índice FHPI mais próximo à variação dos preços efetivamente ocorrida em cada localidade.

#### **4 Conclusões**

Com o objetivo de entender a dinâmica de preço dos imóveis na Cidade do Recife entre 2000 e 2012, este trabalho perseguiu o objetivo específico de (I) estimação de um índice de preços hedônicos para a Cidade do Recife, para a sua subdivisão geográfica de RPA, por meio de uma estrutura de cálculo que permitisse o uso de características estruturais. Tudo isso só tornou-se possível com o uso da base de dados do ITBI, da Prefeitura da Cidade do Recife. Os quais forneceram as informações das características estruturais utilizadas no modelo de preços hedônicos.

Com essas informações foi possível propor um índice de preços dos apartamentos (o que se denominou de FHPI) incluindo variáveis estruturais; contribuindo, assim, para melhor precisão quanto à captação da variação de preço. Dado que, ao aplicar um processo de emparelhamento na base de dados usada para o cálculo do índice, tomando como referência características intrínsecas, tem-se como objetivo obter imóveis mais comparáveis ao longo dos anos, em virtude de não existir venda repetida do mesmo imóvel em cada período de tempo. Garantindo, assim, que se use imóveis semelhantes ao longo do período de tempo analisado. Logo, a variação captada pelo índice não será decorrente de alterações quanto a tipos de atributos presentes nos imóveis, sendo mais preciso quanto a este aspecto. Isso tudo vai além da estrutura de índice de preços de imóveis baseado em medidas de tendência central (por exemplo, média e mediana), comumente usada no mercado da habitação, e possibilita que as mudanças de preço aferidas sejam de fato decorrentes de alterações na precificação do imóvel e não sofram influência de variações nas características consideradas.

A partir do comportamento dos índices estimados (Índice da Mediana (como o mercado em geral faz) e o FHPI (com características estruturais)) ao nível de Regiões Políticas Administrativas (RPAs) foi possível verificar que o comportamento dos preços dos imóveis na cidade do Recife não tem um padrão. Algo que já se acreditava ser factível de encontrar, foi encontrado pelo processo de inferência realizado. As evidências indicam que nas RPAs 1 (Centro) e 4 (Oeste) o controle para as características dos imóveis é bem relevante ao estimar a variação dos preços. Nas RPAs 2 (Nordeste) e 5 (Sudoeste), a consideração das características dos imóveis é fundamental para a mensuração precisa das variações de preços dos imóveis. E por fim, nas RPAs 3 (Norte) e 6 (Sul) as evidências sugerem que os imóveis negociados no período mais recente foram imóveis relativamente mais simples que aqueles dos períodos passados. Logo, constata-se que há modificações importantes de valores que ocorreram na cidade do Recife e que o índice de preço mediano dependendo da localidade não refletiu.

Portanto, as medidas comumente utilizadas, quando consideradas em relação aos valores dos imóveis, não permitem medir adequadamente as variações dos valores reais da economia, dado que não fazem uso de amostras emparelhadas e variáveis estruturais. Os resultados obtidos permitem obter um maior conhecimento sobre a dinâmica do mercado imobiliário em diferentes regiões da cidade do Recife.

O nível de detalhamento da dinâmica de preços dos imóveis na Cidade do Recife, obtido com a proposição de um índice de preços dos imóveis muito mais preciso quanto à captação do que de fato é variação de preço (retirando-se o que consiste em variações de

preços resultantes de modificações em características estruturais) e com a decomposição do diferencial de preço, pode nortear tanto a esfera pública, na concepção de políticas públicas, como a esfera privada, que passa a conhecer fatores conectados com a natureza do mercado de imóveis do Recife e pode orientar suas decisões em algo que representa um dispêndio de quase um terço da renda da família, o imóvel (a decisão de onde morar).

## Referências

ALMEIDA, A.N.; AZZONI, C.R., 2013. *Custo de vida comparativo das regiões metropolitanas brasileiras: 1996–2012*. TDNereus 11, São Paulo.

BACEN, 2015. Indicadores econômicos consolidados. Disponível em: <http://www.bcb.gov.br/?INDECO>. Acessado em: junho de 2015.

BARBOSA, M.R., SILVEIRA NETO, R.M., 2015. *Condicionantes da mobilidade urbana: uma análise empírica para a região metropolitana do Recife*. XIII ENABER, Curitiba (PR).

BRUECKNER, J.K., 2011. *Lectures on Urban Economics*, MIT Press.

COCHRAN, W. G.; RUBIN, D. B., 1973. Controlling bias in observational studies: a review. *Sankhya: The Indian Journal of Statistics*, v.Series A, n.35, p.417-446.

COELHO JUNIOR. A. F.; LOPEZ, E.; SILVEIRA NETO, R.M., 2014. *Neighborhood price index: the case of Recife, Brazil*, 61<sup>st</sup> NARSC, Washington D.C..

COELHO JÚNIOR. A. F.; SILVEIRA NETO, R.M., 2015a. *Variação dos preços de imóveis nas cidades brasileiras: valoração de mercado, características intrínsecas ou amenidades? uma análise para o caso da cidade do Recife*, 43<sup>o</sup> Encontro Nacional de Economia – Anpec, Florianópolis (SC).

COELHO JÚNIOR & SILVEIRA NETO, 2015B *Dinâmica dos preços dos imóveis no mercado formal de residências da cidade do Recife: um estudo de sua evolução e de seus determinantes no período 2000-2012*. 132f. Tese (Programa de Doutorado em Economia). UFPE.

COX, D. R., 1970. *The analysis of binary data*. London: Methuen.

DIEWERT, W. E., 1978. Superlative index numbers and consistency in aggregation. *Econometrica*, v.46, n.4, p.883-900.

EFRON, B., 1979. *Bootstrap methods: another look at the jackknife*. *Ann. Stat.*, v.7, p.1-26.

EFRON, B.; TIBSHIRANI, R.J., 1993. *An introduction to the bootstrap*, Chapman & Hall.

GRILICHES, Z., 1971. *Hedonic price indices for automobiles: an econometric analysis of quality change*. Cambridge, Massachusetts.

LONFORD, N., 2009. *A house price index defined in the potential outcomes framework*. Working paper. Universitat Pompeu Fabra.

LOPEZ, E.; AROCA, P., 2012. Estimación de la inflación regional de precios de la vivienda en Chile. *El Trimestre Económico*, v.79, n.315.

LOPEZ, E.; DU, M. 2014. *Estimating housing price indices for 10 Illinois*.

*Metropolitan Statistical Areas (MSAs)*. Mimeografado. Disponível em: <<http://www.real.illinois.edu/d-paper/14/14-T-1.pdf>>.

MAYO, S. K., 1981. Theory and Estimation in the Economics of Housing Demand. *Journal of Urban Economics*. v.10, n.1, p.95-116.

MCMILLEN, D. P., 2012. Repeated sales as a matching estimator. *Real Estate Economics*. v. 40, n.4, p.745-773.

MEESE, R.; WALLACE, N., 1991. Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *Real Estate Economics*. v.19, n. 3, p.1540-6229.

PAREDES, D.; AROCA, P., 2008. Metodología para estimar un índice regional de costo de vivienda en Chile. *Latin American Journal of Economics-formerly Cuadernos de Economía*, v.45, n.131, p.129-143.

PAREDES, D., 2011. A methodology to compute regional housing price index using matching estimator methods. *The Annals of Regional Science*, v.46, p.139-157.

PREFEITURA MUNICIPAL DO RECIFE, 2014. A cidade  
Disponíveis em: <<http://www2.recife.pe.gov.br/a-cidade/aspectos-gerais/>>  
Acessado em: junho de 2014.

ROSEN, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, v.72, p.34-55.

ROSENBAUM, P.R.; RUBIN, D.B, 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*. v.39, n.1.

SEABRA, D. M.; SILVEIRA NETO, R. M., 2014. *Amenidades urbanas e valor das residências*: Uma análise empírica para a cidade do Recife, 42º Encontro Nacional de Economia, Natal.

SECOVI, 2013. *Indicadores*. Disponível em: <http://www.secovi.com.br/pesquisas-e-indices/indicadores-do-mercado/> Acessado em: dezembro de 2014.

SIRMANS, S. G.; MACPHERSON D. A.; ZIETZ, E. N., 2005. The composition of hedonic pricing models. *Journal of Real Estate Literature*, v.13, n.1, p.1-44.

## Anexo

**Tabela 4. RPA 1 - Comportamento Médio das Variáveis Estruturais (2000 - 2012)**

		Variáveis Estruturais				
		<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
Ano	2000	6.15	28.70	87.36	0.22	5.67
	2001	6.60	29.23	82.72	0.25	5.63
	2002	6.79	29.61	89.04	0.24	5.33
	2003	6.31	29.64	81.89	0.29	6.09
	2004	6.24	30.01	79.31	0.27	6.07
	2005	6.52	31.70	85.14	0.23	5.81
	2006	6.15	32.95	84.33	0.27	5.78
	2007	5.86	33.70	83.49	0.22	5.60
	2008	6.10	33.28	80.09	0.25	5.93
	2009	7.03	31.08	87.50	0.37	5.73
	2010	6.16	33.42	81.71	0.24	6.95
	2011	7.20	29.05	91.21	0.41	5.04
	2012	7.25	33.08	89.06	0.35	5.62

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota: O separador decimal usado foi o “.”.

**Tabela 5. RPA 2 - Comportamento Médio das Variáveis Estruturais (2000 - 2012)**

		Variáveis Estruturais				
		<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
Ano	2000	4.34	16.50	98.41	0.41	3.68
	2001	4.43	15.73	94.09	0.50	3.75
	2002	4.65	14.33	99.34	0.54	3.72
	2003	4.63	13.62	88.01	0.60	3.71
	2004	5.04	15.13	101.03	0.48	3.41
	2005	5.06	16.32	97.36	0.55	3.29
	2006	5.61	15.78	98.12	0.63	3.43
	2007	7.17	14.28	93.22	0.68	3.52
	2008	6.88	14.85	94.36	0.68	3.43
	2009	6.73	11.93	89.76	0.75	3.24
	2010	9.93	8.85	88.83	0.83	3.21
	2011	10.30	9.14	86.79	0.84	3.37
	2012	9.48	10.93	85.74	0.80	3.35

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota: O separador decimal usado foi o “.”.

**Tabela 6. RPA 3 - Comportamento Médio das Variáveis Estruturais (2000 - 2012)**

		Variáveis Estruturais				
		<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
Ano	2000	7.70	11.38	173.95	0.80	2.58
	2001	8.63	10.29	171.29	0.82	2.65
	2002	8.57	10.40	156.54	0.82	2.83
	2003	8.03	11.67	158.93	0.83	2.72
	2004	8.63	11.58	154.13	0.83	2.81
	2005	9.15	11.04	153.60	0.84	2.73
	2006	8.76	11.78	146.95	0.84	2.77
	2007	9.33	11.21	147.20	0.86	2.71
	2008	9.15	11.94	144.02	0.88	2.67
	2009	9.34	10.83	136.57	0.88	2.78
	2010	9.24	10.97	138.57	0.90	2.71
	2011	9.55	9.97	121.18	0.91	2.78
	2012	9.53	10.24	116.82	0.90	2.90

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota: O separador decimal usado foi o “.”.

**Tabela 7. RPA 4 - Comportamento Médio das Variáveis Estruturais (2000 - 2012)**

		Variáveis Estruturais				
		<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
Ano	2000	4.46	11.71	101.80	0.59	4.01
	2001	4.17	10.24	94.66	0.69	3.95
	2002	4.22	10.10	95.96	0.70	4.15
	2003	4.74	11.13	101.81	0.69	4.45
	2004	4.77	11.87	101.53	0.63	3.98
	2005	5.10	11.02	98.51	0.62	3.88
	2006	5.34	12.20	95.07	0.67	3.98
	2007	5.39	12.89	96.65	0.70	4.14
	2008	5.88	10.56	93.03	0.76	3.68
	2009	6.79	11.17	98.98	0.77	3.73
	2010	6.18	10.94	91.01	0.76	3.73
	2011	6.71	9.75	84.76	0.81	3.97
	2012	7.47	11.03	90.09	0.81	3.61

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota 1: O separador decimal usado foi o “.”.

**Tabela 8. RPA 5 - Comportamento Médio das Variáveis Estruturais (2000 - 2012)**

		Variáveis Estruturais				
		<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
Ano	2000	2.53	14.53	75.76	0.30	11.80
	2001	2.67	15.27	77.04	0.26	11.51
	2002	2.37	17.78	76.33	0.23	12.34
	2003	3.43	10.00	78.56	0.54	9.15
	2004	3.63	9.03	79.95	0.62	8.00
	2005	3.61	10.60	72.78	0.57	8.75
	2006	3.05	13.37	69.62	0.50	10.64
	2007	3.73	14.09	72.17	0.51	9.29
	2008	3.98	11.07	73.16	0.65	7.65
	2009	3.31	15.23	69.99	0.53	9.38
	2010	5.20	7.93	68.13	0.78	6.22
	2011	4.26	9.09	65.82	0.76	6.33
	2012	4.40	6.01	61.73	0.86	5.64

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota 1: O separador decimal usado foi o “.”.

**Tabela 9. RPA 6 - Comportamento Médio das Variáveis Estruturais (2000 - 2012)**

		Variáveis Estruturais				
		<i>floor</i>	<i>age</i>	<i>private_bu</i>	<i>standard_construct</i>	<i>unit_den</i>
Ano	2000	7.46	12.74	145.45	0.84	3.70
	2001	7.07	14.06	133.64	0.81	4.07
	2002	7.27	14.06	138.21	0.81	4.07
	2003	7.30	14.34	130.96	0.80	4.17
	2004	7.61	13.43	126.42	0.79	4.06
	2005	7.58	13.65	125.87	0.83	3.78
	2006	7.69	15.18	124.72	0.83	3.69
	2007	7.68	15.92	122.84	0.82	3.67
	2008	8.33	14.84	124.56	0.86	3.44
	2009	8.67	14.62	124.21	0.88	3.27
	2010	8.71	13.52	123.01	0.90	3.39
	2011	9.21	13.15	120.55	0.90	3.40
	2012	9.81	11.70	111.65	0.92	4.07

Fonte: Elaboração própria a partir dos dados do ITBI.

Nota 1: O separador decimal usado foi o “.”.