

**UNIVERSIDADE FEDERAL DO RIO GRANDE – FURG  
INSTITUTO DE CIÊNCIAS ECONÔMICAS, ADMINISTRATIVAS E  
CONTÁBEIS – ICEAC  
CURSO DE CIÊNCIAS ECONÔMICAS**

**MARCUS VINICIUS BASTOS DOS SANTOS**

**UMA AVALIAÇÃO DA CAPACIDADE DO GOOGLE TRENDS EM AUXILIAR NA  
PREVISÃO DA DEMANDA POR IMÓVEIS NO BRASIL**

**Rio Grande  
2017**

Marcus Vinicius Bastos dos Santos

**UMA AVALIAÇÃO DA CAPACIDADE DO GOOGLE TRENDS EM AUXILIAR NA  
PREVISÃO DA DEMANDA POR IMÓVEIS NO BRASIL**

Monografia apresentada como requisito parcial para obtenção de título de Bacharel pelo curso de Ciências Econômicas da Universidade Federal do Rio Grande.

Orientador: Prof. Dr. Cristiano Aguiar de Oliveira.

**Rio Grande  
2017**

## **AGRADECIMENTOS**

Quero agradecer primeiramente a Deus e a meus pais, Célia e Celso, que me proporcionaram todo o suporte necessário para que eu pudesse me empenhar durante todo o curso e especialmente nesta etapa final. Também minha mãe, que sempre com um olhar crítico me ajudou a aprimorar o trabalho imensamente.

Necessário agradecer ainda a todos os professores que fizeram parte dessa caminhada, nos ensinando e passando seus conhecimentos da melhor maneira possível. Em especial, preciso agradecer ao professor Cristiano Oliveira, que além de excelente professor, foi meu orientador e precursor da ideia dessa monografia, me ajudando do início ao fim para que o trabalho tivesse sucesso.

Devo agradecer também a todos os amigos que fiz durante a graduação. Amigos que sempre estiveram presentes nos momentos bons e ruins, e me incentivaram para chegar até aqui. Muito obrigado pelos grupos de estudo, trabalhos e conversas que com certeza fizeram esses anos inesquecíveis.

Finalmente, muito obrigado a todos que participaram de alguma forma para a conclusão do curso e deste trabalho.

*“Life can only be understood backwards; but it must be lived forwards.”*

Søren Kierkegaard

## LISTA DE ILUSTRAÇÕES

<i>GRÁFICO 1: Resposta de <math>d</math> a um choque em <math>s</math>, com intervalo de confiança bootstrap a 90%.</i>	39
<i>GRÁFICO 2: Resposta de <math>d</math> a um choque em <math>u</math>, com intervalo de confiança bootstrap a 90%.</i>	41
<i>GRÁFICO 3: Resposta de <math>d</math> a um choque em <math>w</math>, com intervalo de confiança bootstrap a 90%.</i>	41
<i>GRÁFICO 4: Resposta de <math>d</math> a um choque em <math>i</math>, com intervalo de confiança bootstrap a 90%.</i>	42

## LISTA DE TABELAS

<i>TABELA 1: Fonte dos dados CBIC</i>	30
<i>TABELA 2: P-valor do teste ADF</i>	34
<i>TABELA 3: Estatística do teste do traço, teste do autovalor máximo e teste do traço corrigido para o tamanho da amostra, respectivamente</i>	35
<i>TABELA 4: Estatísticas do modelo VECM</i>	37
<i>TABELA 5: Respostas a um choque de um erro padrão em <math>s</math></i>	40
<i>TABELA 6: Respostas de <math>d</math> ao choque de um erro padrão em <math>u, w, i</math> e <math>d</math></i>	43
<i>TABELA 7: Decomposição do erro da previsão para <math>d</math></i>	44
<i>TABELA 8: Estatísticas de avaliação da previsão</i>	45
<i>TABELA 9: Causalidade de Granger</i>	47

## APÊNDICE

TABELA A 1: Estatísticas descritivas das séries	57
GRÁFICO A. 1: Taxa de desemprego $u$ .	58
GRÁFICO A. 2: Nível de salários $w$ .	58
GRÁFICO A. 3: Taxa de juros $i$ .	59
GRÁFICO A. 4: Volume de buscas $s$ .	59
GRÁFICO A. 5: Demanda imobiliária $d$ .	60

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>10</b>
<b>2 REVISÃO DE LITERATURA</b>	<b>14</b>
<b>3 REFERENCIAL TEÓRICO</b>	<b>16</b>
<b>4 METODOLOGIA</b>	<b>18</b>
4.1 ESTACIONARIEDADE DAS SÉRIES	19
4.2 ANÁLISE DE COINTEGRAÇÃO	20
4.3 ESPECIFICAÇÃO: VAR E VECM	22
4.3.1 AUTOCORRELAÇÃO E NORMALIDADE DOS RESÍDUOS	23
4.4 FUNÇÃO IMPULSO-RESPOSTA	24
4.5 DECOMPOSIÇÃO DA VARIÂNCIA	25
4.6 CAUSALIDADE DE GRANGER	26
<b>5 BASE DE DADOS</b>	<b>29</b>
5.1 DEMANDA POR IMÓVEIS	29
5.2 GOOGLE TRENDS	30
5.3 DADOS MACROECONÔMICOS	32
<b>6 RESULTADOS</b>	<b>34</b>
6.1 ANÁLISE DE ESTACIONARIEDADE	34
6.2 ANÁLISE DE COINTEGRAÇÃO	35
6.3 ESPECIFICAÇÃO: VECM	36



6.3.1 AUTOCORRELAÇÃO E NORMALIDADE DOS RESÍDUOS	37
6.4 FUNÇÃO IMPULSO-RESPOSTA	38
6.5 DECOMPOSIÇÃO DA VARIÂNCIA	44
6.6 CAUSALIDADE DE GRANGER	45
<b>7 CONSIDERAÇÕES FINAIS</b>	<b>49</b>
<b>BIBLIOGRAFIA</b>	<b>50</b>
<b>APÊNDICE</b>	<b>57</b>

## RESUMO

O trabalho visa estimar um modelo econométrico com o intuito de avaliar a capacidade do Google Trends em auxiliar na previsão da demanda por imóveis no Brasil. Com isso, foi definida uma metodologia de série temporal através de um modelo autorregressivo (VAR). Para compor o modelo foi utilizado 5 variáveis (buscas na internet, demanda imobiliária, taxa de juros, taxa de desemprego e nível de salários) durante o período de janeiro de 2004 até dezembro de 2016, totalizando uma amostra com 145 observações mensais. As estimativas apontam que existe uma relação positiva entre o volume de pesquisas e a demanda por imóveis. Como principal contribuição, este artigo aponta que as buscas na internet apresentam informações relevantes na análise da demanda imobiliária brasileira, apesar de conter limitações.

Palavras Chave: Google Trends, Mercado Imobiliário, Série Temporal, Vetor Correção de Erros (VECM).

## **ABSTRACT**

This work aims to estimate an econometric model to evaluate Google Trends capacity in predicting real state demand in Brazil. To do so, has been defined a time series methodology through a vector autoregressive model (VAR). To compose the model 5 variables were used (internet searches, housing sales, interest rate, unemployment rate and wages) during the period of january 2004 to december 2016, totalizing 145 monthly observations sample. Estimates suggests a positive relationship between the searches volume and the housing demand. As main contribution, this article shows that internet searches hold relevant information on the analysis of brazilian real state demand, despite having limitations.

Key words: Google Trends, Housing Market, Time Series, Vector Error Correction Model (VECM).

## 1 INTRODUÇÃO

Atualmente, a velocidade da informação é algo notório e a maneira como os agentes a ofertam e a demandam é também singular. O motivo mais forte dessa afirmação é o advento da internet e a infinidade de recursos e opções que ela possibilita. De acordo com o IBGE, só no Brasil, mais de 50% da população usa a internet e tem acesso a seus recursos. Um dos recursos mais abrangentes são as ferramentas de buscas, como Google, Yahoo, Bing etc. que provêm retorno para quase que qualquer termo pesquisado. Conforme Guzman (2011), a ubiquidade de ferramentas online disponíveis nos permite ter muitas informações sobre nós mesmos, como nossas opiniões e expectativas. Dados sobre essas informações podem mostrar características do comportamento humano que antes eram desconhecidas.

Em relação a dados, o Google disponibiliza uma ferramenta que provê em tempo real, diariamente e semanalmente o volume de qualquer palavra ou expressão que usuários pesquisam no Google em dada localidade, o Google Trends (CHOI; VARIAN, 2011). Essa vasta fonte de informação pode ser útil para analisar diversos assuntos. Como por exemplo, pesquisar o termo “televisão de tela plana” ou “geladeira”, serviria de parâmetro para analisar a demanda por bens duráveis através do tempo (MCLAREN; SHANBHOGE, 2011).

Segundo Preis et al. (2010), os dados encontrados no Google Trends trazem *insights* em nossa vida econômica em diferentes escalas. Um crescente número de usuários da internet utiliza mecanismos de pesquisa. Cada termo pesquisado pode ser visto como um voto, assim, cada termo traz informações sobre nossos interesses na menor escala possível, que são os desejos individuais. Em grande escala, nossos interesses, quando expostos em pesquisas online, revelam tendências que afetam a economia.

Para Goel et al. (2010), com o aumento do volume de buscas na internet a procura de notícias, pesquisas e informação, é intrigante o fato de se poder observar, a qualquer momento no tempo, sobre o que a consciência coletiva da população está preocupada, ou quais são seus interesses e intenções. Também segundo Goel et al. (2010), partindo desse pressuposto, é um pequeno passo concluir que, o que as pessoas estão buscando hoje, é uma predição do que elas irão fazer no futuro próximo.

No caso dos Estados Unidos, Webb (2009) conduziu um experimento para analisar se os dados sobre as pesquisas na internet podiam ser utilizados para indicar tendências reais sobre a inadimplência, e consequentes despejos no mercado hipotecário norte americano. De acordo com seus resultados, foi possível perceber que o volume de pesquisas sobre o termo “despejo” é altamente correlacionado com o número real de despejos ocorridos. Isso traz uma nova perspectiva para a análise de mercados imobiliários em geral.

Geralmente, o governo ou empresas liberam dados e indicadores econômicos com substancial defasagem do período presente (WU et al., 2009), podendo chegar até dois ou três meses no caso de dados sobre o mercado imobiliário. No âmbito desta monografia, busca-se averiguar se pesquisas na internet podem ser consideradas um indicador antecedente de vendas de imóveis no Brasil, uma vez que, como ilustrado por Choie Varian (2009), não só é possível prever o volume de vendas no setor imobiliário, como também no setor automotivo e no varejo, além de também prever níveis de desemprego e os destinos de viagens mais visitados.

No caso do mercado imobiliário, a formação dos preços é obtida através da teoria dos preços hedônicos, na qual os imóveis são um conjunto de características físicas e espaciais. Dada a complexidade envolvida na análise do melhor imóvel, as pesquisas na internet funcionam como um filtro, diminuindo de maneira significativa o tempo e os custos envolvidos nas buscas por um imóvel. Dessa forma, infere-se que teoricamente as pessoas pesquisem pelos imóveis que tem interesse previamente pela internet, e depois de filtrar seus resultados partam para as visitas literais, fora do mundo virtual. Ainda sobre isso, Guzman (2009) coloca que o comportamento em buscas online pode ser interpretado como uma expectativa revelada dos agentes, ou seja, presume-se que os indivíduos pesquisem sobre assuntos que queiram aprender mais, ou sobre assuntos que lhes causem

preocupações, como por exemplo, o que pode significar os sintomas de mal-estar que estejam sentindo.

Para este trabalho, é coerente esperar que os agentes busquem por termos relacionados ao mercado imobiliário, como “casas à venda”, “imóveis na planta”, “imobiliárias”, “terrenos”, “kitnets”, e assim por diante. A busca por tais temas sugere que os indivíduos estejam desejosos de adquirir um imóvel no período seguinte.

Na literatura internacional, o tema é bastante contemporâneo e para literatura brasileira, trata-se de uma abordagem inovadora. Segundo dados da Câmara Brasileira da Indústria da Construção (CBIC), em dezembro de 2016 o Brasil teve mais de 2 bilhões de reais financiados para a aquisição de imóveis com recursos do Sistema Brasileiro de Poupança e Empréstimo (SBPE). Dada a importância do mercado para o país – através do grande volume de investimento e o alto custo de projetos imobiliários envolvidos – observar a formação da demanda, e até mesmo prever como ela se dará, é de grande valia para tomada de decisões por parte de compradores, mas principalmente vendedores.

Dessa forma, o objetivo geral deste trabalho é avaliar a previsibilidade de um Vetor Autorregressivo (VAR) para a demanda no mercado imobiliário brasileiro através da relação entre as buscas agregadas feitas na internet voltadas para o setor de imóveis, providas pelo Google Trends, e o volume de vendas efetivo do mercado. Além dessas variáveis, foram também incluídos dados macroeconômicos que possam juntamente influenciar a demanda por imóveis no Brasil.

Além dessa introdução, o trabalho está dividido em mais 5 seções, na qual a segunda aborda uma revisão literária exemplificando a utilidade do Google Trends, a terceira traz um referencial teórico acerca do consumo de bens imóveis, a quarta apresenta a metodologia e a base de dados empregadas, a quinta expõe os resultados encontrados e por fim as considerações finais são apresentadas.

## 2 REVISÃO DE LITERATURA

O uso do Google Trends para analisar e até prever determinadas tendências de mercado é observado nos estudos de diversos autores, tais como Lallanilla (2013), Goel et al. (2010), McLaren e Shanbhoge (2011), Silver (2015), Web (2009), DeVera et al. (2016), Giannone et al. (2006). Um dos autores precursores nesse tema é o economista chefe do Google, Hal Varian em conjunto com Hyunyoung Choi, que avaliaram a relação entre o nível de buscas online com indicadores econômicos, como o desemprego, volume de vendas de carros e imóveis, e estatísticas sobre o destino de viagens internacionalmente (CHOI; VARIAN, 2009).

O próprio Google analisa os dados sobre pesquisas feitas por termos relacionados com a gripe *influenza* e com a dengue, os chamados *Flu Trends* e *Dengue Trends*, que visavam estimar a proliferação das doenças em tempo real<sup>1</sup>.

D'Amuri e Marcucci (2009) concluem que o índice provido pelo Google é eficiente para prever o nível de desemprego nos EUA. Os modelos criados por eles através do índice tiveram resultados positivos para prever as taxas de desemprego por um, dois e três meses adiante. Segundo os autores, o uso dos dados das buscas online, deverá se tornar muito popular em pesquisas econômicas em um futuro próximo.

Schmidt e Vosen (2009) propuseram a utilização do Google Trends para prever o consumo privado. Através de seus resultados, descobriram que existe uma relação entre as pesquisas e a maneira que os consumidores alocam seu capital. Com isso concluíram que os dados das tendências online fornecem modelos mais precisos.

---

<sup>1</sup> Para mais informações, consultar <http://www.google.org/flutrends/about/>.

Preis et al. (2010) indagaram se havia alguma relação entre volume de pesquisas feitas no Google e as flutuações do mercado financeiro. De acordo com seus estudos, puderam perceber clara evidência que o volume de transações semanais das empresas presentes no S&P 500 estava correlacionado com o volume de pesquisas semanais sobre o nome das respectivas empresas.

Guzmán (2011) supõe uma forma de medir, em tempo real, as expectativas sobre a inflação, baseada em metadados (dados sobre dados), partindo de termos pesquisados no Google. Comparando o Índice de Busca por Inflação do Google (*Google Inflation Search Index – GISI*) com outros 37 índices regulares de expectativa sobre a inflação, ela encontrou duas vantagens fundamentais: a primeira é relativa à periodicidade dos dados, que se dá em tempo real praticamente; e a segunda é a flexibilidade, visto que as séries podem ser facilmente convertidas em séries de menor frequência. Além disso, através de vários testes em cima dos dados ela concluiu que os dados gerados pelo GISI se sobressaem em relação aos demais.

No que diz respeito ao mercado imobiliário, ao que se sabe, um dos poucos estudos que utiliza as buscas é o de Humphrey (2010), que faz uma análise dos dados de pesquisa do Google em modelos de mínimos quadrados ordinários (MQO), para averiguar se teriam performances melhores na previsão de vendas de imóveis existentes. Os resultados mostraram que as variáveis do Google reduzem o valor do resíduo nas estimações, ou seja, melhoram o ajustamento das regressões e tornam os modelos mais coerentes com a realidade.



### 3 REFERENCIAL TEÓRICO

Durante um intervalo de tempo, os participantes do mercado, isto é, os compradores e os vendedores, podem estabelecer relações positivas ou negativas, de acordo com sua percepção de oportunidades. Pessoas interagem umas com as outras, e esse fato geralmente produz o que os economistas chamam de efeito “manada”. A decisão de comprar ou vender é influenciada pelas pessoas à nossa volta como também por notícias e informações que temos acesso. Se ouvirmos notícias ruins, podemos ficar tentados a vender. Assim, a relação positiva ou negativa entre os agentes é uma função do estado coletivo de todos os agentes. (PREIS et al., 2010).

As maneiras mais tradicionais de previsão da demanda se dão em volta dos indicadores macroeconômicos, que sinalizam qual a situação geral do mercado, e também na teoria do consumidor, que dá luz às preferências que cada indivíduo tem frente a um bem ou serviço, dado sua renda – ou nesse caso, dado sua renda e seu acesso ao crédito para financiar seu consumo.

Além desses fatores, há também a formação dos preços imobiliários, que se dão através da teoria dos preços hedônicos. Tal teoria postula que os preços são formados levando-se em conta as características dos bens (RYVERA, 2007), ou seja, sua localização, idade, tamanho, número de peças, estética etc.

Conforme Tavares et al. (2012, p. 148):

Muitos problemas do mercado imobiliário resultam da informação imperfeita que o afeta e a obtenção de informação traduz-se num custo, que por vezes pode ser elevado e afasta a transação do imóvel do seu justo valor. Os problemas de assimetria de informação no mercado imobiliário e o desvio que provoca no justo valor da propriedade têm sido invocados como um dos causadores da crise do *Subprime*, pois o valor de mercado pode desviar-se

da utilização de expectativas otimistas e critérios subjetivos que os vieram a inflacionar [...].

Dada a complexidade que envolve a tomada de decisões referentes ao mercado imobiliário, e a falta de informações existente, tem-se muitos custos, como o deslocamento e tempo perdidos, para sanar esses problemas e tomar a melhor decisão. Assim, é evidente que a informação carrega um papel fundamental no processo de tomada de decisão. Quanto maior for a discrepância nos preços, maior o esforço para obtê-la, conforme coloca Humphrey (2010).

Ele continua: Indivíduos maximizam sua utilidade por fazerem pesquisas maiores atrás de informações, afim de que bens mais caros estejam de acordo com suas preferências. Um exemplo disso é o tempo que um indivíduo toma pesquisando um imóvel e o tempo que ele toma pesquisando um grampeador. Dada a magnitude de valores entre os bens e o quanto o indivíduo economizaria fazendo uma escolha acertada, pressupõe-se que ele dedique uma parcela maior do seu tempo pesquisando imóveis. Imóveis são bens mais heterogêneos do que grampeadores e geralmente impactam de maneira mais acentuada o nível de utilidade do indivíduo (HUMPHREY, 2010). O que as buscas fazem é reduzir o tempo despendido procurando por imóveis fisicamente, pois se pode descartar opções ruins sem ter que visitá-los. Outro ponto é que ajudam a se ter uma visão melhor do mercado (preços) sem a necessidade de se ter a ajuda de um corretor.

Ainda para Humphrey (2010), a internet é a nova ferramenta que possibilita a coleta de informações a um custo baixíssimo. Dados sobre vendas de imóveis em pesquisas no Google apresentaram clara relação com o volume de vendas de imóveis efetivo, juntamente com o índice de preços dos imóveis, segundo o trabalho de Wu e Brynjolfsson (2009) no mercado imobiliário norte americano. Assim, buscas na Web podem ser usadas para prever o futuro da atividade econômica, como por exemplo, quando a economia irá se recuperar de uma recessão (WU; BRYNJOLFSSON, 2009).

## 4 METODOLOGIA

Para determinar a influência das buscas online na demanda imobiliária do país, será utilizado um modelo econométrico de Vetores Autorregressivos (VAR) devido a sua capacidade, tanto na realização de previsões – superior aos modelos de equações simultâneas (SIMS, 1980; MCNEES, 1986) – como também em demonstrar a trajetória<sup>2</sup> das séries, dado um choque estrutural (BUENO, 2008). Os modelos VAR, conforme Medeiros et al. (2011), uma generalização natural dos modelos autorregressivos univariados<sup>3</sup> (AR), buscam responder as deficiências dos modelos estruturais de natureza keynesiana que se ancoram em teorias econômicas. Por sua vez, os modelos VAR são tidos como não teóricos (Andrade, 2004).

Os modelos VAR possuem três formas básicas: o VAR irrestrito, o VAR estrutural (SVAR) e o Modelo de Correção de Erro Vetorial (VECM). Para a escolha do melhor modelo é necessário a análise das séries no que tange a estacionariedade, o número certo de defasagens e a cointegração. Tais inferências permitem determinar se o modelo será aplicado com as variáveis em nível ou em suas diferenças e, por fim, qual será o modelo mais adequado.

É importante avaliar se as séries apresentam tendência de crescimento ao passar do tempo, pois pode indicar algum tipo de causalidade entre elas. Não atentar ao fato de duas sequências apresentarem tendência similar ou opostas pode

---

<sup>2</sup> Conforme Bueno (2008), trajetória indica quando um choque afeta uma série, se ela muda de patamar ou não, para que patamar vai, além de outras informações.

<sup>3</sup> Uma regressão univariada é um modelo linear com uma única equação e única variável, na qual seu valor é explicado somente por sua própria defasagem (*lag*) passada. No caso do modelo linear VAR, temos *n*-equações e *n*-variáveis, na qual cada variável é explicada por seus próprios valores defasados, como também pelos valores contemporâneos e passados das variáveis *n*-1 restantes. (STOCK; WATSON, 2001, p. 101)

acarretar em, erroneamente, inferir que variações em uma delas seja causada pela outra, quando na verdade elas apresentam correlação somente devido a outros fatores não observados (WOOLDRIDGE, 2012, p. 363).

A análise de modelos regressivos proporciona a investigação do relacionamento entre as variáveis, através do efeito causal que cada variável impõe a outra no sentido de precedência temporal. A metodologia utilizada nesse sentido, conforme definida por Granger (1969), implica o termo causalidade como a variável  $x$  causar a variável  $y$  se o valor presente de  $y$  pode ser previsto mais precisamente pelos valores passados de  $x$  do que sem eles, com tudo o mais constante.

#### 4.1 Estacionariedade das séries

Conforme Bueno (2008), quando a série temporal não é estacionária, é impossível estimar todos os momentos da série e, assim, surge a possibilidade de se ter regressões espúrias. Diz-se que uma série é estacionária ou possui fraca estacionariedade quando obedece a três condições básicas: a média e a variância devem ser constantes ao longo do tempo e o valor da covariância entre dois períodos de tempo depende somente na diferença ou defasagem entre os dois períodos e não no tempo em que a covariância é computada (GUJARATI, 2004).

Para se testar a estacionariedade das séries, é necessário testar a presença de raízes unitárias. Um dos testes mais utilizado para este fim é o teste *Augmented Dickey-Fuller* (ADF) ou Teste Aumentado de Dickey-Fuller, que compreende estimar a seguinte equação:

$$\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^{p-1} \lambda_i \Delta y_{t-1} + \varepsilon_t \quad (1)$$

em que  $\mu$  é o intercepto,  $y_t$  é a série a ser testada,  $\alpha$  é o operador da 1ª diferença e  $\varepsilon_t$  é a estrutura do resíduo – que se assume ser um ruído branco identicamente e independentemente distribuído. O uso de defasagens procura eliminar a autocorrelação dos resíduos, o qual é determinado pelo menor valor do critério de informação (MEDEIROS et al., 2011). As hipóteses nula e alternativa são, respectivamente,  $H_0: \alpha = 0$  e  $H_1: \alpha < 1$ . A hipótese nula testa a presença de uma raiz

unitária, o que significa que é não-estacionária quando o valor estatístico é menor que o valor crítico. Além disso, o ADF também permite considerar a presença de constante e/ou tendência, avaliar o parâmetro de raiz unitária e a presença ou não de intercepto ou tendência (MARGARIDO et al., 2004).

## 4.2 Análise de cointegração

Quando duas séries não estacionárias tendem a se mover juntas através do tempo, diz-se que elas são cointegradas (ADKINS, 2014), ou seja, se relacionam no longo prazo. Segundo Enders (1995, p. 385), o teste de cointegração de Engle e Granger<sup>4</sup> apresenta algumas limitações, como não identificar quais vetores são cointegrados quando existe mais de duas variáveis no modelo e, se qualquer erro for cometido no primeiro estágio da estimação, onde se avalia a estacionariedade dos resíduos da regressão, todo o procedimento falha, visto que no estágio seguinte se aplica uma regressão dos próprios resíduos.

Assim, para avaliar a presença de cointegração aplicou-se o teste de máxima verossimilhança proposto por Johansen (1988) e Johansen e Juselius (1990), o que, conforme Enders (1995, p. 386), não passa de uma generalização multivariada do teste de Dickey-Fuller. Enders (1995, p. 386-392) também traz uma breve descrição do teste, na qual no caso multivariado, é possível ver a estacionariedade de  $y_t$  como dependente da magnitude de  $(\alpha_t - 1)$ , conforme segue:

$$\Delta y_t = (\alpha_t - 1)y_{t-1} + \varepsilon_t \quad (2)$$

Se  $(\alpha_t - 1) = 0$ ,  $y_t$  tem uma raiz unitária. Tirando o caso em que  $(\alpha_t - 1) \neq 0$ , na qual o processo é explosivo, podemos concluir que a série é estacionária. Aplicando-se uma simples generalização para  $n$  variáveis, temos:

$$x_t = A_1 x_{t-1} + \varepsilon_t$$

---

<sup>4</sup> Bueno (2008, p. 241) traz a definição de Engle e Granger (1987) para cointegração, na qual os elementos do vetor  $X_t$ ,  $n \times 1$ , são ditos cointegrados de ordem  $(d, b)$ , denotados por  $X_t \sim CI(d, b)$ , se todos os elementos de  $X_t$  são integrados de ordem  $d$ , isto é, são  $I(d)$ ; e se existe um vetor não nulo,  $\beta$ , tal que  $u_t = X_t' \beta \sim I(d - b)$ ,  $b > 0$ .

$$\begin{aligned}
\Delta x_t &= A_1 x_{t-1} - x_{t-1} + \varepsilon_t \\
\Delta x_t &= (A_1 - I)x_{t-1} + \varepsilon_t \\
\Delta x_t &= \pi x_{t-1} + \varepsilon_t
\end{aligned} \tag{3}$$

onde  $\pi = (A_1 - I)$ ,  $x_t$  e  $\varepsilon_t$  são vetores  $(n \times 1)$ ,  $A_1$  é uma matriz de parâmetros  $(n \times n)$  e  $I$  é uma matriz identidade  $(n \times n)$ . O *rank* de  $(A_1 - I)$  é igual ao número de vetores cointegrados.

Há diversas maneiras de generalizar a equação (3) e a mesma pode facilmente ser modificada para permitir a presença de tendência e constante. Conforme o teste ADF, o modelo multivariado também pode ser generalizado para permitir um modelo autorregressivo de ordem  $I(p)$ :

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + \varepsilon_t \tag{4}$$

na qual  $x_t$  é um vetor  $(n \times 1)$  indicado por  $(x_{1t}, x_{2t}, \dots, x_{nt})'$ ;  $\varepsilon_t$  é um vetor  $n$ -dimensional independentemente e identicamente distribuído com média zero e matriz de variância  $\Sigma_t$ .

A equação (4) pode ser posta em uma forma mais prática por se subtrair de cada lado  $x_{t-1}$ , em seguida somar e diminuir  $(A_1 - I)x_{t-2}$ , repetindo o processo para  $(A_2 + A_1 - I)x_{t-3}$ , até obter-se:

$$\Delta x_t = \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \pi x_{t-p} + \varepsilon_t \tag{5}$$

em que:

$$\pi = -(I - \sum_{i=1}^p A_i) \tag{6}$$

$$\pi_i = -(I - \sum_{j=1}^i A_j) \tag{7}$$

Novamente,  $\pi$  é o *rank* (posto) da matriz, o qual é igual ao número de vetores cointegrados independentes. Se o  $\text{rank}(\pi) = 0$ , a matriz é nula e a equação (5) é o modelo VAR usual nas primeiras diferenças. Em casos intermediários, se o  $\text{rank}(\pi) = 1$ , há apenas um único vetor cointegrado e a expressão  $\pi x_{t-p}$  é o vetor

correção de erro. Para outros casos na qual  $1 < \text{rank}(\pi) < n$ , existem múltiplos vetores cointegrados.

O número de vetores cointegrados pode ser obtido pelo teste estatístico do traço (*trace statistics*  $\lambda_{\text{trace}}$ ), e a estatística de autovalor máximo (*maximum eigenvalue* =  $\lambda_{\text{max}}$ ) usando seus respectivos valores críticos<sup>5</sup> (JOHANSEN; JUSELIUS, 1990). As estatísticas dos testes  $\lambda_{\text{trace}}$  e  $\lambda_{\text{max}}$  podem ser escritas como:

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^g \ln(1 - \hat{\lambda}_i) \quad (8)$$

$$\lambda_{\text{max}}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (9)$$

em que  $r$  é o número de vetores cointegrados na hipótese nula;  $T$  é o número de observações; e  $\hat{\lambda}_i$  é o valor estimado do autovalor (*eigenvalue*)  $n$ , sendo que os autovalores estão em ordem decrescente. A hipótese nula coloca que o número de vetores cointegrados é menor ou igual a  $r$  e a hipótese alternativa propõe que existem mais que  $r$  vetores. O teste do autovalor máximo é feito para cada valor separadamente, tendo como hipótese nula que o número de vetores cointegrados é igual a  $r$ , contra a hipótese alternativa de que existem  $r+1$  vetores cointegrados (JOHANSEN; JUSELIUS, 1990).

#### 4.3 Especificação: VAR e VECM

Os testes de estacionariedade e cointegração permitem avaliar qual forma do modelo VAR propiciará melhores resultados. Caso as variáveis sejam estacionárias em nível, uma modelagem a partir do VAR irrestrito será aplicada. Através de uma notação matricial, o modelo VAR é expresso por:

$$y_t = A_0 + A_1 y_{t-1} + \dots + A_p y_{t-p} + B_0 z_t + B_1 z_{t-1} + \dots + B_p z_{t-p} + \varepsilon_t \quad (10)$$

---

<sup>5</sup>Johansen e Juselius (1990) proveem os valores críticos das estatísticas dos testes do traço ( $\lambda_{\text{trace}}$ ) e do máximo autovalor ( $\lambda_{\text{max}}$ ) obtidos usando estudos simulados. Os valores críticos estão descritos em Enders (1995 p. 419).

onde  $y$  é um vetor  $n \times 1$  que inclui as variáveis endógenas do modelo;  $z$  é um vetor  $m \times 1$  cujos elementos são variáveis exógenas do modelo;  $A_0$  é um vetor de intercepto  $n \times 1$ ;  $A_1, \dots, A_p$  são matrizes  $n \times n$  de coeficientes que associam valores defasados de variáveis endógenas a seus valores atuais;  $B_1, \dots, B_p$  são matrizes  $n \times m$  de coeficientes que associam valores atuais de variáveis exógenas a valores de variáveis endógenas; e  $\varepsilon_t$  é um vetor  $n \times 1$  de resíduos (inovações) aleatórios  $IID \sim N(0, \sigma^2)$  (MEDEIROS et al., 2011).

Uma crítica aos modelos VAR irrestrito e VAR estrutural é que eles não levam em consideração, de forma explícita, relações de longo prazo entre as variáveis (HAUER, 2007). A análise de cointegração visa testar os resíduos, para constatar tratar-se de uma variável estacionária e, uma vez que o resíduo é estacionário, determinar qual o melhor ajustamento para o modelo VAR, denominado por causa desse ajuste de quase VAR. Quando se utiliza a informação decorrente desse fato, tem-se o modelo VECM, em que se incorpora o erro de equilíbrio, daí a denominação de modelo vetor de correção de erros (BUENO, 2008, p. 243). Dessa forma, caso existam relações de cointegração no sistema, um modelo de Correção de Erro Vetorial (VECM) será utilizado. O modelo VECM pode ser descrito como:

$$\Delta y_t = \Pi_1 y_{t-k} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_{k-1} \Delta y_{t-(k-1)} + u_t \quad (11)$$

em que  $\Pi = (\sum_{j=1}^k \beta_j) - I_g$ ;  $\Gamma_i = (\sum_{j=1}^i \beta_j) - I_g$ ;  $\Delta y_t$  é um vetor com  $n$  variáveis, sendo  $u_t \sim (0, \Sigma)$ , onde  $\Sigma$  é uma matriz de  $u_t$  variâncias com  $E(u_t u_s) = 0$ ,  $\forall t \neq s$  (JOHANSEN; JUSELIUS, 1990).

#### 4.3.1 Autocorrelação e normalidade dos resíduos

Para garantir a melhor especificação do modelo é importante atentar para a análise da autocorrelação dos coeficientes oriundos da regressão, bem como a normalidade dos seus resíduos e, assim, evitar incorrer no erro de gerar regressões espúrias.

A estatística de Ljung-Box,  $Q$ , permite testar em conjunto a hipótese de que todos os coeficientes  $\rho_k$  até determinada defasagem são simultaneamente



iguais a zero (GUJARATI, 2004, p. 813). O teste consiste em estimar a seguinte equação:

$$LB = n(n + 2) \sum_{k=1}^m \left( \frac{\hat{\rho}_k^2}{n - k} \right) \sim \chi^2 m \quad (12)$$

onde  $n$  é valor da amostra e  $m$  é o número de defasagens. É possível testar a presença de heterocedasticidade condicional, na qual a hipótese nula é que a soma das autocorrelações é estatisticamente diferente de zero (BUENO, 2008, p. 285):

$$H_0: \sum_{k=1}^m \hat{\rho}_k = 0 \quad H_A: \hat{\rho}_k \neq 0, \text{ para algum } k.$$

Caso o modelo apresente problemas de autocorrelação, pode-se aumentar o número de defasagens, no entanto é importante ressaltar que as defasagens só devem ser acrescentadas até sanarem o problema, pois à medida que se aumenta as defasagens, perdem-se graus de liberdade na estimação do modelo.

Existem diversos testes para avaliar a normalidade dos resíduos, no entanto foi elencado o teste de Doornik-Hansen, por ser um teste multivariado e estatisticamente robusto. Resumidamente, o teste consiste em estimar a seguinte estatística:

$$E_p = Z'_z Z_1 + Z'_2 Z_2 \sim \chi^2(2p) \quad (13)$$

onde  $Z'_1 = (z_{11}, \dots, z_{1p})$  e  $Z'_2 = (z_{21}, \dots, z_{2p})$ . Após a transformação dos dados em distribuições aproximadamente normais, é aplicado o teste univariado para cada uma das dimensões<sup>6</sup>.

#### 4.4 Função impulso-resposta

Da mesma forma que os modelos autorregressivos possuem uma média móvel, os vetores autorregressivos também podem ser escritos como vetores média móveis (VMA) (ENDERS, 1995, p. 305). Tal representação desempenha um dos principais atributos da metodologia de Sims (1980), pois permite traçar o caminho

<sup>6</sup> Para uma explicação detalhada, vide HANSEN, Henrik; DOORNIK, Jurgen A. An omnibus test for univariate and multivariate normality. **Economics Papers Series** (Nuffield College, University of Oxford), 1994.

das variáveis do sistema VAR através de choques em vários períodos (ENDERS, 1994, p.305).

Para exemplificar a matriz que representa o VMA em um modelo bivariado de primeira ordem, temos:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \epsilon_{yt-i} \\ \epsilon_{zt-i} \end{bmatrix} \quad (14)$$

os 4 elementos  $\phi_{jk}(0)$  são multiplicadores de impacto ou funções impulso resposta<sup>7</sup>. O coeficiente  $\phi_{12}(0)$  representa o impacto instantâneo de um desvio padrão de  $\epsilon_{zt}$  e no valor de  $y_t$ .  $\phi_{11}(1)$  e  $\phi_{12}(1)$  indicam o impacto de um desvio padrão no período seguinte em  $\epsilon_{yt-1}$  e  $\epsilon_{zt-1}$ , respectivamente, em  $y_t$ . Atualizando em um período indica que  $\phi_{11}(1)$  e  $\phi_{12}(1)$  também representa o impacto de um desvio padrão de  $\epsilon_{yt}$  e  $\epsilon_{zt}$  em  $y_{t+1}$ . O efeito acumulado dos choques pode ser obtido através da soma dos coeficientes das funções impulso resposta (ENDERS, 1995, p. 306).

Conforme Margarido (2004) é necessário diagonalizar as matrizes de variância-covariância dos resíduos para prevenir que os choques estejam contemporaneamente correlacionados. Para isso, é utilizado o método de decomposição de Cholesky, o qual permite inferir, período a período, o efeito que o choque de um desvio padrão em uma variável impõe as demais variáveis do modelo. No entanto, a ordenação das equações pode influenciar no resultado da função impulso resposta. Assim, a decomposição de Cholesky é um método arbitrário que exige atenção na interpretação dos resultados.

#### 4.5 Decomposição da variância

De acordo com Enders (1995, p. 310-311) tomando a equação (14) em seu formato compacto e a condicionando para se ter o erro da previsão de  $n$ -períodos futuros,  $x_{t+n} - E_t x_{t+n}$ , temos:

<sup>7</sup> Os passos detalhados para se chegar na matriz que representa o VMA encontram-se em Enders, 1995, p 305-306.

$$y_{t+n} - E_t y_{t+n} = \sum_{i=0}^{n-1} \phi_i \epsilon_{t+n-i} \quad (15)$$

Denotando-se a variância do erro da previsão para  $n$ -períodos de  $y_{t+n}$  como  $\sigma_y(n)^2$ :

$$\begin{aligned} \sigma_y(n)^2 = & \sigma_y^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2] \\ & + \sigma_z^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2] \end{aligned} \quad (16)$$

Como todos os valores de  $\phi_{jk}(i)^2$  são positivos, a variância do erro da previsão aumenta a medida que o horizonte de previsões  $n$  aumenta. A proporção da variância do erro da previsão,  $\sigma_y(n)^2$ , para cada  $n$ -período futuro em função de cada um dos choques,  $\epsilon_{yt}$  e  $\epsilon_{zt}$ , é respectivamente:

$$\frac{\sigma_y^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2]}{\sigma_y(n)^2} \quad (17)$$

$$\frac{\sigma_z^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2]}{\sigma_y(n)^2} \quad (18)$$

Desta forma, a decomposição do erro da previsão indica a proporção do movimento que uma variável terá através de choques em si mesma versus choques nas demais variáveis, ou seja, ela demonstra qual a proporção em que uma variável explica a si mesma e as demais em  $n$  períodos.

#### 4.6 Causalidade de Granger

Conforme colocado anteriormente, uma variável  $x_t$  é causal para uma variável  $y_t$ , se os valores passados de  $x_t$  auxiliam na previsão de  $y_t$ . De acordo com o detalhamento feito por Lütkepohl e Krätzig (2004), admitindo-se que  $y_{t+h|\Omega_t}$  é o melhor predictor para  $h$  períodos de  $y_t$  na origem  $t$  baseado em todas as informações relevantes do universo  $\Omega_t$ , podemos definir que  $x_t$  não Granger-causa  $y_t$  se, e somente se  $y_{t+h|\Omega_t} = y_{t+h|\Omega_t \setminus \{x_s | s \leq t\}}$ ,  $h = 1, 2, \dots$ . Desta forma, a hipótese nula de que  $x_t$  não Granger-causa  $y_t$  é aceita caso removendo o passado de  $x_t$  do conjunto de informações não se altera a previsibilidade de  $y_t$  para nenhum horizonte de

períodos. Em contrapartida,  $x_t$  Granger-causa  $y_t$  se uma melhor previsão for gerada pelo uso de informações passadas de  $x_t$  para um horizonte de períodos futuros. Se  $\Omega_t$  contém valores passados de  $y_t$  e  $x_t$  apenas, isto é,  $\Omega_t = \{(y_s, x_s)' | s \leq t\}$ , e  $(y_t, x_t)'$  é gerado por um VAR( $p$ ) bivariado denotado por:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \sum_{i=1}^p \begin{bmatrix} \alpha_{11,i} & \alpha_{12,i} \\ \alpha_{21,i} & \alpha_{22,i} \end{bmatrix} \begin{bmatrix} y_{t-i} \\ x_{t-i} \end{bmatrix} + u_t, \quad (19)$$

então,  $y_{t+h} | \Omega_t = y_{t+h} | \Omega_t \setminus \{x_s | s \leq t\}$ ,  $h = 1, 2, \dots$  é equivalente a  $\alpha_{12,i} = 0$ ,  $i = 1, 2, \dots, p$ . Nota-se que  $x_t$  não Granger-causa  $y_t$ , se suas defasagens não aparecem na equação de  $y_t$ . Analogamente,  $y_t$  não Granger-causa  $x_t$  se seus lags não se encontram na segunda equação.

Da mesma forma que para o VAR a causalidade de Granger também pode ser investigada no VECM. Porém, a maneira a qual ela foi testada seguiu a metodologia proposta por Toda e Yamamoto (1995). Segundo eles, o primeiro passo compreende testar a ordem de integração de cada uma das séries do modelo, na qual a maior ordem encontrada seja denotada por  $m$ .

A seguir é necessário aplicar um modelo VAR com as séries em nível, independente de sua ordem de integração<sup>8</sup>. O número máximo de defasagens deve ser estabelecido conforme colocado anteriormente, utilizando-se o menor valor para os critérios de informação e aquele o qual sana os problemas de autocorrelação do modelo. O valor da máxima defasagem encontrada será  $p$ .

O próximo passo consiste em estimar novamente um modelo VAR, porém adicionando-se como exógenas ao modelo o mesmo conjunto de variáveis com o número de defasagens igual ao número encontrado anteriormente mais a ordem de integração máxima<sup>9</sup>, isto é,  $p + m$ . O teste de não causalidade de Granger é feito através de um teste de Wald dos coeficientes estimados no modelo apenas para as variáveis  $p$  defasadas. A hipótese nula é de que os primeiros valores  $p$  defasados de

<sup>8</sup> Mesmo que duas ou mais séries apresentem a mesma ordem de integração e cointegrem entre si deve-se aplicar um modelo VAR em nível para testar a causalidade de Granger.

<sup>9</sup> Esse artifício permite corrigir para uma distribuição  $\chi^2$  assintótica no teste de Wald.

cada uma das variáveis são zero nas demais<sup>10</sup>. Rejeitar a hipótese nula significa não rejeitar a presença de causalidade de Granger<sup>11</sup>.

---

<sup>10</sup> A estatística do teste de Wald é assintoticamente  $\chi^2$  distribuída com  $p$  graus de liberdade na hipótese nula.

<sup>11</sup> Para uma explicação detalhada do teste de causalidade de Granger, consultar TODA, Hiro Y.; YAMAMOTO, Taku. Statistical inference in vector autoregressions with possibly integrated processes. **Journal of econometrics**, v. 66, n. 1, p. 225-250, 1995.

## 5 BASE DE DADOS

O trabalho visa estimar um modelo que permita inferir sobre a utilidade das buscas na internet para prever a demanda imobiliária. No entanto, outras variáveis se fazem importantes no modelo, uma vez que a demanda imobiliária depende de outros fatores ligados à conjuntura econômica do país, tais como o nível de desemprego, os salários e as taxa de juros. Todas as séries são mensais e compreendem janeiro de 2004 até janeiro de 2016. A periodicidade foi definida dada a maior disponibilidade dos dados possível, a fim de incluir a maior quantidade de dados pertinentes ao modelo.

No total, foram computados 725 observações referentes as 5 variáveis, no período de 145 meses que abrangeram a amostra. Para facilitar a análise, as séries foram indicadas de acordo com sua primeira letra em inglês (por ser mais usual na literatura). Assim, a taxa de desemprego será chamada de “*u*”; os salários “*w*”; a taxa de juros selic “*i*”; as buscas no Google “*s*”; e a demanda imobiliária “*d*”.<sup>12</sup>

Conforme o intuito do trabalho foi elencado um modelo Vetor Autorregressivo (VAR), contudo a forma como ele se dará (VAR Estrutural ou Vetor Correção de Erro – VECM) dependerá do comportamento das séries analisadas.

### 5.1 Demanda por imóveis

A primeira série do modelo é a demanda imobiliária em. Ela é importante para avaliar se o comportamento das buscas e das demais variáveis influenciam no seu próprio comportamento.

---

<sup>12</sup> A estatística descritiva das séries e os respectivos gráficos temporais estão apresentados no apêndice desse trabalho.

A demanda efetiva dos imóveis, por se tratar de dados escassos, ficou restrita a 8 capitais brasileiras: Belo Horizonte, Fortaleza, Goiânia, Maceió, Porto Alegre, Recife, Rio de Janeiro e São Paulo.

Todos os valores foram agrupados e elaborados pelo banco de dados da Câmara Brasileira da Indústria da Construção (CBIC), porém, provém de fontes distintas de acordo com a cidade.

TABELA 1: Fonte dos dados CBIC

Cidade	Fonte dos dados
Belo Horizonte	IPEAD-UFMG
Fortaleza	SINDUSCON-CE/FIEC/IEL
Goiânia	ADEMI-GO
Maceió	ADEMI-AL/SINDUSCON-AL
Porto Alegre	SINDUSCON-RS
Recife	FIEPE
Rio de Janeiro	ADEMI-RJ
São Paulo	SECOVI-SP/EMBRAESP

Fonte: Elaboração própria.

Foi aplicado o somatório do valor das vendas de todas as cidades em cada mês, com o intuito de expressar, o mais próximo possível, o comportamento das vendas de imóveis no país. As séries têm periodicidade mensal, iniciando em janeiro de 2004 e se estendendo até janeiro de 2016. É importante ressaltar que no caso da cidade de Fortaleza, a metodologia de pesquisa dos dados foi alterada em abril de 2013, impossibilitando a comparação das duas séries. Porém, ela segue refletindo a demanda imobiliária em algum grau, o que é suficiente para atingir os objetivos do trabalho.

## 5.2 Google Trends

As buscas no Google aparecem como um potencial indicador antecedente para as vendas dos imóveis, pois, como outros bens, os imóveis também recebem uma pesquisa prévia que procura filtrar os locais passíveis de serem adquiridos futuramente.

A principal variável deste estudo será o volume de pesquisas relacionadas com o mercado imobiliário, disponível pelo Google Trends. Os dados da plataforma são uma amostra aleatória das buscas feitas no próprio Google, na qual uma porcentagem das buscas é usada para computar os dados do Google Trends.

De acordo com o volume de buscas do termo pesquisado, a plataforma provê diária, semanal ou mensalmente relatórios estatísticos. As buscas são proporcionais ao tempo e a localidade geográfica na qual são feitas. Cada dado gerado é dividido pelo total de buscas feitas no período e localidade a qual representam, para comparar a popularidade relativa do termo pesquisado, pois de outra forma, locais com maior volume de pesquisas seriam elencados no topo. Os resultados são então escalados em um índice de 0 a 100 baseados na proporção de um tópico específico para todas as buscas de todos os tópicos. Diferentes regiões que mostram o mesmo número de buscas para algum termo não necessariamente apresentam o mesmo volume de buscas total<sup>13</sup>.

Os termos podem ser pesquisados através de categorias específicas. No caso deste trabalho, a pesquisa por “imóveis”, pode ser feita amplamente, como termo de pesquisa; pode ser referida a bens imobiliários, sob o tópico de indústria imobiliária, pode se relacionar a registros de imóveis, corretor de imóveis, e outros. No entanto, algumas categorias enquadram pesquisas não relacionadas apenas a demanda por imóveis – como, por exemplo, pesquisando-se “imóveis” através da ótica de registro de imóveis, englobaria terrenos e locais vazios e não imóveis físicos propriamente.

É intuitivo esperar que as pessoas pesquisem termos como, “imóveis a venda”, “venda de imóveis”, “imóveis”, “casas a venda”, “venda de apartamentos”, etc. Cada termo apresenta resultados diferentes e a maneira como são escritos (grafia e ordem de palavras) também alteram os retornos. A ideia principal é englobar termos que mais se aproximem das pesquisas que antecedem a compra de imóveis, por isso, inicialmente foram adotados os termos “imóveis”, “imoveis”, sua soma e a média, “apartamentos a venda”, “casas a venda”, sua soma e média, “aluguel” (imobiliárias), “a venda” (imobiliárias), “venda” (imobiliárias), “a venda” (imóveis), “venda” (imóveis), “imóveis a venda”, “imoveis a venda”, sua soma e sua

---

<sup>13</sup> Para mais informações, consultar <https://support.google.com/trends/?hl=en#topic=>



média. Os termos entre parênteses representam uma subcategoria a qual o termo em si foi computado, os resultados foram filtrados para responderem somente às pesquisas feitas no Brasil.

Após testes preliminares, a variável definida para as buscas na internet foi o termo “a venda” retirado da subcategoria “imóveis”. Os dados apresentam periodicidade mensal e compreendem janeiro de 2004 até janeiro de 2016.

### 5.3 DADOS MACROECONÔMICOS

Variáveis macroeconômicas indicam o estado que a economia do país se encontra, ou seja, o estágio do ciclo de negócios que se encontra, e apontam que tipo de política está sendo adotada pelo governo. O nível de desemprego da economia pode indicar um momento de expansionismo ou recessão, bem como o nível dos salários. Conforme a teoria do consumidor, dado um aumento da renda, os agentes tenderiam a consumir mais, isto se refletiria, por exemplo, no aumento de financiamentos imobiliários. Nessa análise, os juros entram como custo de oportunidade: por um lado oferecem taxas de retorno ao investimento e por outro precificam o financiamento.

Alguns dados macroeconômicos corroboram para a avaliação do melhor modelo econométrico bem como para lhe conferir robustez. Além disso, uma análise empírica do cenário econômico do país permite inferir de maneira mais aguçada na interpretação dos resultados.

Todos os dados foram extraídos do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil<sup>14</sup> e abrangem janeiro de 2004 até janeiro de 2016, conforme as demais séries.

As variáveis consideradas mais pertinentes foram a taxa percentual de desemprego (na semana) das regiões metropolitanas do Brasil; o rendimento médio total real efetivo das pessoas ocupadas em reais; e a taxa de juros SELIC acumulada no mês e anualizada.

De acordo com a teoria econômica, basicamente a taxa de desemprego afeta negativamente o nível de salários, ou seja, quanto mais pessoas desempregadas, mais mão de obra disponível e assim, menor são os salários

---

<sup>14</sup> Disponível em: <https://www3.bcb.gov.br/sgspub/localizarseries/>

ofertados. Com mais pessoas sem uma fonte de renda e com menos pessoas com uma fonte de renda menor, o consumo tende a ser reduzido, inclusive o consumo de bens imobiliários. Analogamente, o cenário pode ser de baixo desemprego, salários mais altos e maior consumo dentro da economia.

Segundo Mendonça (2013), o mercado imobiliário brasileiro é impulsionado principalmente pelo governo. Em 2008, com a explosão da crise do *subprime* norte americano, medidas anticíclicas atuaram nesse contexto, elevando a disponibilidade de crédito em todos os setores, conforme Machado et al. (2014).

Os juros atuam de modo a precificar o financiamento. Mendonça (2013) aponta para uma quebra estrutural para as concessões de financiamentos imobiliários, ocorrida em meados de 2005, onde, além de outros fatores, mudanças na legislação referentes ao crédito direcionado ao setor refletiram um cenário de alta expansão.

Obviamente essas variáveis também são afetadas por outros fatores, como taxa de inflação, mudanças do cenário econômico internacional, expansão ou retração do PIB, etc.

## 6 RESULTADOS

### 6.1 Análise de estacionariedade

O teste de raiz unitária foi aplicado nas variáveis em nível a um grau de significância de 5%. A tabela abaixo ilustra os resultados do teste ADF. Os testes com a presença de *dummies* sazonais foram feitos apenas nas séries que graficamente apresentam esse comportamento.

TABELA 2: P-valor do teste ADF

Especificação do teste	<i>u</i>	<i>w</i>	<i>i</i>	<i>s</i>	<i>d</i>
	-1,87151	0,221393	-0,494142	0,5978	-0,146997
Sem constante	(0,05847)	(0,7506)	(0,5024)	(0,8455)	(0,6331)
	-2,63739		-1,72895		-1,88978
Com constante	(0,08543)	-	(0,4166)	-	(0,3375)
	0,176249		-1,64644		-0,838291
Com constante e tendência	(0,9977)		(0,7746)		(0,9609)
		-1,72845		-0,686973	
Com constante e <i>dummies</i> de sazonalidade	-	(0,4169)	-	(0,8482)	-
		0,615351		-2,55641	
Com constante e tendência <i>dummies</i> de sazonalidade	-	(0,9996)	-	(0,3007)	-

Fonte: Elaboração própria com o auxílio do auxílio do software Gretl 2016b.

De acordo com os resultados, não se rejeita a hipótese nula das séries serem não estacionárias em nível, a um grau de significância de 5%, ou seja, considera-se que nenhuma delas é estacionária em nível.

## 6.2 Análise de cointegração

A análise de cointegração foi feita sob a ótica do teste de Johansen, com ordem de defasagem igual a 3<sup>15</sup>, presença de *dummies* sazonais e conforme o caso 3, constante sem restrições. Ela acompanha os seguintes resultados:

TABELA 3: Estatística do teste do traço, teste do autovalor máximo e teste do traço corrigido para o tamanho da amostra, respectivamente

<i>Rank</i> ( <i>r</i> )	Teste Traço	Teste Autovalor Máx	Teste Traço*
0	0,0053	0,1222	0,0081
1	0,0324	0,1053	0,0400
2	0,1796	0,4894	0,1900
3	0,1632	0,3515	0,1689
4	0,0596	0,0596	0,0623

Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

Nota: \*Corrigido para o tamanho da amostra ( $gl = 115$ ).

O número de relações de cointegração selecionadas ao nível de 5% de significância indicam para o *rank* igual a 2, através da estatística do teste do traço, na qual se aceita a hipótese de que o número de vetores cointegrados é menor ou igual a *r*.

O teste do autovalor máximo é feito para cada valor separadamente, a hipótese nula põe que o número de vetores cointegrados é igual a *r*, contra a

<sup>15</sup> De acordo com os testes preliminares, que envolveram 17 combinações de diferentes termos de buscas, o número de 3 defasagens mostrou ser um valor intermediário, e, para esse caso, apresentou o resultado mais parcimonioso na especificação final do modelo, como é exposto a seguir.

hipótese alternativa de que existem  $r + 1$  vetores cointegrados. Desta forma, o teste assume que não há presença de vetores cointegrados no modelo.

Contudo, o teste do traço corrigido para o tamanho da amostra aponta para a presença de relação de cointegração para o *rank* igual a 2 a um nível de significância de 5%. Dado que o teste corrigido para o tamanho da amostra tende a ser mais robusto, a ordem de cointegração do modelo foi definida como  $rank = 2$ .

A presença de cointegração entre as variáveis indica que elas apresentam relacionamento no longo prazo. Assim, um modelo VAR irrestrito não levaria em consideração essa relação de forma explícita, sendo necessária a aplicação de um modelo VECM.

### 6.3 Especificação: VECM

De acordo com os resultados encontrados nos testes de estacionariedade e cointegração das variáveis, o modelo mais indicado para o propósito do trabalho é o modelo vetor correção de erro – VECM. Foi atribuído para o teste de Johansen o número de 3 defasagens, *dummies* sazonais e constante sem restrições. Doravante, é muito importante que se siga a mesma especificação ao se estimar o modelo VECM.

Desta forma, será estimado um sistema VECM, com grau de defasagem igual a 3, observações de abril de 2004 até janeiro de 2016, totalizando 142 observações. A ordem de cointegração definida no teste de Johansen foi *rank* 2 e tem-se o caso 3, constante sem restrições. A tabela a seguir mostra os resultados obtidos do modelo.

TABELA 4: Estatísticas do modelo VECM

Estatística	<i>u</i>	<i>w</i>	<i>i</i>	<i>s</i>	<i>d</i>
Média var. dependente	-0,030986	3,589155	-0,014366	0,563380	-2,788732
D.P. var. dependente	0,248451	190,5922	0,336411	6,370883	1686,262
Soma resíduos quadrados	6,970096	198092,7	3,739704	776,5754	1,15e+08
E.P. da regressão	0,244077	41,14730	0,178783	2,576315	989,4443
R-quadrado	0,199177	0,961324	0,765643	0,864305	0,714307
R-quadrado ajustado	0,034905	0,953391	0,717570	0,836470	0,655704
Rô	-0,028248	-0,007524	-0,040189	-0,009191	0,012428
Durbin-Watson	2,030691	2,005440	2,075493	1,979908	1,964283

Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

### 6.3.1 Autocorrelação e normalidade dos resíduos

Para avaliar a autocorrelação dos resíduos, foi utilizado o teste Ljung-Box. Segue a lista das equações e seus respectivos resultados.

- Equação *u*:  
 $Q' = 11,9143$  com  $p\text{-valor} = P(Q_{ui} - \text{quadrado}(12) > 11,9143) = 0,453$
- Equação *w*:  
 $Q' = 17,8179$  com  $p\text{-valor} = P(Q_{ui} - \text{quadrado}(12) > 17,8179) = 0,121$
- Equação *i*:  
 $Q' = 17,1624$  com  $p\text{-valor} = P(Q_{ui} - \text{quadrado}(12) > 17,1624) = 0,144$
- Equações:  
 $Q' = 18,6638$  com  $p\text{-valor} = P(Q_{ui} - \text{quadrado}(12) > 18,6638) = 0,097$
- Equação *d*:  
 $Q' = 10,6663$  com  $p\text{-valor} = P(Q_{ui} - \text{quadrado}(12) > 10,6663) = 0,558$

Foi determinado o número de 12 defasagens para testar a hipótese de que todos os coeficientes são simultaneamente iguais a zero. De acordo com o

resultado, percebe-se a ausência de autocorrelação nos resíduos de todas as equações a um nível de significância de 5%.

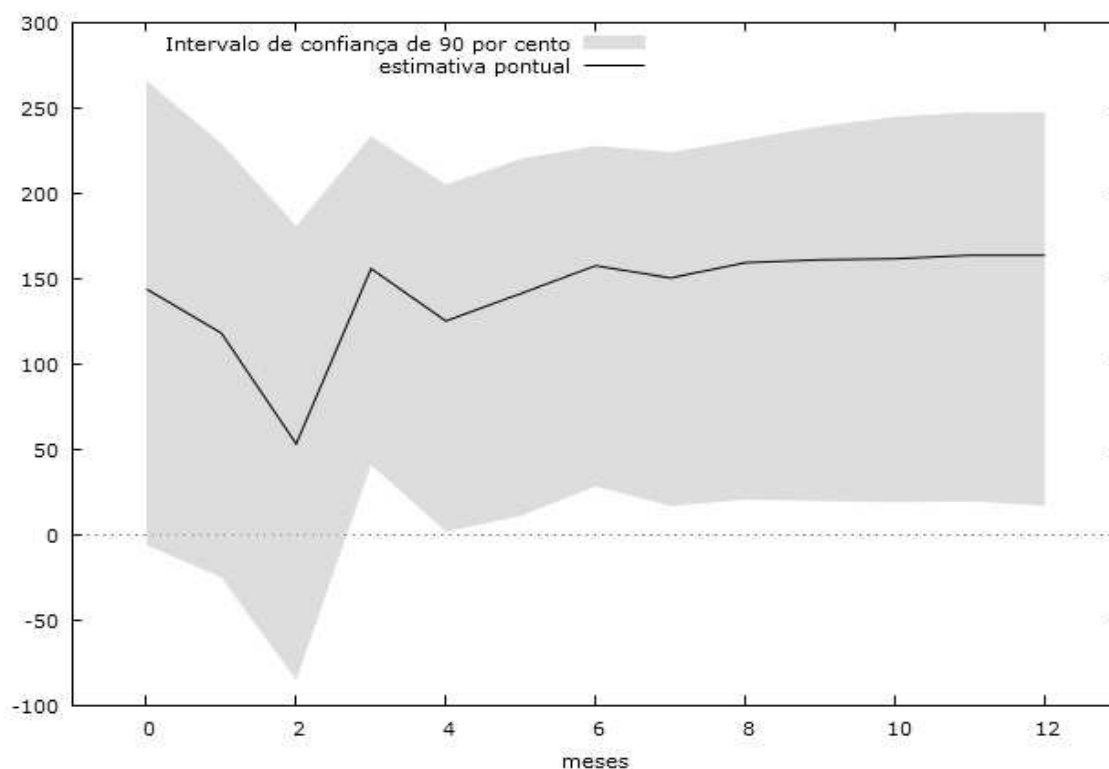
Já o teste de normalidade foi feito através da estatística de Doornik-Hansen, apontando para a presença de forte anormalidade nos resíduos a 5% de significância:  $qui - quadrado(10) = 52,8673 [0,0000]$ . Ao investigar-se a origem da anormalidade, notou-se apenas as séries de salários ( $w$ ) e demanda imobiliária ( $d$ ) como sendo as responsáveis. Para resolver o problema, algumas *dummies* de choque foram introduzidas e, de fato, corroboraram para a normalidade do modelo. No entanto, a ordem de cointegração pode diferir quando observado com a presença de *dummies* e, uma vez que o teste de cointegração não é robusto o suficiente ao seu uso, o modelo não foi alterado.

#### 6.4 Função impulso-resposta

A função impulso-resposta é importante para se analisar qual a intensidade que o choque de um desvio padrão em cada variável através de  $n$ -períodos.

Conforme exposto na metodologia, é importante atentar para a ordenação das equações ao se aplicar a decomposição de Cholesky. Empiricamente foi adotada uma sequência cronológica das equações, porém cabe ressaltar que foi uma decisão arbitrária. A ordem foi definida pela taxa de desemprego ( $u$ ), o nível de salários ( $w$ ), a taxa de juros ( $i$ ), as buscas na internet ( $s$ ) e, por fim, a demanda imobiliária ( $d$ ). O gráfico a seguir demonstra a trajetória de  $d$  após os choques em  $s$ .

GRÁFICO 1: Resposta de  $d$  a um choque em  $s$ , com intervalo de confiança *bootstrap* a 90%



Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

A análise do gráfico permite inferir que a partir do 4º período, todas as variações de  $d$  são significativas ao choque. Para explicitar os valores encontrados graficamente, segue a tabela:



TABELA 5: Respostas a um choque de um erro padrão em  $s$ 

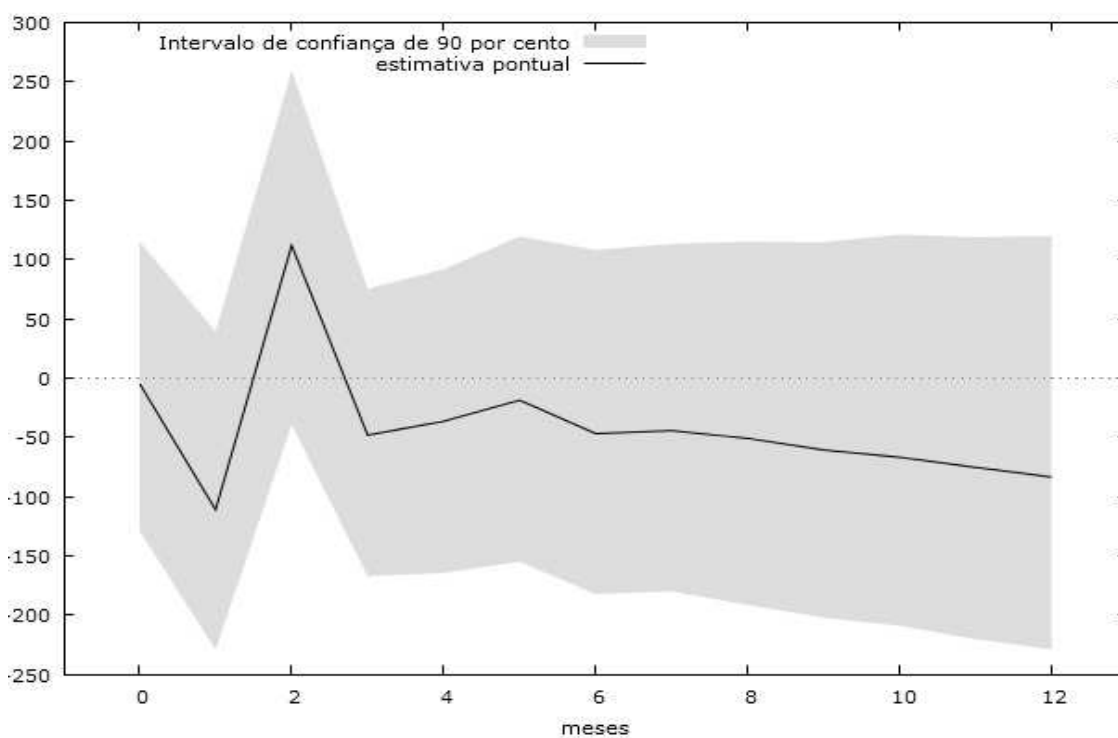
Período	$d$
1	143,82
2	118,02
3	53,051
4	155,77
5	125,04
6	141,11
7	157,49
8	150,29
9	159,30
10	160,98
11	161,56
12	163,62

Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

Dada a significância dos resultados a partir do 4<sup>o</sup> mês até o fim do período, é possível estimar os efeitos do choque de um desvio padrão nas buscas na internet na demanda imobiliária para o período de um ano. Isso é feito somando-se todos os períodos significativos, ou seja, os valores oriundos do choque a partir do 4<sup>o</sup> até o 12<sup>o</sup> mês. Desta maneira, pode-se concluir que o efeito do choque de um erro padrão nas buscas online acarreta na venda de aproximadamente 1375 imóveis no Brasil em um ano.

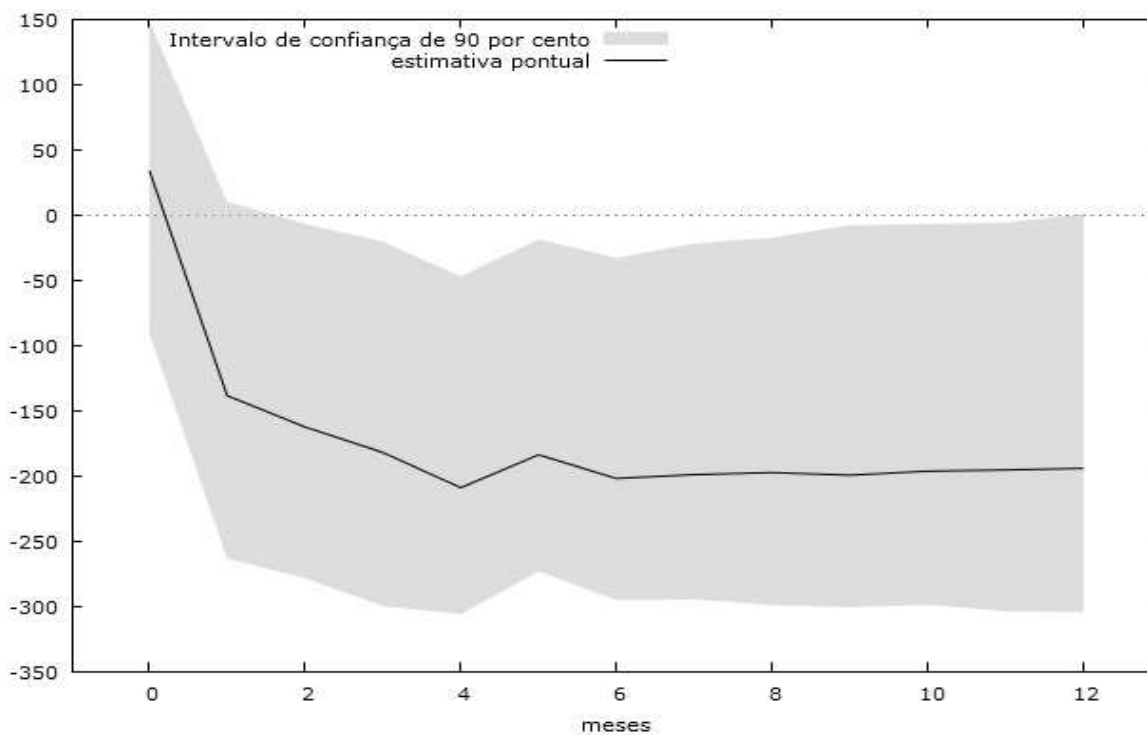
Além da resposta da demanda ao choque nas buscas, tem-se também os efeitos das demais variáveis. Os gráficos a seguir mostram o comportamento da demanda imobiliária frente aos choques nas variáveis desemprego, salários, taxa de juros e na própria demanda respectivamente.

GRÁFICO 2: Resposta de  $d$  a um choque em  $u$ , com intervalo de confiança *bootstrap* a 90%



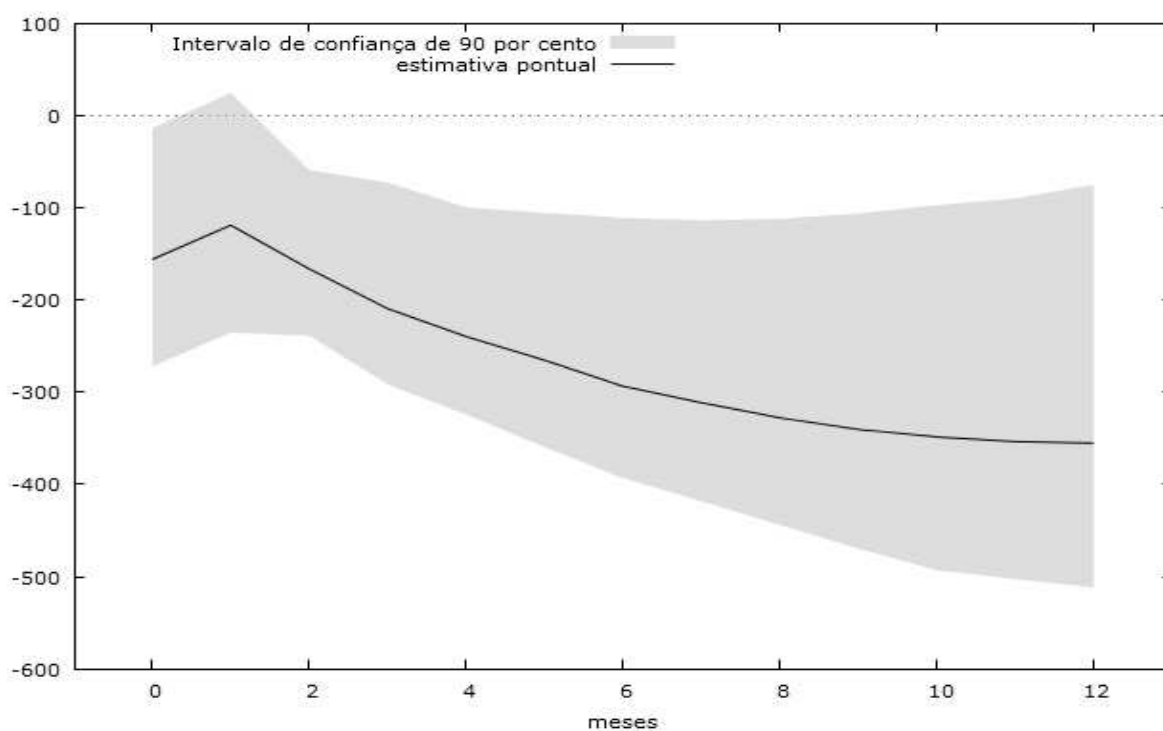
Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

GRÁFICO 3: Resposta de  $d$  a um choque em  $w$ , com intervalo de confiança *bootstrap* a 90%



Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

GRÁFICO 4: Resposta de  $d$  a um choque em  $i$ , com intervalo de confiança *bootstrap* a 90%



Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

Como feito anteriormente, uma tabela refletindo os valores dos gráficos é adequada para a análise individual dos efeitos dos choques. Para se distinguir em qual variável o choque foi aplicado, elas foram colocadas ao lado da demanda imobiliária.

TABELA 6: Respostas de  $d$  ao choque de um erro padrão em  $u$ ,  $w$ ,  $i$  e  $d$ 

Período	$u_d$	$w_d$	$i_d$
1	-4,8310	34,006	-155,84
2	-111,24	-138,62	-118,95
3	111,82	-162,66	-166,35
4	-48,233	-182,23	-209,47
5	-36,637	-209,21	-239,68
6	-18,803	-183,98	-265,38
7	-46,864	-202,16	-293,57
8	-44,456	-199,23	-311,43
9	-51,036	-197,58	-327,87
10	-60,821	-199,57	-340,50
11	-66,927	-196,53	-348,49
12	-75,449	-195,66	-353,56

Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

Os gráficos permitem avaliar os períodos significativos, através do intervalo de 90% de confiança. Na análise dos choques na taxa de desemprego, nota-se que não há períodos significativos na demanda imobiliária.

Tomando os choques no nível de salários, tem-se, pelo menos do 4<sup>o</sup> ao 11<sup>o</sup> período, algum grau de significância, o que implica em aproximadamente uma diminuição de 1570 imóveis demandados em um ano.

Para a taxa de juros, a demanda imobiliária é ainda mais sensível. Os períodos significativos compreendem do 3<sup>o</sup> mês até o 12<sup>o</sup>, resultando em 2856 imóveis a menos demandados no Brasil.

Os choques na própria demanda são os mais significativos, englobando todo o período analisado. Isso quer dizer que em um ano, aproximadamente 3898 imóveis são adquiridos no Brasil, dado um choque de um desvio padrão na própria demanda imobiliária. Uma explicação econômica para esse fato pode se dar no

“efeito manada”, na qual os indivíduos passam a seguir o comportamento de outros, por todos estarem aparentemente se movendo na mesma direção.

### 6.5 Decomposição da variância

A decomposição do erro da previsão demonstra qual a proporção em que uma variável explica a si mesma e as demais em  $n$  períodos. A tabela a seguir denota os resultados encontrados para, em um horizonte de 12 períodos, a decomposição da variância para a demanda de imóveis. A ordenação das equações seguiu o mesmo padrão da análise da função impulso-resposta.

TABELA 7: Decomposição do erro da previsão para  $d$

Período	Erro padrão	$u$	$w$	$i$	$s$	$d$
1	898,132	0,0029	0,1434	3,0107	2,5643	94,2787
2	974,033	1,3068	2,1474	4,0512	3,6484	88,8463
3	1034,03	2,3290	4,3800	6,1827	3,5005	83,6078
4	1152,3	2,0506	6,0279	8,2833	4,6461	78,9920
5	1233,21	1,8787	8,1409	11,0094	5,0845	73,8866
6	1313,38	1,6768	9,1396	13,7892	5,6371	69,7574
7	1401,01	1,5855	10,1140	16,5089	6,2175	65,5741
8	1481,03	1,5089	10,8602	19,1950	6,5936	61,8423
9	1561,1	1,4650	11,3765	21,6873	6,9758	58,4954
10	1639,45	1,4659	11,7970	23,9775	7,2892	55,4704
11	1714,63	1,4925	12,0990	26,0518	7,5519	52,8049
12	1787,5	1,5515	12,3307	27,8831	7,7866	50,4481

Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

De acordo com a decomposição do erro da previsão para  $d$ , percebe-se que, após um ano, aproximadamente 1,55% das variações da demanda por imóveis devem-se a variações da taxa de desemprego, 12,33% a variações nos salários,

27,88% a variações nas taxas de juros<sup>16</sup>, 7,78% a variações nas buscas e 50,44% a variações na própria demanda imobiliária.

É razoável que quase 8% das variações da demanda sejam explicadas por variações nas buscas. Para reforçar essa evidência, foi feita a análise de previsão dentro da amostra para a demanda de imóveis, com e sem a presença da variável *buscas* durante todo o período analisado (jan. 2004 – dez. 2016). O teste foi feito para um intervalo de 95% de confiança. O resultado segue na tabela a seguir.

TABELA 8: Estatísticas de avaliação da previsão

	Ausência da variável <i>buscas</i>	Presença da variável <i>buscas</i>
Erro médio	4,952e-010	1,5671e-010
Erro Quadrado médio	8,2934e+005	8,0664e+005
Erro Unitário Médio Quadrado	910,68	898,13
Erro Absoluto Médio	715,64	694,61
Erro Percentual Médio	-1,8496	-1,6698
Erro Percentual Médio Absoluto	14,534	13,99

Fonte: Elaboração própria com o auxílio do software Gretl 2016b.

De acordo com as estatísticas da avaliação da previsão, percebe-se que, em todos os parâmetros, a presença da variável *buscas* obteve uma performance melhor, mesmo que pequena.

Tal resultado é altamente relevante, pois indica que as buscas na internet e o seu comportamento temporal, oriundo do Google Trends, corroboram para auxiliar a previsão da demanda imobiliária no Brasil.

## 6.6 Causalidade de Granger

Para a análise de causalidade, foram adotados os passos propostos por Toda e Yamamoto (1995), na qual deve ser gerado um modelo VAR irrestrito em

<sup>16</sup> O resultado encontrado na variação da demanda imobiliária derivado das variações na taxa de juros após um ano faz muito sentido do ponto de vista econômico, pois os juros precificam o dinheiro, sendo assim, é fator determinante para a tomada de financiamentos.

nível, e as mesmas medidas de especificação devem ser seguidas conforme o VECM anterior.

A melhor especificação encontrada foi um VAR com 5 defasagens, presença de constante e *dummies* sazonais, juntamente com o mesmo conjunto de variáveis defasadas 6 períodos como variáveis exógenas ao modelo<sup>17</sup>. Após estimar a regressão, foi aplicado um teste de Wald nas variáveis defasadas 5 períodos. O resultado indica um teste de causalidade de Granger, denotado na seguinte tabela.

---

<sup>17</sup> Conforme destacado na metodologia de Toda e Yamamoto (1995), as variáveis exógenas devem ser defasadas de acordo com o máximo grau de integração que possuem, nesse caso  $I(1)$  mais o número defasagens que melhor especifica o modelo VAR. Assim, tem-se  $1 + 5 = 6$  defasagens nas variáveis exógenas.

TABELA 9: Causalidade de Granger

	Qui-quadrado	Probabilidade
$w \rightarrow u$	0,644694	0,9859
$i \rightarrow u$	3,182694	0,6718
$s \rightarrow u$	3,52733	0,6195
$d \rightarrow u$	5,868826	0,3192
$u \rightarrow w$	9,885115	0,0786*
$i \rightarrow w$	3,246929	0,6620
$s \rightarrow w$	2,269126	0,8108
$d \rightarrow w$	7,636217	0,1775
$u \rightarrow i$	4,829050	0,4371
$w \rightarrow i$	2,091864	0,8363
$s \rightarrow i$	2,498980	0,7766
$d \rightarrow i$	4,459086	0,4854
$u \rightarrow s$	7,836651	0,1655
$w \rightarrow s$	14,34378	0,0136**
$i \rightarrow s$	6,520335	0,2588
$d \rightarrow s$	8,58254	0,1269
$u \rightarrow d$	12,1563	0,0327**
$w \rightarrow d$	3,947031	0,5571
$i \rightarrow d$	14,12545	0,0148**
$s \rightarrow d$	1,443568	0,9195

Fonte: Resultados do Estudo.

Nota: \*Significativo a 10%, \*\*Significativo a 5%, \*\*\*Significativo a 1%. Fonte: Resultados do Estudo.



Analisando-se a estatística do teste a um nível de significância de 5%, tem-se que as relações de causalidade significativas se dão apenas em  $w \rightarrow s$ ,  $u \rightarrow d$  e  $i \rightarrow d$ , ou seja, rejeita-se a hipótese nula de não causalidade.

Apesar de as buscas não Granger-causarem a demanda de acordo com a estatística do teste de Wald, as relações encontradas fazem muito sentido do ponto de vista econômico. Os salários precedem temporalmente as buscas, talvez porque os indivíduos só cogitem procurar por imóveis se efetivamente tiverem disponíveis recursos para adquirir um. Doravante, a taxa de desemprego, por estar relacionada com o nível de salários, também Granger-causa a demanda imobiliária, e, por fim, a taxa de juros – como está fortemente ligada ao “preço” dos financiamentos, neste caso, imobiliários – apresenta precedência temporal à demanda imobiliária.

## 7 CONSIDERAÇÕES FINAIS

O trabalho trouxe como proposta principal averiguar a capacidade do Google Trends em prever a demanda de imóveis no Brasil. O modelo VECM com 3 defasagens – apesar de apresentar problemas de normalidade em função da série de salários ( $w$ ) e demanda imobiliária ( $d$ ), que apresentam comportamento mais estocástico – foi capaz de apresentar resultados significativos.

A análise da função impulso-resposta mostrou que a demanda imobiliária é sensível ao choque de um desvio padrão nas buscas, o que permite inferir que de fato as pesquisas na internet podem ser consideradas um antecessor na variação das vendas dos imóveis. Além disso, a decomposição do erro da previsão também aponta nesse sentido, uma vez que após um ano, 7,78% das variações na demanda imobiliária são explicadas pelas buscas online.

Outro ponto importante são as estatísticas de avaliação da previsão. Quando confrontados o uso e o não uso da variável buscas ( $s$ ), percebe-se uma leve vantagem quando a variável é empregada. Tais inferências demonstram que o Google Trends é útil na previsão da demanda, porém a causalidade de Granger não foi estatisticamente significativa para as buscas.

Além disso, outras limitações foram observadas. A escolha dos termos que possivelmente foram mais utilizados para se referir a procura por imóveis é muito abrangente e não existe uma teoria ou regra para serem definidos. Dessa forma, a escolha foi de certa forma empírica e arbitrária.

Também foi enfrentado o problema de periodicidade menor do que a indicada para variáveis relevantes para o trabalho, como o índice de preços dos imóveis FIPEZAP, que só passou a ser computado em 2012, e os índices de crédito direcionado para o setor imobiliário. Além disso, a própria demanda por imóveis

também foi computada de apenas 8 capitais brasileiras em função do mesmo problema: não abranger todo o período analisado.

De toda maneira, concluiu-se que o objetivo foi alcançado. Foi realizada uma avaliação da capacidade das buscas auxiliarem na previsão de vendas de imóveis. Os resultados que aqui foram encontrados mostram que as buscas possuem um papel relevante nas vendas futuras. Sendo assim, é possível concluir que se trata de uma ferramenta que pode balizar não somente decisões no setor imobiliário brasileiro, mas também podem ser aplicados para diversas outras áreas.

Fica claro que ainda existe muito espaço para melhorias e avanços, os quais ficarão aos encargos de futuros estudos e novos trabalhos, porém o potencial que o Google Trends oferece pode proporcionar decisões cada vez mais corretas de formadores de políticas públicas, empresas, corretoras e até mesmo indivíduos.

## BIBLIOGRAFIA

- ADKINS, Lee C. **Using Gretl for Principles of Econometrics**. 4. Ed. Oklahoma State University, 2014. Disponível em: <[http://www.learneconometrics.com/gretl/using\\_gretl\\_for\\_POE4.pdf](http://www.learneconometrics.com/gretl/using_gretl_for_POE4.pdf)>. Acesso em: 26 ago. 2016.
- ALVES, J. A.; LIMA, RICARDO CHAVES. Transmissão de preços entre os mercados de açúcar espacialmente separados no Brasil: uma análise de co-integração. **Congresso brasileiro de economia e sociologia rural**, 2010. Disponível em: <<http://www.sober.org.br/palestra/15/1036.pdf>>. Acesso em: 05 ago. 2016.
- ANDRADE, João Sousa. **Apontamentos de econometria aplicada**. Faculdade de Economia da Universidade de Coimbra, mar. 2004. Disponível em: <<http://www.academia.edu/download/32398060/econometria.pdf>>. Acesso em: 20 ago. 2016.
- BANCO CENTRAL DO BRASIL (BC). BC: Brasília, jan. 2016. Disponível em <<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>>. Acesso em: 27 abr. 2016.
- BRASIL. **Google Trends**. Mountain View: Google, 2016. Disponível em: <[www.google.com/trends](http://www.google.com/trends)>. Acesso em: 20 abr. 2016.
- BUENO, Rodrigo de Losso da Silveira. **Econometria de Séries Temporais**. 2. Ed. São Paulo: Cengage Learning, 2008.
- CÂMARA BRASILEIRA DA INDÚSTRIA DA CONSTRUÇÃO (CBIC). CBIC: Brasília, 30 abr. 2016. Disponível em: <<http://www.cbicdados.com.br/menu/mercado-imobiliario/mercado-imobiliario>>. Acesso em: 30 abr. 2016.
- CHOI, Hyunyoung; VARIAN, Hal. **Predicting the Present with Google Trends**. Anais eletrônicos. 2009. Disponível em: <[https://static.googleusercontent.com/media/www.google.com/pt-BR/googleblogs/pdfs/google\\_predicting\\_the\\_present.pdf](https://static.googleusercontent.com/media/www.google.com/pt-BR/googleblogs/pdfs/google_predicting_the_present.pdf)>. Acesso em: 05 abr. 2016.
- CHOI, Hyunyoung; VARIAN, Hal. **Predicting the Present with Google Trends**. Anais eletrônicos. 2011. Disponível em:

<<http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf>>. Acesso em: 05 abr. 2016.

D'AMURI, Francesco; MARCUCCI, Juri. "Google it!" – Forecasting the US Unemployment Rate with a Google Job Search Index. **ISER Working Paper Series**. Nov. 2009. Disponível em: <<https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2009-32.pdf>>. Acesso em: 30 abr. 2016.

DEVERA, Aaron Stephen; O'BRIEN, Joseph Michael. **Hacking Google Trends to Predict Voter Turnout**. Fordham College Rose Hill. 25 jan. 2016. Disponível em: <[http://aaronsdevera.com/public/files/01252016/hacking\\_google\\_report.pdf](http://aaronsdevera.com/public/files/01252016/hacking_google_report.pdf)>. Acesso em: 25 abr. 2016.

DICKEY, D.; FULLER, W. A. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. **Econometrica**, vol. 49, n. 4, 1981. pp. 1057-72. Disponível em: <<http://www.jstor.org/stable/1912517>>. Acesso em: 24 ago. 2016.

HANSEN, Henrik; DOORNIK, Jurgen A. An omnibus test for univariate and multivariate normality. **Economics Papers Series** (Nuffield College, University of Oxford), 1994. Disponível em: <<https://www.doornik.com/research/normal2.pdf>>. Acesso em: 08 out. 2016.

ENDERS, W. **Applied econometric time series**. United States: John Wiley & Sons, 1995. 433 p.

FAVERO, Luiz Paulo Lopes; BELFIORE, Patrícia Prado; LIMA, Gerlando A. S. Franco de. Modelos de precificação hedônica de imóveis residenciais na região metropolitana de São Paulo: uma abordagem sob as perspectivas da demanda e da oferta. **Estudos Econômicos**, São Paulo, v. 38, n. 1, p. 73-96, mar. 2008. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0101-41612008000100004&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-41612008000100004&lng=en&nrm=iso)>. Acesso em: 01 mai. 2016.

GIANNONE, Domenico; REICHLIN, Lucrezia; SMALL, David H. Nowcasting GDP and inflation: The Real-Time Informational Content of Macroeconomic Data Releases. **Working Paper Series** n. 633, mai. 2006. European Central Bank, 2006. Disponível em: <<https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp633.pdf?491207c32ca40dfb26f764d8c4e71db4>>. Acesso em: 30 abr. 2016.

GOEL, Sharad; HOFMAN, Jake M.; LAHAIE, Sebastien; PENNOCK, David M.; WATTS, Duncan J. Predicting Consumer Behavior with Web Search. Proceedings of the National Academy of Sciences (PNAS). Microeconomics and Social Systems, Yahoo! Research, 111 West 40th Street, New York, NY 10018.29 abr. 2010. Disponível em: <<http://www.pnas.org/content/107/41/17486.full.pdf>>. Acesso em: 07 abr. 2016.

- GRANGER, C. W. J. Investigating Causal Relationships by Econometric Models and Cross-Spectral Methods. **Econometrica**, vol. 37, n. 3, pp. 424-38, 1969. Disponível em: <<http://www.jstor.org/stable/1912791>>. Acesso em: 01 set. 2016.
- GUJARATI, D. **Basic Econometrics**. New York: McGraw-Hill Book Company, 2004.
- GUZMAN, Giselle. Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations. **The Journal of Economic and Social Measurement**, Vol. 36, No. 3, Dezembro de 2011. Disponível em: <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2004598](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2004598)>. Acesso em: 07 abr. 2016.
- HAUER, Mariana. **Os modelos VAR e VEC espaciais: uma abordagem bayesiana**. 2007. 201 f. Dissertação (Mestrado em Economia) – Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, Porto Alegre, 2008. Disponível em: <<http://hdl.handle.net/10183/12585>>. Acesso em: 11 abr. 2016.
- HUMPHREY, Brian D. **Forecasting Existing Home Sales using Google Search Engine Queries**. Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University. Duke University. Durham, North Carolina, 2010. Disponível em: <<https://econ.duke.edu/uploads/assets/dje/2010/100419%20Humphrey%20The%20Final.pdf>>. Acesso em: 30 abr. 2016.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). IBGE: Rio de Janeiro, 2016. Disponível em: <<http://www.ibge.gov.br/>>. Acesso em: 27 abr. 2016.
- JOHANSEN, S. Statistical Analysis of Cointegrating Vectors. **Journal of Economic Dynamics and Control**, vol. 12, n. 2-3, 1988. pp. 231-254.
- JOHANSEN, S.; JUSELIUS, K. Maximum likelihood estimation and inference on cointegration with application to the demand for money. **Oxford Bulletin of Economics and Statistics**, vol. 52, n. 2, pp. 169-209, 1990. Disponível em: <<http://nhjy.hzau.edu.cn/kech/hgjxx/Article/UploadFiles/tese/xuexiziyuan/jdwxxd/lwl/8.pdf>>. Acesso em: 17 ago. 2016.
- LALLANILLA, Marc. Can Google Predict Drug Side Effects? **Live Science**. 07 mar. 2013. Disponível em: <<http://www.livescience.com/27732-google-drug-side-effects.html>>. Acesso em: 05 abr. 2016.
- LEINWEBER, David. Big Data Gets Bigger: Now Google Trends Can Predict The Market. **Forbes**, 26 abr. 2013. Disponível em: <<http://www.forbes.com/sites/davidleinweber/2013/04/26/big-data-gets-bigger-now-google-trends-can-predict-the-market/#3228b845613a>>. Acesso em: 05 abr. 2016.
- LÜTKEPOHL, H.; KRÄTZIG, M. **Applied Time Series Econometrics**. New York: Cambridge University Press, 2004.

- MACHADO, Mari Eldionara Rosa; CERETTA, Paulo Sérgio; VIEIRA, Kelmara Mendes. A relação entre as variáveis macroeconômicas e a concessão de crédito no mercado imobiliário brasileiro. **Revista de Gestão, Finanças e Contabilidade**, v. 4, n. 3, p. 64, 2014.
- MARGARIDO, Mario A.; BUENO, Carlos R. F.; MARTINS, Vagner A.; CARNEVALLI, Luciana B. Análise dos efeitos de preço e câmbio sobre o preço do óleo de soja na cidade de São Paulo: Uma aplicação do modelo VAR. **Pesquisa & Debate**, SP, v. 15, n. 1(25), p. 69-106, 2004. Disponível em: <<http://revistas.pucsp.br/index.php/rpe/article/view/11942/8647>>. Acesso em: 07 ago. 2016.
- MARGARIDO, Mario A. Teste de Co-integração de Johansen Utilizando o SAS. **Agricultura em São Paulo**, São Paulo, v. 51, n. 1, p. 87-101, 2004. Disponível em: <[https://www.researchgate.net/profile/Mario\\_Margarido/publication/228449115\\_TestedeCo-integracao\\_de\\_Johansen\\_Utilizando\\_o\\_SAS/links/56bcb86d08ae5e7ba40f3f4b.pdf](https://www.researchgate.net/profile/Mario_Margarido/publication/228449115_TestedeCo-integracao_de_Johansen_Utilizando_o_SAS/links/56bcb86d08ae5e7ba40f3f4b.pdf)>. Acesso em: 07 set. 2016.
- MCLAREN, Nick; SHANBHOGE, Rachana. **Using Internet Search Data as Economic Indicators**. Quarterly Bulletin 2011 Q2. Bank of England. Junho, 2011. Disponível em: <<http://www.bankofengland.co.uk/publications/Documents/quarterlybulletin/qb110206.pdf>>. Acesso em: 07 abr. 2016.
- MEDEIROS, Otávio Ribeiro; DOORNIK, Bernardus Ferdinandus Nazar Van; OLIVEIRA, Gustavo Rezende. Modelando e estimando as demonstrações financeiras de uma empresa com o modelo VAR – VECM. **Brazilian Business Review**, Vitória, ES, v. 8, n. 3, p. 20 – 40, jul/set 2011. Disponível em: <[http://repositorio.unb.br/bitstream/10482/14580/1/ARTIGO\\_ModelandoEstimandoDemonstracoes.pdf](http://repositorio.unb.br/bitstream/10482/14580/1/ARTIGO_ModelandoEstimandoDemonstracoes.pdf)>. Acesso em: 20 ago. 2016.
- MENDONÇA, Mário Jorge Cardoso de. O crédito imobiliário no Brasil e sua relação com a política monetária. **Rev. Bras. Econ.**, Rio de Janeiro, v. 67, n. 4, p. 457-495, dez. 2013. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0034-71402013000400005&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71402013000400005&lng=pt&nrm=iso)>. Acesso em: 05 mai. 2017.
- O Globo. Relembre o que é Subprime. **O Globo Economia**. 23 mar. 2009. Atualizado 05 jan. 2012. Disponível em: <<http://oglobo.globo.com/economia/relembre-que-subprime-3144507>>. Acesso em: 27 abr. 2016.
- OREIRO, José Luís. **Bolhas, Incerteza e Fragilidade Financeira: Uma Abordagem Pós-Keynesiana**. UCAM – Universidade Cândido Mendes, Mestrado em Economia Empresarial. Rio de Janeiro, 2001. Disponível em: <[http://www.ie.ufrj.br/images/pesquisa/publicacoes/rec/REC%205/REC\\_5.2\\_01\\_Bolhas\\_incerteza\\_e\\_fragilidade\\_financeira\\_uma\\_abordagem\\_pos\\_keynesiana.pdf](http://www.ie.ufrj.br/images/pesquisa/publicacoes/rec/REC%205/REC_5.2_01_Bolhas_incerteza_e_fragilidade_financeira_uma_abordagem_pos_keynesiana.pdf)>. Acesso em: 30 abr. 2016.

- PAIXÃO, Luís Felipe Marques. **Mercado Imobiliário**: Uma Discussão a Respeito da Bolha Brasileira, e o Comportamento dos Consumidores Porto-Alegrenses Frente ao Cenário de Incertezas. 2014. Trabalho de Conclusão de Curso (Bacharel em Administração). – Curso de Administração, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2014.
- PAPPAS, Stephanie. Google Predicts Stock-Market Crashes, Study Suggests. **Yahoo! News**. LiveScience Senior Writer. 25 abr. 2013. Disponível em: <<https://www.yahoo.com/news/google-predicts-stock-market-crashes-study-suggests-131040530.html>>. Acesso em: 05 abr. 2016.
- PREIS, Tobias; MOAT, Helen Susannah; STANLEY, H. Eugene. Quantifying Trading Behavior in Financial Markets Using *Google Trends*. **Nature**. Scientific Reports. Fev. 2013. Disponível em: <<http://www.nature.com/articles/srep01684>>. Acesso em: 05 abr. 2016.
- PREIS, Tobias; REITH, Daniel; STANLEY, H. Eugene. Complex Dynamics of Our Economic Life on Different Scales: Insights from Search Engine Query Data. **The Royal Society Publishing**. 15 nov. 2010. Disponível em: <<http://rsta.royalsocietypublishing.org/content/roypta/368/1933/5707.full.pdf>>. Acesso em: 12 abr. 2016.
- RIVERA, Edward Bernard Bastiaan de Rivera y. Preços Hedônicos: Teoria e Aplicação no Setor Imobiliário na Cidade de São Paulo (1995-2004). **Revista Jovens Pesquisadores** Ano IV, No. 7, jul./dez. 2007. Disponível em: <<http://www.mackenzie.br/dhtm/seer/index.php/jovenspesquisadores/article/view/900/413>>. Acesso em: 27 abr. 2016.
- SHIMSHONE, Yair; EFRON, Niv; MATIAS, Yossi. **On the Predictability of Search Trends**. Technical report, Google, 2009. 17 ago. 2009. Disponível em: <[http://static.googleusercontent.com/media/research.google.com/pt-BR//archive/google\\_trends\\_predictability.pdf](http://static.googleusercontent.com/media/research.google.com/pt-BR//archive/google_trends_predictability.pdf)>. Acesso em: 07 abr. 2016.
- SCHMIDT, Torsten; VOSEN, Simeon. Forecasting Private Consumption: Survey-based Indicators vs. Google Trends. **Ruhr Economic Papers**. Ruhr-Universität Bochum (RUB), Department of Economics. Germany, 2009. Disponível em: <[http://repec.rwi-essen.de/files/REP\\_09\\_155.pdf](http://repec.rwi-essen.de/files/REP_09_155.pdf)>. Acesso em: 05 abr. 2016.
- SILVA, Daniel Santos. **Análise do Comportamento Imobiliário na Cidade de Porto Alegre**: Um Estudo sob o Foco da Teoria Econômica. Trabalho de Conclusão de Curso (Bacharel em Administração). – Curso de Administração, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2014. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/117323/000965743.pdf?sequence=1>>. Acesso em: 30 abr. 2016.
- SIMS, Christopher A. **Macroeconomics and reality**. *Econometrica: Journal of the Econometric Society*, v. 48, n. 1, p. 1-48, 1980. Disponível em: <<http://www.jstor.org/stable/1912017>>. Acesso em: 16 jul. 2016.



- SILVER, Nate. How Google Search Can Predict Hockey Ticket Sales. **Five ThirtyEight**. 24 abr. 2015. Disponível em: <<http://fivethirtyeight.com/datalab/how-google-searches-can-predict-hockey-ticket-sales/>>. Acesso em: 24 abr. 2016.
- STOCK, James H.; WATSON, Mark W. Vector Autoregressions. **The Journal of Economic Perspectives**, v. 15, n. 4, p. 101-115, 2001. Disponível em: <<http://hassler-j.iies.su.se/courses/macroi/papers/StockWatsonJEP.pdf>>. Acesso em: 23 ago. 2016.
- TAVARES, Fernando António de Oliveira; MOREIRA, António Carrizo; PEREIRA, Elisabeth Teixeira. Assimetria de Informação no Mercado Imobiliário: Uma Revisão da Literatura. **Revista Universo Contábil**, ISSN 1809-3337, FURB, Blumenau, v. 8, n. 2, p. 146-164, abr./jun., 2012. Disponível em: <<http://proxy.furb.br/ojs/index.php/universocontabil/article/view/2704/1970>>. Acesso em: 30 abr. 2016.
- TODA, Hiro Y.; YAMAMOTO, Taku. Statistical inference in vector autoregressions with possibly integrated processes. **Journal of econometrics**, v. 66, n. 1, p. 225-250, 1995.
- WEBB, G. Kent. **Internet Search Statistics as a Source of Business Intelligence: Searches on Foreclosure as an Estimate of Actual Home Foreclosures**. San Jose State University, SJSU Scholar Works. 01 jan. 2009. Disponível em: <[http://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1014&context=mis\\_public](http://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1014&context=mis_public)>. Acesso em: 24 abr. 2016.
- WOOLDRIDGE, Jeffrey M. **Introductory Econometrics: A Modern Approach**. 5 Ed. South-Western, Cengage Learning, 2012.
- WU, Lynn; BRYNJOLFSSON, Erik. **The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales**. 2 dez. 2009. Disponível em: <<https://ssrn.com/abstract=2022293>>. Acesso em: 30 abr, 2016.

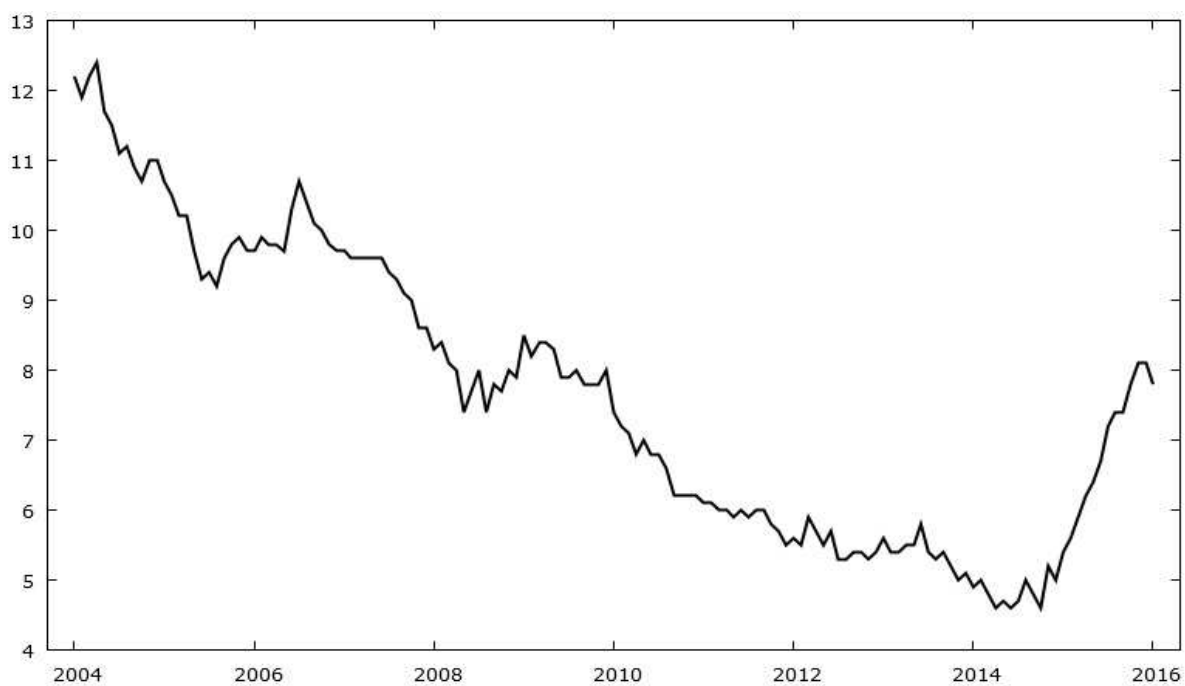
## APÊNDICE

A seguir segue a estatística descritiva das séries, usando as observações de janeiro de 2004 até janeiro de 2016, e seus respectivos gráficos temporais.

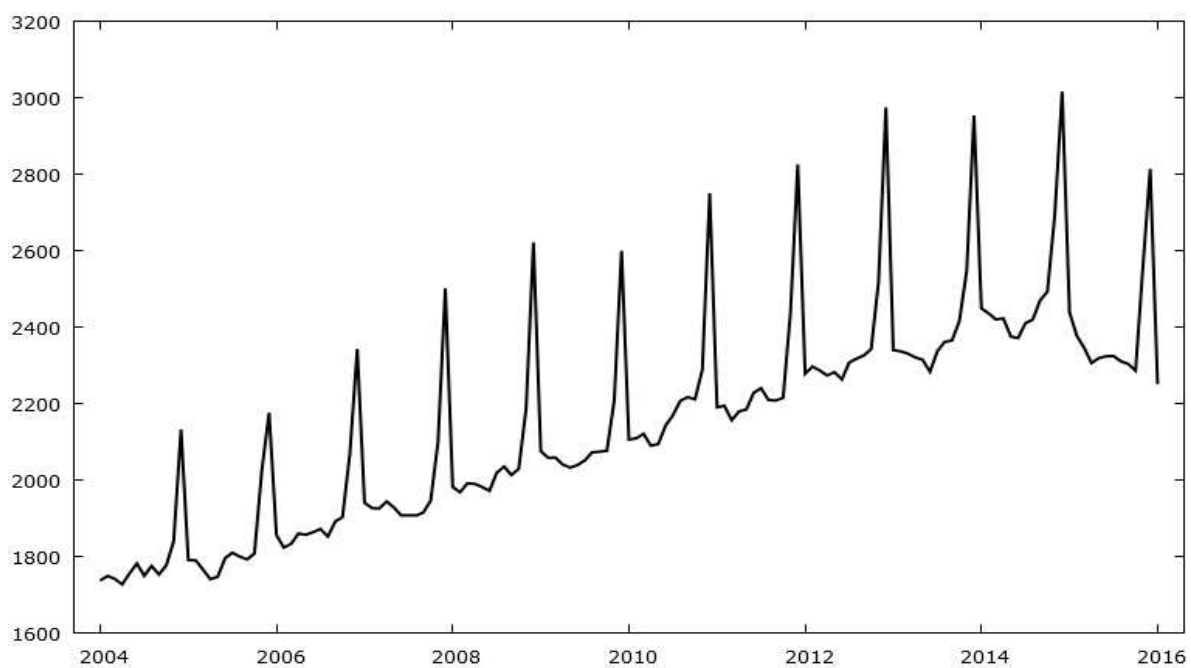
TABELA A 1: Estatísticas descritivas das séries

	<i>u</i>	<i>w</i>	<i>i</i>	<i>s</i>	<i>d</i>
Média	7,6503	2156,6	12,300	48,476	5178,4
Mediana	7,7000	2169,8	11,700	53,000	4951,0
Mínimo	4,6000	1727,2	7,1100	12,000	2029,0
Máximo	12,400	3016,0	19,750	100,00	10100,0
Desvio padrão	2,0883	279,76	3,2723	24,122	1845,2
C. V.	0,27297	0,12972	0,26604	0,49762	0,35633
Enviesamento	0,33740	0,57031	0,56071	0,059608	0,40910
Curtose Ex.	-1,0190	0,24597	-0,35769	-1,2848	-0,60401
Percentil de 5%	4,8300	1750,9	7,2990	16,000	2652,6
Percentil de 95%	11,170	2671,8	19,166	84,700	8334,0
Intervalo interquartil	3,9000	402,14	3,9850	45,000	3112,5

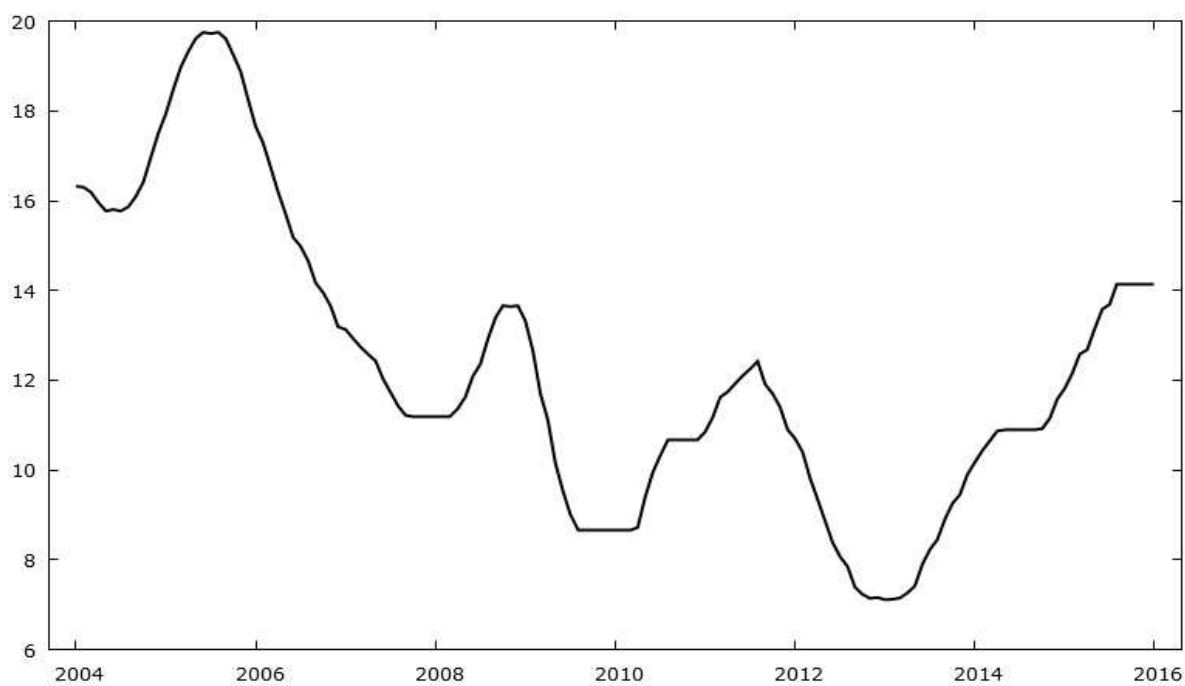
Fonte: Elaboração própria.

GRÁFICO A. 1: Taxa de desemprego  $u$ 

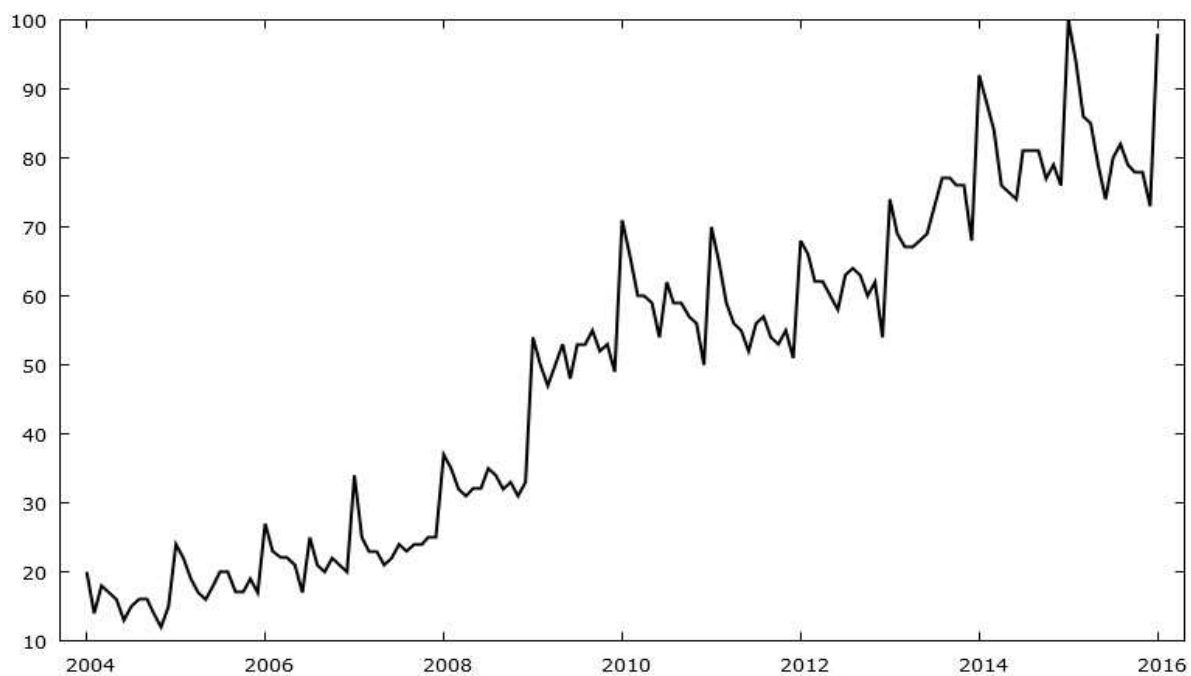
Fonte: PME-IBGE

GRÁFICO A. 2: Nível de salários  $w$ 

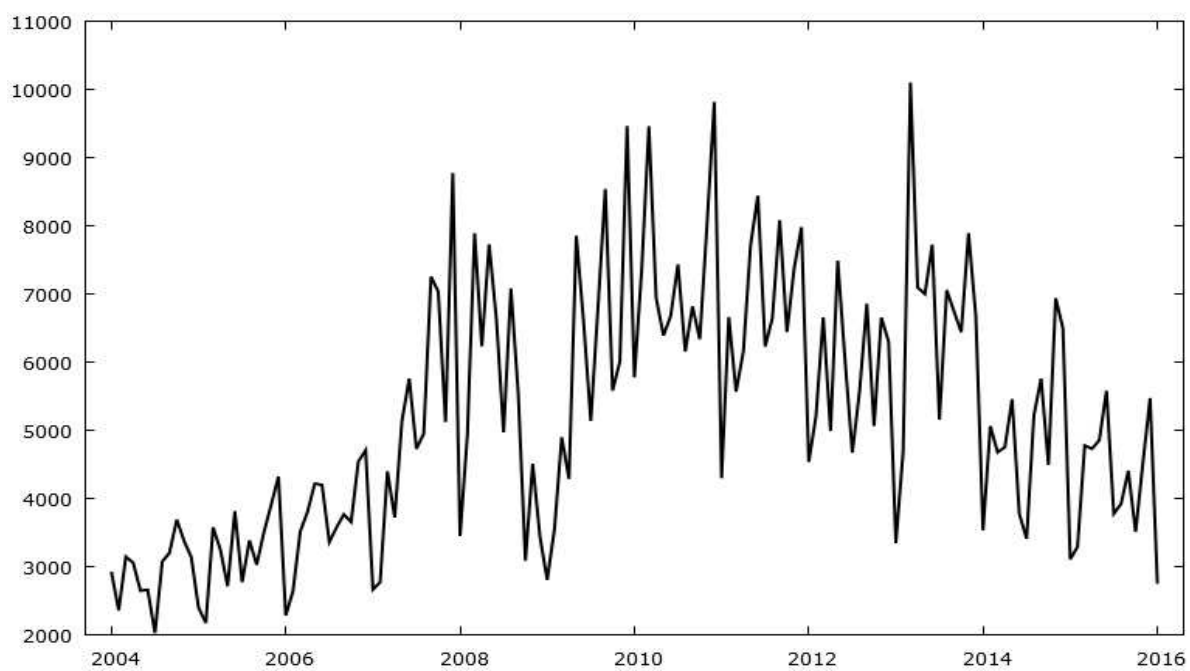
Fonte: PME-IBGE.

GRÁFICO A. 3: Taxa de juros  $i$ 

Fonte: BACEN.

GRÁFICO A. 4: Volume de buscas  $s$ 

Fonte: Google.

GRÁFICO A. 5: Demanda imobiliária *d*

Fonte: CBIC.