



Artigo

Informações e Política Econômica: um teste para aperfeiçoamento de erros de previsão a partir da utilização do *Google Trends*

Information And Economic Policy: a test for improvement of forecast errors from the use of Google Trends

Informaciones y Política Económica: un teste para el perfeccionamiento de errores de previsión a partir de la Utilización del Google Trends

Cláudio D. Shikida¹, Renato Moreira Byrro², Márcio Antônio Salvato³, Ari Francisco de Araujo Junior⁴

¹ Filiação institucional. Doutor em Economia pela Universidade Federal do Rio Grande do Sul, Brasil, e docente do Instituto Brasileiro de Mercado de Capitais

Correspondência: E-mail: ClaudioDS@ibmecmg.br

Instituição de correspondência

Rua Rio Grande do Norte, 300 – Funcionários – Belo Horizonte – 30.130-13 – MG – Brasil

² Filiação institucional. Bacharel em Economia pelo Instituto Brasileiro de Mercado de Capitais, Brasil.

Correspondência: Instituição de correspondência

Rua Rio Grande do Norte, 300 – Funcionários – Belo Horizonte – 30.130-13 – MG – Brasil

³ Filiação institucional. Doutor em Economia pela Fundação Getúlio Vargas, Brasil, e docente do Instituto Brasileiro de Mercado de Capitais, Brasil.

Correspondência: E-mail: marcio.salvato@gmail.com

Instituição de correspondência

Rua Rio Grande do Norte, 300 – Funcionários – Belo Horizonte – 30.130-13 – MG – Brasil

⁴ Filiação institucional. Mestre em Economia pela e docente do

Correspondência: E-mail: arifaj@ibmecmg.br

Instituição de correspondência

Rua Rio Grande do Norte, 300 – Funcionários – Belo Horizonte – 30.130-13 – MG – Brasil



Resumo

Este artigo tem por objetivo replicar os testes de Choi & Varian (2009a) para variáveis da economia brasileira analisando modelos de previsão de séries temporais do tipo ARIMA com o objetivo de avaliar se a inclusão dos índices de pesquisas do *Google Trends* a tais modelos pode reduzir os erros de previsão, ou seja, poderiam ser usados como *leading indicators*. Para isto, foram utilizadas séries relacionadas ao mercado de trabalho e de crédito. Para o mercado de trabalho, a previsão da série de requerimentos de benefícios do seguro-desemprego foi aprimorada a partir da inclusão do *Google Trends* aos modelos. O mesmo não foi observado no caso da previsão da taxa de desemprego. No mercado de crédito, foram testadas duas séries: concessão de financiamentos vinculados a cartões de crédito e de financiamentos imobiliários, sendo que nenhuma delas teve a previsão aprimorada pelo *Google Trends*.

Palavras-chave: Google Trends, Previsões, Indicadores Antecedentes, Política Econômica, Políticas Públicas.

Abstract

This article's aim is to replicate the tests of Choi & Varian (2009a) for Brazilian economic variables, analyzing ARIMA time series models and evaluating the reduction in forecasting errors when introducing a Google Trends variable in order to check if this is a good leading indicator. We used time series related to the labor market and credit market. For the first, the forecasting of the unemployment insurance had a better performance after the inclusion of the Google Trends. Regarding the unemployment rate, the performance was not good. For the credit market we used two series: concession of lendings linked to credit card and mortgages. Their forecasting were not better after the inclusion of the Google Trends.

Keywords: Google Trends, Forecasting, Leading Indicators, Economic Policy, Public Policies

Resumen

Este artículo tiene por objetivo replicar los testes de Choi & Varian (2009a) para variables de la economía brasileña analizando modelos de previsión de series temporales del tipo ARIMA con el objetivo de evaluar se la inclusión de los índices de la búsquedas de *Google Trends* a tales modelos puede reducir los errores de previsión, o sea, podrían ser usados como *leading indicators*. Para eso, fueron utilizadas series relacionadas al mercado de trabajo y de crédito. Para el mercado de trabajo, la previsión de la serie de requerimientos de beneficios del seguro desempleo fue mejorada a partir de la inclusión del *Google Trends* a los modelos. El mismo no fue observado en el caso de la previsión de la tasa de desempleo. En el mercado de crédito, fueran testadas dos series: concesión de financiamientos vinculados a tarjetas de crédito y de financiamientos inmobiliarios, siendo que ninguna de ellas tuvo la previsión mejorada por el *Google Trends*.

Palabras-clave: Google Trends, Previsiones, Indicadores Antecedentes, Política económica, Políticas Públicas.

Introdução

A internet é um dos principais meios de acesso à informação na sociedade atual¹, sendo o mecanismo de busca do *Google* o mais popular para busca de dados dispostos na rede mundial – 47%³, 49%² e 63%³ de *market-share* nas buscas mundiais em 2005, 2006 e 2008, respectivamente. Hipoteticamente, se um indivíduo busca na internet, por exemplo, informações sobre um modelo de automóvel, há uma boa chance de que ele esteja interessado em adquirir aquele bem num futuro próximo; ou, se são buscadas informações sobre o requerimento de benefício do seguro-desemprego, é possível que o indivíduo tenha sido demitido recentemente. Os índices de buscas no *Google* ofereceriam, assim, informações sobre as ações dos agentes econômicos com uma defasagem muito pequena – se comparada aos índices oficiais –, tendo em vista a publicação semanal dos dados do *Google Trends*.

O *Google Trends* é uma ferramenta do sistema de pesquisas do *Google* que, grosso modo, informa o volume de pesquisas para uma dada palavra-chave num certo período de tempo, em uma região específica. Por meio dele, é possível, por exemplo, saber a evolução das pesquisas que os usuários de internet fizeram no *Google* sobre os candidatos de uma eleição a cargos públicos, sobre um novo produto lançado no mercado, um clube de futebol, ou qualquer outro tema. Os índices do *Google Trends* são calculados pela razão entre o volume de pesquisas com uma dada palavra-chave e o volume total de pesquisas com todas as palavras-chave pesquisadas na mesma região e mesmo período de tempo. O índice é semanal e calculado em relação à média observada no mês de janeiro de 2004 (igual a um), quando as séries do *Trends* têm início. O *Google* também disponibiliza uma ferramenta denominada “*Insights for Search*”, que utiliza a mesma base de dados do *Google Trends* e possui agregadores de informações. O *Insights* agrupa automaticamente as palavras-chave da base de dados em categorias e subcategorias como “Alimentos e bebidas”, “Finanças e Seguros”, etc. No entanto, não são informadas ao usuário final quais palavras-chave compõem cada categoria. Neste trabalho serão utilizados tanto os índices do *Trends* quanto do *Insights*.

A utilização do *Google Trends* para fins de política pública não é uma novidade. Por meio de sua divisão, *Google.org*, pesquisas são feitas para fins de prevenção de doenças (gripe e dengue, atualmente) e catástrofes naturais. O acompanhamento dos casos de gripe nos EUA e sua evolução por esta ferramenta teve início em 2006 tem sido utilizado pelo

¹ No Brasil, entretanto, o acesso à internet é bem mais restrito que em países desenvolvidos, onde já houve pesquisas com o *Google Trends*, como EUA e Alemanha, que possuíam, em 2007, 62% e 71% dos domicílios com acesso à internet, respectivamente, contra 17% no Brasil (vide Tabela A.1 do Anexo).

² Pesquisa *Nielsen Online*, 2006. Disponível em: <<http://searchenginewatch.com/2156451>>. Acesso em: 25 jun. 2010.

³ Pesquisa *Nielsen Online*, 2008. Disponível em: <<http://searchenginewatch.com/3632382>>. Acesso em: 25 jun. 2010.

Center for Disease Control (CDC) como uma espécie de indicador antecedente para a doença. Atualmente o CDC e a Google trabalham em parceria neste aspecto⁴.

Em 2009, o economista chefe da Google, Hal Varian, publicou um artigo intitulado “*Predicting the present with Google Trends*” (Choi & Varian, 2009a), no qual é destacado o potencial desta ferramenta para a previsão de variáveis econômicas ou como *leading indicator*. Além disso, são apresentados alguns exercícios econométricos, sugerindo o uso do *Google Trends* na previsão de variáveis.

Este trabalho tem por objetivo replicar os testes de Choi e Varian (2009a) para variáveis da economia brasileira e testar estatisticamente a hipótese de melhoria dos erros de previsão de variáveis econômicas, de tal forma a comparar modelos de previsão de séries temporais ARMA e avaliar se a inclusão dos índices de pesquisas do *Google Trends* aos modelos pode reduzir os erros de previsão. Para isto, foram utilizadas séries relacionadas ao mercado de trabalho e de crédito.

1. Revisão de Literatura

Choi e Varian (2009a) foram os primeiros a testar a hipótese de que os dados de pesquisas do *Google Trends* estão correlacionados com atividades dos agentes econômicos e que, portanto, podem ser utilizados como auxiliares em previsões econométricas para variáveis econômicas reais. O título do artigo, ‘*Predicting the Present*,’ refere-se à proposta dos autores de se usar termos de pesquisa do *Google Trends* como uma espécie de indicadores das atividades econômicas no presente, antecipando movimentos que seriam publicados posteriormente (com defasagem maior) por estatísticas oficiais sobre os mesmos eventos. O primeiro exemplo usado pelos autores foi a série de vendas de automóveis da marca Ford nos EUA (Estados Unidos da América), entre janeiro/2004 e agosto/2008. A metodologia aplicada é bastante simples, e consiste em comparar dois modelos para a mesma série:

- a) O primeiro, um modelo AR sazonal, em logaritmo, que usa as vendas do mês anterior e também as de doze meses passados para prever a venda de uma dada série (y). Basicamente:

⁴ Ver Carmody (2013).

$$\log(y_t) = \log(y_{t-1}) + \log(y_{t-12}) + e_t \quad (1)$$

- b) O segundo modelo incorpora termos de pesquisa do *Google Trends* relacionados à variável estudada (no exemplo, vendas da Ford nos EUA):

$$\log(y_t) = \log(y_{t-1}) + \log(y_{t-12}) + x_t^{(1)} + e_t \quad (2)$$

em que:

x é o índice de pesquisas do *Google Trends* com a palavra-chave escolhida e (1) é a primeira semana do mês⁵.

A comparação entre os modelos foi feita utilizando o erro absoluto médio das previsões realizadas dentro da amostra de cada modelo. O artigo mostrou que a inclusão da variável do *Google Trends* neste modelo foi capaz de reduzir o erro de previsão em 3%.

Outros exemplos analisados neste artigo, sob a mesma metodologia, mostraram resultados similares. Por exemplo, para o setor de varejo automobilístico, as estimações com a inclusão do índice de pesquisas do *Google Trends* geraram reduções de 15% e 18% nos erros, mas as previsões permaneceram muito otimistas em relação à variável real a partir do início de 2008, indicando que os modelos não foram capazes de captar o arrefecimento que a economia norte-americana experimentava. Também foram obtidas melhorias nos erros de previsão com modelos de previsão de vendas de outras montadoras além da Ford. No segmento de vendas de imóveis nos EUA, os autores alcançaram uma redução do erro absoluto médio em 12% com a incorporação do *Google Trends*.

Em Choi e Varian (2009b), os mesmos autores realizaram uma pesquisa sobre o mercado de trabalho norte-americano. Utilizando a mesma metodologia do artigo anterior, foi analisada a eficácia do *Google Trends* na redução dos erros de previsão de pedidos de seguro-desemprego. Foram estimados dois modelos com períodos de tempo distintos: desde janeiro/2004, cobrindo toda a série disponível no *Trends*, e desde dezembro/2007, assumindo a hipótese de que houve quebra estrutural no mercado de trabalho norte-americano a partir desta data, devido à recessão econômica, ambas seguindo até junho/2009. Utilizando duas categorias de agrupamento de palavras-chave do *Google Trends*, 'Jobs' e 'Welfare & Unemployment', os autores conseguiram reduzir os erros de previsão, considerando os últimos seis meses dentro da amostra, em 15,74% para o modelo de período mais longo e em 12,9% para o modelo com período desde dezembro/2007.

⁵ Os dados do *Trends* são semanais.

Askitas & Zimmermann (2009), seguindo o exercício de Choi & Varian (2009b), estudaram o mercado de trabalho alemão. Neste mercado, em particular, o índice de desemprego é medido ao longo de um mês cheio e divulgado quase sempre ao final de cada mês. Assim, os autores optaram por utilizar os índices do *Trends* nas duas últimas semanas do mês (imediatamente anteriores ao anúncio) e as duas primeiras do mês seguinte (imediatamente posteriores ao anúncio). Por meio de testes de causalidade de Granger, analisou-se o índice de desemprego alemão no período entre janeiro/2004 e abril/2009, e concluiu-se que a utilização das duas últimas semanas imediatamente anteriores à divulgação do índice oficial é estatisticamente aceitável como previsão. Isto torna os dados do *Trends* úteis, pois permitem antecipar os movimentos que serão publicados pelo governo a respeito do mercado de trabalho alemão. Os autores, no entanto, não utilizam nenhum critério para avaliar a melhoria nos erros de previsão dos modelos com o *Google Trends*.

No estudo de Suhoj (2009) também foi utilizado o teste de causalidade de Granger como metodologia para avaliar a existência de relação entre o *Trends* e as variáveis econômicas de setores da economia israelense. O autor ressalta que “*Granger causality tests provide evidence that corresponding indices contain cyclical components which conform with cycles of economic growth*” (Suhoj, 2009:29). Além disso, foi estimado no artigo um modelo indicador de probabilidade de recessão, utilizando os dados do *Google Trends* como *leading indicator*. Também foram feitas referências a outros estudos, na área de *marketing*, segundo os quais as intenções de compra dos consumidores podem explicar as visitas a páginas na internet, e que “[...] *intentions are transformed into purchases within a short time horizon*” (Suhoj, 2009:7). O autor conclui que os resultados para os dados de Israel não rejeitam a hipótese de que os índices do *Google Trends* podem auxiliar como *leading indicator* para o crescimento econômico corrente.

A metodologia do presente trabalho se aproxima da que foi utilizada nos estudos de Choi e Varian (2009a, 2009b), mas a especificação dos modelos não será *ad hoc* e seguirá critérios de avaliação tradicionais quanto ao número de defasagens em modelos ARIMA. A comparação entre a qualidade de previsão dos modelos será feita por meio do Erro Absoluto Médio (EAM) e também do Coeficiente de Desigualdade de Theil.

2. Metodologia

Neste artigo segue-se Choi e Varian (2009a, 2009b). Assim, por exemplo, suponha o modelo utilizado por estes autores em seu estudo sobre as vendas de automóveis da Ford, nos EUA, entre 2004 e 2008:

$$\log(y_t) = \log(y_{t-1}) + \log(y_{t-12}) + e_t \quad (3)$$

em que:

y_t : variável em análise no período t.

e_t : termo aleatório (ruído branco).

Incorporando a variável de um índice do *Google Trends* que, espera-se, esteja relacionado à série analisada:

$$\log(y_t) = \log(y_{t-1}) + \log(y_{t-12}) + x_t^{(s)} + e_t \quad (4)$$

em que:

y_t : variável em análise no período t.

$x_t^{(s)}$: índice do *Google Trends* referente à semana s do mês t, sendo $1 \leq s \leq 4$, assumindo valores inteiros neste intervalo⁶.

e_t : termo aleatório (ruído branco).

Para avaliar a importância da inclusão da variável $x_t^{(s)}$: como indicador antecedente, compara-se o erro de previsão dos dois modelos usando os critérios Erro Absoluto Médio (EAM) e Coeficiente de Desigualdade de Theil (CDT):

$$EAM = \frac{1}{T} \sum_{t=1}^T (y_t^* - y_t) \quad (5)$$

$$CDT = U = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (y_t^* - y_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^T (y_t^*)^2 + \frac{1}{T} \sum_{t=1}^T (y_t)^2}} \quad (6)$$

em que:

y_t^* : previsão da série para o instante de tempo t.

y_t : série de tempo original no instante de tempo t.

T: número de períodos previstos dentro da amostra:

$$0 \leq U \leq 1$$

⁶ Também podem ser utilizadas as médias dos índices semanais em cada mês; neste caso, será indicado com a letra 'm' no sobrescrito da variável: x_t^m .

O EAM foi o critério adotado pelos estudos anteriores sobre o *Google Trends*. Além deste, neste artigo, utiliza-se também o CDT, já que este fornece a decomposição do erro de previsão e análise da participação proporcional do viés (U^B), variância (U^V) e erro não-sistemático (U^S), da seguinte maneira (Pindick & Rubinfeld, 2009):

$$\frac{1}{T} \sum_{t=1}^T (y_t^* - y_t)^2 = (\bar{y}^* - \bar{y})^2 + (\sigma_{y^*} - \sigma_y)^2 + 2(1 - \rho) \sigma_{y^*} \sigma_y \quad (7)$$

em que:

$\bar{y}^*, \bar{y}, \sigma_{y^*}, \sigma_y$ são a média e o desvio-padrão das séries y^* e y , respectivamente, e ρ é o coeficiente de correlação entre ambas as séries⁷.

As proporções do erro podem ser definidas como:

$$U^B = \frac{(\bar{y}^* - \bar{y})^2}{\frac{1}{T} \sum (y_t^* - y_t)^2} \quad (8)$$

$$U^V = \frac{(\sigma_{y^*} - \sigma_y)^2}{\frac{1}{T} \sum (y_t^* - y_t)^2} \quad (9)$$

$$U^S = \frac{2(1-\rho) \sigma_{y^*} \sigma_y}{\frac{1}{T} \sum (y_t^* - y_t)^2} \quad (10)$$

A proporção do viés (U^B) é um indicador de erro sistemático do modelo predictor, já que mede o desvio entre a média de y^* e y . A proporção de variância (U^V) indica a capacidade do modelo em replicar o grau de variância da série analisada. O último componente (U^S), como o próprio nome diz, mede o erro não-sistemático. A soma dos três indicadores deve ser igual a um e, idealmente, espera-se que U^B e U^V sejam iguais a zero e que, conseqüentemente, U^S seja igual a 1, para qualquer valor de U . (Pindyck & Rubinfeld, 2009)

Desta forma, a decomposição fornecida pelo CDT permite-nos analisar se o *Google Trends* melhora o ajustamento do modelo para fins de previsão reduzindo o erro sistemático e/ou de variância. Se o *Google Trends* puder captar as mudanças na variabilidade das séries estudadas, seria um instrumento muito interessante de auxílio a previsões em períodos voláteis, notadamente mais difíceis para se trabalhar.

⁷ Isto é: $\rho = (1/\sigma_{y^*} \sigma_y T) \sum (y_t^* - \bar{y}^*) (y_t - \bar{y})$.

3. Dados

O *Google Trends* não retorna o volume absoluto de pesquisas realizadas com uma dada palavra-chave, mas sim um índice que representa a razão entre o volume de pesquisas com uma palavra-chave e o volume total de pesquisas com todas as palavras-chave pesquisadas na mesma região e mesmo período de tempo. O índice é semanal e calculado em relação à média observada no mês de janeiro de 2004 (igual a um), quando as séries do *Trends* têm início. Se for observada, por exemplo, uma semana indicando valor 1,5, significa que a participação do volume de pesquisas com aquela palavra-chave, naquela semana, foi 50% maior que a média de janeiro de 2004.

A tabela 4.1 mostra os dados utilizados neste trabalho (o período amostral para todas as séries é de Janeiro/2004 a Dezembro/2009).

Tabela 4.1 – Descrição das variáveis utilizadas

Variável	Descrição
Seguro-Desemprego	<p>Número de requerimentos de indenização do Seguro Desemprego (tendo como referência o mês de demissão do requerente)</p> <p>Número de requerimentos de indenização do Seguro Desemprego (tendo como referência o mês de requerimento do benefício)</p> <p>Fonte: Ministério do Trabalho e Emprego (SAEG.net)</p> <p>Google Trends: palavra-chave "seguro desemprego"</p>
Taxa de Desemprego	<p>Taxa de desocupação (% da população economicamente ativa)</p> <p>Fonte: Pesquisa Mensal de Emprego (IBGE)</p> <p>Google Insights: categorias "Sociedade" → "Trabalho e Sindicatos"</p>
Cartão de Crédito	<p>Volume de concessões de crédito em operações de Cartão de Crédito (recursos livres)</p>

	Fonte: Banco Central do Brasil Google Insights: categorias "Finanças e Seguros" → "Crédito e empréstimo" → "Cartões de Crédito"
Crédito Imobiliário	Operações de crédito imobiliário com recursos livres Fonte: Banco Central do Brasil Google Insights: categoria "Imobiliário" → "Financiamento de imóveis residenciais"

Fonte: Elaboração própria.

O *Google* também disponibiliza uma ferramenta denominada "*Insights for Search*", que utiliza a mesma base de dados do *Google Trends* e possui agregadores de informações. O *Insights* agrupa, automaticamente, as palavras-chave da base de dados em categorias e subcategorias como "Alimentos e bebidas", "Finanças e Seguros", etc. No entanto, não são informadas ao usuário final quais palavras-chave compõem cada categoria. Neste artigo, foram utilizados dados tanto do *Google Trends* quanto do *Insights for Search*.

4. Estimações e Resultados

Para cada série analisada, foram estimados seis modelos de séries temporais: utilizando a própria série e, em seguida, incorporando a média dos índices semanais do *Google Trends* no mês e o índice da primeira, segunda, terceira e quarta semanas do mês, individualmente.

A especificação dos modelos seguiu o seguinte esquema:

1. Verificação de estacionariedade das séries;
2. A definição dos termos AR e MA dos modelos foi realizada com auxílio do algoritmo de otimização de modelos do pacote econométrico *Easyreg*, com base nos critérios *Akaike* e *Schwarz*;
3. Algumas alterações na especificação dos modelos foram realizadas posteriormente, utilizando o pacote econométrico *Eviews*, seguindo a

metodologia *Box-Jenkins*, para melhor adequação dos termos AR e MA ao correlograma (autocorrelação e autocorrelação parcial) das séries;

Espera-se que o sinal do coeficiente estimado para a variável *Google Trends* seja positivo, ou seja, que um aumento no volume de pesquisas com a palavra-chave “seguro-desemprego” esteja associado a uma elevação nos requerimentos de benefícios do seguro desemprego, por exemplo. Após a estimação dos modelos, foram geradas previsões dentro da amostra para um período de 12 meses (janeiro/2009 a dezembro/2009), e então foram comparados os resultados entre o modelo original, sem o *Google Trends*, e os demais. Para cada série analisada, foi selecionado o melhor modelo com *Google Trends*, pelo critério de EAM e CDT, e traçado um gráfico entre a previsão deste modelo, a previsão do modelo sem *Google Trends* e a série original.

As séries de taxa de desemprego, cartão de crédito e crédito imobiliário apresentaram raiz unitária no teste *Augmented Dickey-Fuller*⁸ (ADF) e, desta forma, os modelos relacionados a estas três séries foram estimados em primeira diferença, mas as previsões para o ano de 2009 e os critérios de comparação EAM e CDT foram calculados para as séries em nível. Os testes ADF, com a série em nível, foram realizados com intercepto e tendência; com a série em primeira diferença, foram realizados somente com intercepto. Os resultados dos testes estão dispostos na Tabela 4.1.

Tabela 4.1 – Resultados dos testes ADF para estacionariedade das séries

Séries	Valor-p ADF
	Em nível
Taxa de desemprego	0,1065*
Seguro Desemprego	0,0000
Cartão de Crédito	0,7273*
Crédito Imobiliário	0,3469*

Fonte: cálculos próprios.

* Em primeira diferença, o valor-p do teste ADF foi menor que 0,001.

⁸ A hipótese nula do teste ADF é de presença de raiz unitária; para valor-p menor que 0,01, rejeitamos a hipótese nula com 99% de confiança.

Todos os resultados estão dispostos em tabelas, da seguinte forma: a primeira coluna se refere ao modelo original estimado para a série, sem a variável do *Google Trends*. Os resultados de EAM e CDT se referem à previsão dentro da amostra. As colunas seguintes se referem aos resultados do mesmo modelo, mas agora incorporando a variável do *Google Trends* indicada na primeira linha, que pode ser: a média das semanas do mês, a primeira, segunda, terceira ou quarta semana do mês. Na segunda e terceira linhas apresentamos o valor-p e o sinal de cada coeficiente do *Google Trends*, respectivamente. Nas linhas seguintes, seguem os resultados para o EAM e CDT, bem como seus componentes (viés, variância e erro não-sistemático). A numeração das equações apresentadas nesta seção será modificada para facilitar a sua identificação. Para a taxa de desemprego, por exemplo, teremos a sigla 'txdes' seguida do número do modelo dentre os doze estimados no total para esta montadora; por exemplo: 'txdes1'.

4.1 Taxa de Desemprego

Todos os coeficientes do *Google Trends* tiveram sinal positivo nos modelos estimados para a série da taxa de desemprego, mas o nível de significância estatística foi baixo. O modelo 'txdes5', que incorporou o índice do *Google Trends* da terceira semana foi o que apresentou os melhores resultados: queda de 8% no EAM e no CDT. A decomposição do CDT permaneceu praticamente inalterada.

Tabela 4.1.1 – Taxa de Desemprego

Modelos:	txdes1	txdes2	Txdes3	txdes4	txdes5	txdes6
Variável Trends		média mensal	Semana 1	semana 2	semana 3	semana 4
P-valor Trends		0,5024	0,8359	0,7616	0,1969	0,596
Sinal Trends		+	+	+	+	+
EAM	0,861288	0,833167	0,863537	0,856509	0,788249	0,847756
CDT	0,059103	0,057464	0,05931	0,058822	0,054358	0,058432
<i>Viés</i>	0,902407	0,890083	0,901071	0,9004	0,88513	0,892981
<i>Variância</i>	0,07411	0,084968	0,075392	0,075877	0,087251	0,08401

<i>Não-sistem.</i>	0,023484	0,024948	0,023537	0,023723	0,027619	0,023009
--------------------	----------	----------	----------	----------	----------	----------

Fonte: cálculo próprio.

As equações do modelo original (txdes1) e do modelo com *Google Trends* que apresentou melhor desempenho (txdes5) seguem detalhadas abaixo:

$$d(\hat{y}_t) = 0,85 \cdot d(y_{t-12}) - 0,87 \cdot d(\varepsilon_{t-12}) \quad (\text{txdes1})$$

p-valor (0,0000) (0,0000)
R²: 0,72 Akaike: 0,23 Schwarz: 0,30

$$d(\hat{y}_t) = 0,84 \cdot d(y_{t-12}) - 0,87 \cdot d(\varepsilon_{t-12}) + 0,0038 \cdot d(x_t^{s3}) \quad (\text{txdes5})$$

p-valor (0,0000) (0,0000) (0,1969)
R²: 0,73 Akaike: 0,24 Schwarz: 0,34

em que:

\hat{y}_t = nº estimado da taxa de desemprego no mês t
 x_t^{s3} = índice *Google Trends* na terceira semana do mês t

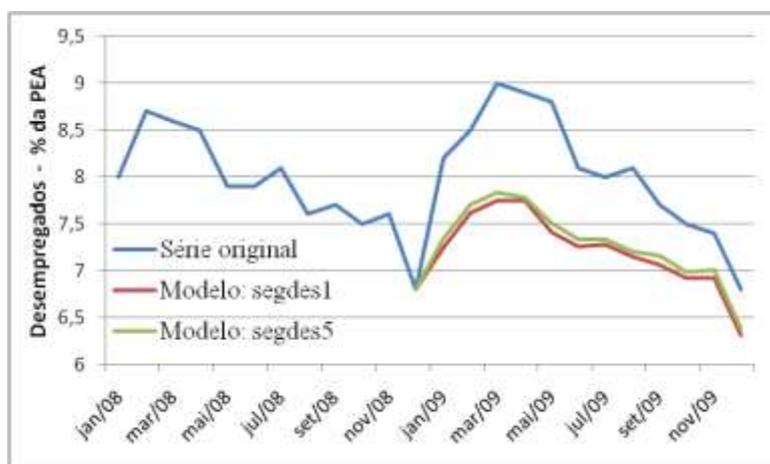


Gráfico 4.1.1 – Taxa de desemprego

Fonte: IBGE (série original); cálculo próprio (modelos)

Das variáveis ligadas ao mercado de trabalho, o desempenho do *Google Trends* foi melhor na série de requerimentos do seguro desemprego, conforme os dados apresentados a seguir.

4.2 Seguro Desemprego

Para a série de requerimentos do seguro desemprego por data de demissão, em nenhum dos modelos estimados obtivemos coeficientes com sinal positivo para os índices de pesquisas. Já para a série calculada pela data de requerimento do benefício, o modelo 'segdes6', que utiliza o índice do *Google Trends* referente à quarta semana do mês, proporcionou uma queda de 18% no EAM e de 21% no CDT. Apesar disto, observamos um aumento de 6,5 p.p. e de 1,3 p.p. , respectivamente, na participação do viés e da variância, simultaneamente a uma queda de quase 8 p.p. no erro não-sistemático.

Tabela 4.2.1 – Requerimentos do Seguro Desemprego (por data de requerimento)

Modelos:	segdes1	segdes2	segdes3	segdes4	segdes5	segdes6
Variavel Trends		média mensal	semana 1	semana 2	semana 3	semana 4
P-valor Trends		0,0288	0,4822	0,0962	0,0788	0,0052
Sinal Trends		+	+	+	+	+
EAM	58520,91	51218,08	55067,51	54613,57	50541,45	48031,27
CDT	0,055278	0,047776	0,052403	0,051608	0,048318	0,043431
<i>Viés</i>	0,008139	0,039516	0,000135	0,012601	0,023367	0,073081
<i>Variância</i>	0,638676	0,70509	0,705447	0,669719	0,703795	0,652177
<i>Não-sistem.</i>	0,353185	0,255395	0,294418	0,317679	0,272838	0,274742

Fonte: cálculo próprio.

As equações dos modelos 'segdes1' e 'segdes6' seguem abaixo:

$$\hat{y}_t = 1503170 + 0,34 \cdot y_{t-2} + 0,63 \cdot y_{t-12} + 0,26 \cdot \varepsilon_{t-1} \quad (\text{segdes1})$$

p-valor (0,6021) (0,0003) (0,0000) (0,0521)
 R²: 0,76 Akaike: 24,13 Schwarz: 24,27

$$\hat{y}_t = 521985,7 + 0,34 \cdot y_{t-2} + 0,63 \cdot y_{t-12} + 0,26 \cdot \varepsilon_{t-1} + 66799,14 \cdot x_t^{s4}$$

(segdes6)
 p-valor (0,0000) (0,0397) (0,0000) (0,2564) (0,0052)
 R²: 0,79 Akaike: 24,03 Schwarz: 24,20

em que:

\hat{y}_t = volume estimado de requerimentos de seguro desemprego no mês t
 x_t^{s4} = índice *Google Trends* na quarta semana do mês t

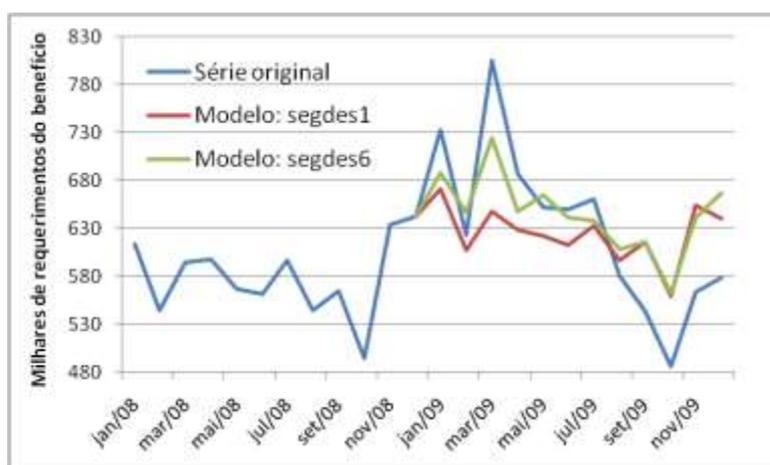


Gráfico 4.2.1 – Requerimentos de benefício do seguro desemprego

Fonte: Ministério do Trabalho (série original); cálculo próprio (modelos)

4.3 Cartão de Crédito

O modelo 'cartao3', que incorpora o índice do Google Trends referente à primeira semana do mês, trouxe queda de 4% no EAM e de 2,4% no CDT, sem modificações expressivas na participação do viés, variância e erro não-sistemático.

Tabela 4.3.1 – Concessões de crédito em operações de cartão de crédito

Modelos:	cartao1	cartao2	cartao3	cartao4	cartao5	cartao6
Variável Trends		média mensal	semana 1	semana 2	semana 3	semana 4
P-valor Trends		0,4249	0,2731	0,4577	0,7404	0,6425
Sinal Trends		+	+	+	+	+
EAM	706858	693234,7	677875,7	702889,4	702830,4	698842,4
CDT	0,036459	0,036069	0,035596	0,36266	0,036415	0,036159
Viés	0,138498	0,151495	0,144228	0,150385	0,143916	0,144536
Variância	0,453166	0,435075	0,459413	0,446854	0,441506	0,430004
Não-sistem.	0,408336	0,41343	0,396359	0,402761	0,414578	0,425459

Fonte: cálculo próprio.

Aqui temos:

$$d(\hat{y}_t) = 164374 + 0,39 \cdot d(\varepsilon_{t-1}) \quad (\text{cartao1})$$

p-valor (0,0000) (0,0000)

R²: 0,35 Akaike: 29,44 Schwarz: 29,51

$$d(\hat{y}_t) = 158933,7 - 0,67 \cdot d(\varepsilon_{t-1}) + 3158,9 \cdot d(x_t^{s1})$$

(cartao3)

p-valor (0,0000) (0,0000) (0,2731)

R²: 0,36 Akaike: 29,45 Schwarz: 29,55

em que:

\hat{y}_t = volume estimado de operações de financiamento com cartão de crédito no mês t

x_t^{s1} = índice *Google Trends* na primeira semana do mês t

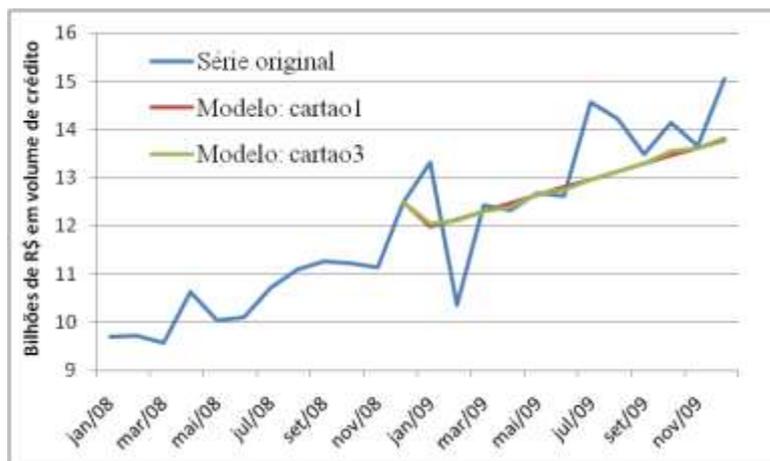


Gráfico 4.3.1 – Volume de operações de crédito com cartão de crédito

Fonte: Banco Central do Brasil (série original); cálculo próprio (modelos)

Nos testes para a série de crédito imobiliário, cujos resultados são apresentados na subseção a seguir, foram obtidos resultados superiores à série de cartão de crédito.

4.4 Crédito Imobiliário

O modelo 'credimob2' trouxe pequena contribuição para a previsão da série de concessão de empréstimos habitacionais: queda de quase 10% no EAM e no CDT. No entanto, podemos observar que a participação da variância aumentou em 10 p.p., acompanhada de uma queda pela metade da participação do erro não-sistemático.

Tabela 4.4.1 – Crédito imobiliário

Modelos:	credimob1	credimob2	credimob3	credimob4	credimob5	credimob6
Variável Trends		média mensal	semana 1	semana 2	semana 3	semana 4
P-valor Trends		0,4457	0,4747	0,4675	0,4703	0,4596
Sinal Trends		+	+	+	+	+
EAM	51702,77	46707,91	47139,8	47044,69	47135,69	47046,09
CDT	0,182106	0,164135	0,165751	0,165403	0,165747	0,165435

<i>Viés</i>	0,530144	0,510774	0,513092	0,512801	0,513542	0,513588
<i>Variância</i>	0,31011	0,412659	0,404437	0,40641	0,404626	0,406446
<i>Não-sistem.</i>	0,159746	0,076567	0,08247	0,080789	0,081831	0,079966

Fonte: cálculo próprio.

Para crédito:

$$d(\hat{y}_t) = 0,40 \cdot y_{t-2} - 0,92 \cdot d(\varepsilon_{t-1}) + 0,44 \cdot d(\varepsilon_{t-4}) \quad (\text{credimob1})$$

p-valor (0,0014) (0,0000) (0,0000)

R²: 0,50 Akaike: 22,66 Schwarz: 22,76

$$d(\hat{y}_t) = 0,38 \cdot y_{t-2} - 0,92 \cdot d(\varepsilon_{t-1}) + 0,44 \cdot d(\varepsilon_{t-4}) + 2,38 \cdot d(x_t^m) \quad (\text{credimob2})$$

p-valor (0,0027) (0,0000) (0,0000) (0,4457)

R²: 0,51 Akaike: 22,68 Schwarz: 22,81

em que:

\hat{y}_t = volume estimado de operações de financiamento com cartão de crédito no mês t

x_t^{s1} = índice *Google Trends* na primeira semana do mês t

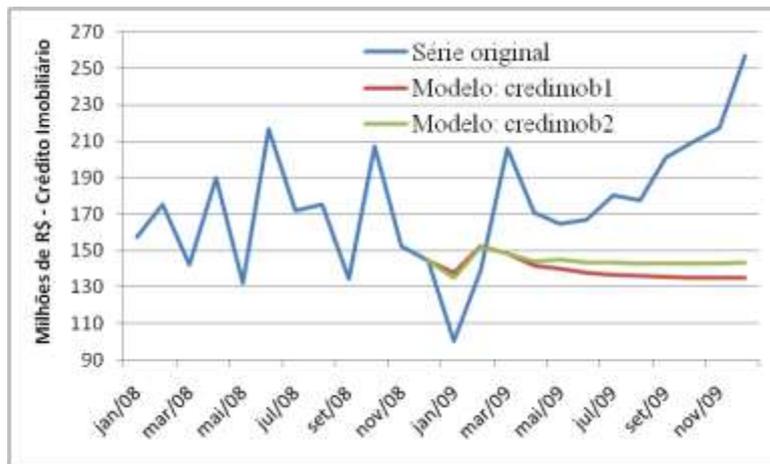


Gráfico 4.4.1 – Volume de operações de financiamento imobiliário

Fonte: Banco Central do Brasil (série original); cálculo próprio (modelos)

Apesar da queda de aproximadamente 10% nos erros de previsão, o coeficiente estimado da variável *Google Trends* foi estatisticamente significativo a um nível de confiança de apenas 45% no modelo 'credimob2'.

Conclusões

Os resultados obtidos demonstram que, em alguns casos – como nas séries de requerimentos do seguro desemprego –, a inclusão do *Google Trends* nos modelos de previsão ARMA traz melhorias do ponto de vista do erro de previsão, mas em muitos não há um ganho expressivo. Para a série de requerimentos do seguro desemprego, foi observada queda do EAM e CDT em torno de 20%, mas com queda relativa de 8 p.p. na participação do erro não-sistemático.

Para a série da taxa de desemprego, alcançou-se redução de 8% nos erros de previsão, associada a uma queda de 0,5 p.p. na participação do erro não-sistemático. Na série de concessões de financiamentos imobiliários, houve redução de 10% nos erros de previsão, mas com queda de 8 p.p. na participação do erro não-sistemático, e para a série de financiamentos em cartão de crédito houve queda de apenas 4% nos erros de previsão.

Analisando a decomposição do coeficiente CDT, foi possível observar que a inclusão do *Google Trends*, para as séries e períodos analisados neste trabalho, tende a contribuir mais na redução do viés do que da variância na composição do erro de previsão, em termos relativos. Esperava-se que, por captar melhor as oscilações dos mercados em períodos mais voláteis, como o do ano estudado, de 2009⁹, seria possível obter uma redução proporcionalmente maior na participação da variância.

O fato de o Brasil possuir uma renda mais baixa e distribuída de forma desigual, em comparação a outros países em que o *Google Trends* já foi estudado, como EUA e Alemanha (Tabela A.1 do Anexo), e ter uma penetração menor da internet nas camadas da população de baixa renda (Tabela A.2 do Anexo), faz com que o potencial do *Google Trends* para previsão de séries brasileiras fique ainda prejudicado.

⁹ A economia brasileira viveu um período mais volátil durante o ano de 2009, devido à quebra do banco *Lehman Brothers*, nos EUA, e o conseqüente agravamento da crise financeira internacional no final do ano de 2008.

Além disso, talvez a maior limitação seja a própria construção dos índices do *Google Trends*, pelo fato de ser divulgada a razão entre as buscas com uma palavra-chave e o volume total de buscas na mesma região e período, e não o volume absoluto de pesquisas com a palavra-chave. O índice de pesquisas sobre um tema qualquer pode ter caído num determinado período não porque há menos pessoas pesquisando a respeito, mas porque houve um aumento atípico no volume total de pesquisas na mesma região e período analisados, seja por qualquer causa. Nestes casos, a informação que o *Google Trends* antecipa aos modelos econométricos é de que há uma menor procura por aquele tópico, quando na realidade isto pode não estar acontecendo.

Apesar do desempenho do *Google Trends* ter sido satisfatório em algumas séries, para este trabalho, em outros casos, o desempenho não foi tão satisfatório. A ferramenta parece ter um potencial muito expressivo e espera-se que ainda poderá contribuir bastante para a análise, observação e acompanhamento dos movimentos dos mercados e dos agentes econômicos para fins de análise de mercado ou de políticas públicas. Entretanto, mais testes são necessários para se verificar o poder de previsão da ferramenta. Além disso, a modificação na forma de cálculo dos índices divulgados, como mencionado acima, seria fundamental para a exploração do potencial do *Google Trends*.

Referências

- Askitas, Nikolaos,. & Zimmermann, Klaus. (2009). Google econometrics and unemployment forecasting. Discussion Paper nº 4201, IZA – Institute for the Study of Labor.
- Carmody, Tim. (2013). Google and CDC show US flu epidemic among worst in a decade. *The Verge*. Acessado em 22/06/2013, de <http://www.theverge.com/2013/1/10/3861538/google-cdc-show-us-flu-epidemic-among-worst-in-decade>
- Choi, H., & Varian, H. (2009a). *Predicting the present with Google Trends*. Technical Report, Google.
- Choi, Hyunyoung., & Varian, Hal. (2009b). *Predicting inicial claims for unemployment benefits*. Technical Report, Google.

- Efron, Niv., Matias, Yossi., & Shimshoni, Yair. (2009). *On the predictability of search trends*. Google, Israel Labs.
- Ginsberg, Jeremy. Mohebbi, Matthew H., Patel, Rajan S., Brammer, Lynnette., Smolinski, Mark S., & Brilliant, Larry. (2009). Detecting influenza epidemics using search query data. *Nature*, v. 457, p. 1012-1014.
- Pindyck, Robert S., & Rubinfeld, Daniel L., (1991). *Econometric models and economic forecasts*. 3. ed. Boston: McGraw-Hill.
- Suhoy, Tanya. (2009). Query indices and 2008 downturn: Israeli data. Discussion Paper nº 2009.06, Research Department, Bank of Israel.

Recebido em 01/12/2012
Revisado em 13/01/2013
Aceito 22/01/2013