2064
TEXTO PARA DISCUSSÃO

ipea

APLICAÇÃO DO MODELO FATORIAL DINÂMICO PARA PREVISÃO DA RECEITA TRIBUTÁRIA NO BRASIL

Mário Jorge Mendonça
Luis Alberto Medrano

ipea

2064

TEXTO PARA DISCUSSÃO

Rio de Janeiro, março de 2015

APLICAÇÃO DO MODELO FATORIAL DINÂMICO PARA PREVISÃO DA RECEITA TRIBUTÁRIA NO BRASIL

Mário Jorge Mendonça[1]
Luis Alberto Medrano[2]

1. Técnico de Planejamento e Pesquisa da Diretoria de Estudos e Políticas Macroeconômicas (Dimac) do Ipea.
E-mail: mario.mendonca@ipea.gov.br

2. Assistente de Pesquisa da Dimac do Ipea e professor da Universidade Federal Rural do Rio de Janeiro (UFRRJ).
E-mail: lmedrano10@msn.com

Governo Federal

Secretaria de Assuntos Estratégicos da Presidência da República
Ministro Roberto Mangabeira Unger

ipea Instituto de Pesquisa Econômica Aplicada

Fundação pública vinculada à Secretaria de Assuntos Estratégicos da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente
Sergei Suarez Dillon Soares

Diretor de Desenvolvimento Institucional
Luiz Cezar Loureiro de Azeredo

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia
Daniel Ricardo de Castro Cerqueira

Diretor de Estudos e Políticas Macroeconômicas
Cláudio Hamilton Matos dos Santos

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais
Rogério Boueri Miranda

Diretora de Estudos e Políticas Setoriais de Inovação, Regulação e Infraestrutura
Fernanda De Negri

Diretor de Estudos e Políticas Sociais, Substituto
Carlos Henrique Leite Corseuil

Diretor de Estudos e Relações Econômicas e Políticas Internacionais
Renato Coelho Baumann das Neves

Chefe de Gabinete
Ruy Silva Pessoa

Assessor-chefe de Imprensa e Comunicação
João Cláudio Garcia Rodrigues Lima

Ouvidoria: http://www.ipea.gov.br/ouvidoria
URL: http://www.ipea.gov.br

Texto para Discussão

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

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2015

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

ISSN 1415-4765

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

CDD 330.908

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Assuntos Estratégicos da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

JEL: H20; H22; C32.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO 7

2 COMPOSIÇÃO DA CTBB E REVISÃO DA LITERATURA 10

3 O MODELO FATORIAL 13

4 MLD 18

5 BASE DE DADOS 19

6 RESULTADOS 19

7 COMENTÁRIOS FINAIS 23

REFERÊNCIAS 24

APÊNDICE A 26

APÊNDICE B 29

APÊNDICE C 34

SINOPSE

Este artigo tem por objetivo estimar o modelo fatorial dinâmico (MFD) bayesiano para previsão de arrecadação de uma amostra de tributos representativa da carga tributária brasileira, com dados mensais para o período de 2001 a 2013. O emprego do modelo fatorial possibilita reduzir a dimensionalidade do elevado número de tributos, levando em consideração as informações contidas nas relações existentes entre eles e permitindo a identificação dos fatores não correlacionados que trazem informação relevante subjacente à dinâmica dos tributos. Além disso, o componente sazonal das séries é modelado endogenamente, permitindo a obtenção de estimativas mais bem ajustadas aos dados e predições mais confiáveis – uma vez que a sazonalidade é uma característica marcante de certas séries de tributos. Confrontamos as previsões obtidas por meio do MFD com aquelas geradas pelo modelo linear dinâmico (MLD) aplicado para cada imposto separadamente e verificamos que o modelo fatorial traz ganhos consideráveis em termos de eficiência e previsão.

Palavras-chave: carga tributária; sazonalidade; fator dinâmico; filtro de Kalman e amostrador de Gibbs.

ABSTRACT

This article aims to estimate the dynamic factor model for prediction tax receipts in Brazil using monthly data for the period 2001-2013. The factorial model allows to reduce the dimensionality of the high number of taxes taking into account the information contained in the existing interrelations between them and allowing to identify only the relevant information through the variables named factors. Further, in our model the seasonal component of the series of taxes is treated endogenously. This procedure permits to obtain better data fitting and more reliable predictions – once seasonality is a hallmark of certain series of tributes. We confront the predictions of the factorial model with those generated by linear dynamic model applied to each tribute separately and found that the factor model brings considerable gains in terms of efficiency and prediction.

Keywords: gross tax burden; seasonality; dynamic factor model; Kalman filter and Gibbs sampling.

1 INTRODUÇÃO

A carga tributária bruta brasileira (CTBB) é dada pela soma dos recursos que o governo recolhe compulsoriamente do setor privado. Seu aumento eleva, salvo exceções,[1] os recursos à disposição do governo, permitindo a diminuição do endividamento público e/ou aumentos na oferta de bens e serviços públicos à população. A ação da CTBB é, portanto, um determinante crucial das dinâmicas de diversas outras grandezas macroeconômicas relevantes.

Embora pareça razoável supor – como primeira aproximação, pelo menos – que as arrecadações dos diferentes tributos que compõem a CTBB acompanhem o movimento geral da economia (Santos, Ribeiro e Gobetti, 2008), é fácil perceber que as heterogeneidades existentes entre eles fazem com que os resultados de modelos agregativos sejam usualmente pouco precisos. As seguidas mudanças na legislação tributária, por sua vez,[2] fazem com que haja bons motivos para se acreditar que as elasticidades e sazonalidades possam variar com o tempo. Assim, um instrumento econométrico que objetive modelar os vários componentes da CTBB deve não apenas lidar com a alta dimensionalidade inerente aos variados tipos de impostos mas também saber tratar as características sazonais específicas a cada um deles.

Não é trivial, entretanto, modelar o comportamento dinâmico da CTBB. Ela é, na verdade, um complexo agregado de dezenas de tributos diferentes, cujas arrecadações dependem de diversas variáveis diferentes e, portanto, apresentam dinâmicas – e sazonalidades – sensivelmente diferentes. Além disso, a modelagem econométrica da CTBB e de seus componentes é muito dificultada pelas seguidas mudanças em alíquotas e bases de incidência de tributos preexistentes – e, vez por outra, também pela eliminação,

1. Note-se que nem todo recurso recolhido compulsoriamente da sociedade é receita do governo. As contribuições para o Fundo de Garantia do Tempo de Serviço (FGTS), por exemplo, são recolhidas compulsoriamente da sociedade – e por isso entram no cálculo da carga tributária bruta (CTB) –, mas não são "receitas públicas", visto que são de propriedade dos trabalhadores. Por outro lado, nem toda receita do governo é recolhida compulsoriamente da sociedade. Quando o cidadão paga entrada em um museu público, por exemplo, está aumentando a receita do governo, mas não a CTB. Daí que aumentos na CTB não necessariamente implicam aumentos nas receitas públicas, e aumentos nas receitas públicas não necessariamente implicam aumentos na CTB. Convém destacar ainda que, apesar disso, a maior parte das receitas públicas brasileiras é de natureza tributária, e a maior parte da CTBB consiste em receitas públicas.

2. Rezende, Oliveira e Araújo (2008, cap. 2) fazem um excelente resumo das inúmeras mudanças na legislação tributária brasileira no período 1998-2006.

temporária ou não, desses últimos e/ou a criação de novos tributos – que tem caracterizado a administração tributária do país no período pós-Plano Real.[3]

Os estudos implementados anteriormente têm seguido três caminhos distintos. Inicialmente, optou-se pela análise da CTBB segundo um ponto de vista agregado. Para isso, Mendonça, Medrano e Sachsida (2011) fizeram uso do modelo linear dinâmico (MLD), com coeficientes variáveis para uma série trimestral agregada da CTBB no período 1995-2009. Seguindo o mesmo modelo, a análise do comportamento da CTBB foi feita também com base no desempenho individual de um conjunto de tributos[4] que existem na economia brasileira e formam uma amostra fortemente representativa da CTBB no período de 1995 a 2010 (Mendonça, Medrano e Sachsida, 2013). Esta amostra pode ser entendida como uma *proxy* da CTBB, uma vez que responde por cerca de 80% da arrecadação tributária no Brasil. A análise econométrica da CTBB, tendo em vista a avaliação de cada tributo tomado separadamente, vem dar continuidade a uma sequência de trabalhos, já implementados no Brasil, cujos objetivos são a análise e a projeção da CTBB.

Uma deficiência com relação aos dois tipos de abordagens citados, sobretudo para a gestão da política econômica, é que, no caso do exercício de previsão, é necessário se conhecer de antemão o valor das variáveis exógenas do modelo para o período escolhido, o que naturalmente contribui para diminuir a sua confiabilidade. Foi nesse sentido que Mendonça, Santos e Martins (2009) estimaram o modelo fatorial dinâmico (MFD), utilizando dados mensais para o período de 1995 a 2007. O MFD incorpora, no conjunto de variáveis a serem modeladas, as variantes macroeconômicas, que, tal como as receitas tributárias, passam a ser tratadas como endógenas. Portanto, a partir da aplicação do MFD, não é necessário estabelecer uma trajetória para qualquer variável para fazer previsão da arrecadação tributária.

Outro ponto importante é que o uso do MFD permite percepções importantes por meio da redução de dimensionalidade (isto é, de aumento no grau de agregação) do modelo. Se, de algum modo, a informação contida nas relações entre os vários

3. E mesmo antes disso, como bem apontam Rezende, Oliveira e Araújo (2008).

4. Os tributos formam a receita da União, estados e municípios e abrangem impostos, taxas, contribuições e empréstimos compulsórios. Por exemplo, o Imposto de Renda (IR) é um tributo, assim como Contribuição para o Financiamento da Seguridade Social (Cofins).

tributos puder ser passada para um conjunto menor de variáveis, nossa habilidade de identificar a "verdadeira" estrutura dos dados será melhor. Com efeito, quanto menor for a dimensionalidade do modelo mais fácil será entendê-lo.

Este estudo tem como objetivo fazer uso do MFD para previsão da arrecadação tributária no Brasil. Entretanto, este trabalho tem pelo menos dois fatores de inovação em relação ao estudo de Mendonça, Santos e Martins (2009). Primeiro, expandimos a amostra de modo a incluir informação. Esta amostra se compõe de dados mensais de 2001 a 2013. Segundo, além de reestimar o modelo proposto por Mendonça, Santos e Martins (2009), estimamos outro, onde os fatores comuns são modelados por variáveis exógenas. Por fim, comparamos as previsões do modelo fatorial com as obtidas pelo modelo linear dinâmico (MLD).

Talvez por conta das dificuldades mencionadas, no que diz respeito a modelar a arrecadação tributária, poucos trabalhos com especificações econométricas explícitas sobre esses dados foram publicados nos últimos anos.[5] Esses textos diferem consideravelmente entre si – trabalhando com frequências, amostras e técnicas econométricas diferentes –, mas nenhum deles procura explorar as informações contidas nas relações entre as arrecadações dos vários tributos.

Diferentemente do que em geral é feito na literatura (Sáfadi e Pena, 2007), o componente sazonal das diferentes séries é modelado endogenamente, permitindo obter um melhor ajustamento e previsões mais precisas sobre a dinâmica futura dos impostos. Aplicamos, finalmente, procedimentos de inferência bayesiana – Kim e Nelson (1999); West e Harrison (1997); Carter e Kohn (1994); Gamerman e Lopes (2007), entre outros – que, entre outras vantagens, permitem resultados mais confiáveis em contextos em que a dimensão temporal da base de dados é pequena em relação ao espaço da dimensionalidade dos parâmetros – além de outras dificuldades, dificilmente tratáveis por meio de procedimentos "frequencistas" (Gelman *et al.*, 2003, p. 696).

Em relação aos resultados obtidos, verificou-se que, de acordo com os critérios de *performance* de cálculo adotados, o MFD traz ganhos substanciais de eficiência, tendo mostrado, na análise fora da amostra para o ano de 2013, que o erro de previsão foi

5. Hernández (1998), Portugal e Portugal (2001), Melo (2001), Carvalho (2001) e Siqueira (2002) deram importantes contribuições a essa literatura. Santos, Ribeiro e Gobetti (2008) apresentam uma resenha detalhada da mesma última.

menor que 10% para a maioria dos tributos analisados. Como forma de testar a robustez do modelo fatorial, foi feita uma estimativa com dados até 2011, fazendo previsão para o ano de 2012. Por todos os critérios de previsão adotados e para todos os tributos, o MFD obteve desempenho superior ao MLD.[6]

Além desta introdução, o texto inclui seis seções. A seção 2 apresenta uma descrição sucinta acerca da composição da CTBB no período 1995-2008. Em seguida, as seções 3 e 4 apresentam uma descrição do MFD e do MLD, respectivamente. A seção 5 descreve a base de dados. Os resultados e a análise dos resultados do modelo são mostrados na seção 6. Por fim, a seção 7 apresenta as principais conclusões do artigo.

2 COMPOSIÇÃO DA CTBB E REVISÃO DA LITERATURA

Conforme já assinalado, a CTBB é composta de um número elevado de impostos. A seguir, será feito um resumo da estrutura tributária nacional.[7] Em termos gerais, a CTBB é formada por quatro grandes agregados de tributos, quais sejam, *i)* impostos sobre produtos; *ii)* outros impostos ligados à produção; *iii)* impostos sobre a renda, o patrimônio e o capital (IRPCs); e *iv)* contribuições previdenciárias.

São sete os principais impostos sobre produto:

- Imposto sobre Circulação de Mercadorias e Serviços (ICMS), estadual;
- Cofins;
- Imposto sobre Produtos Industrializados (IPI);
- Imposto sobre Serviços (ISS), municipal;
- Imposto sobre Importações (II);

6. Para mais detalhes, ver Mendonça, Medrano e Sachsida (2011), Hernández (1998) e Portugal e Portugal (2001).

7. Para uma descrição sucinta da estrutura tributária no Brasil, ver Santos, Ribeiro e Gobetti (2008).

- Imposto sobre Operações de Crédito, Câmbio e Seguros (IOF); e

- Contribuição de Intervenção do Domínio Econômico (Cide).[8]

Apenas o ICMS e a Cofins, somados, responderam por mais de três quartos de todos os impostos sobre produtos e por pouco menos de um terço da CTBB total em 2006. Juntos, os impostos sobre produtos – cujo valor está implícito no preço dos produtos e, portanto, muitas vezes passa despercebido pelo consumidor – respondem por pouco mais de 40% da CTBB.

São quatro os principais tributos brasileiros sobre a renda, o patrimônio e o capital, quais sejam, o IR, a Contribuição Social sobre o Lucro Líquido das pessoas jurídicas (CSLL), o Imposto sobre Propriedade de Veículos Automotores (IPVA) (estadual) e o Imposto Predial Territorial Urbano (IPTU) (municipal). A arrecadação do IR é dividida em seis componentes – o IR pessoa física (IRPF); o IR pessoa jurídica (IRPJ); o IR retido na fonte (IRRF) sobre os rendimentos do trabalho; o IRRF sobre ganhos de capital; o IRRF sobre remessas de divisas ao exterior; e os outros tipos de IRRF – e atingiu pouco menos de 60,5% da arrecadação total dos IRPCs, enquanto a CSLL, o IPTU e o IPVA representaram, respectivamente, 11,9%, 5,9% e 5,5% do total dos IRPCs nesse mesmo ano. Juntos, os IRPCs representam pouco menos de 30% da CTBB.

As contribuições previdenciárias (CPs) respondem por cerca de um quarto da CTBB. Como ocorre com os demais componentes da CTBB, a lista de tributos relacionados às CPs é bastante extensa,[9] ainda que possam ser divididas em: *i)* contribuições de patrões e empregados para o INSS; *ii)* contribuições para o Programa de Integração Social-Programa de Formação do Patrimônio do Servidor Público (PIS-Pasep); *iii)* contribuições para o FGTS; e *iv)* contribuições para a aposentadoria dos funcionários públicos (pagas pelos próprios e pelo governo). A maior parte incide sobre a folha de pagamentos de empresas e administrações públicas (isto é, sobre o "rendimento do fator de produção trabalho").

8. A Cide incide sobre a importação e a comercialização de gasolina e suas correntes, *diesel* e suas correntes, querosene de aviação e outros querosenes, óleos combustíveis, gás liquefeito de petróleo (GLP), inclusive o derivado de gás natural e de nafta, e álcool etílico combustível.

9. Para uma lista que exclui as contribuições para o PIS-Pasep, que foram reclassificadas como contribuições previdenciárias no Sistema de Contas Nacionais (SCN) 2005, ver o anexo 8 de IBGE (2004).

Por fim, o último e quantitativamente menos importante componente da CTBB é o item outros impostos ligados à produção, composto basicamente por taxas diversas (taxa de fiscalização das telecomunicações ou taxa de vigilância sanitária, por exemplo) e por contribuições incidentes sobre a folha de pagamento de empresas e administrações públicas (como para o salário família e para o Sistema S).[10]

Percebe-se, assim, a flagrante heterogeneidade dos componentes individuais da CTBB. A tabela 1 apresenta a amostra dos tributos usada neste estudo – composta por vários dos principais componentes dos quatro grandes tipos de tributo já mencionados – bem como o percentual que cada um deles representa em relação ao total. Esta amostra composta basicamente de tributos federais cuja base de dados é gerenciada pela Secretaria da Receita Federal (SRF), mas inclui também o ICMS, estadual, com dados mensais disponibilizados pelo Conselho Nacional de Política Fazendária (Confaz). Estes vinte impostos representam pouco menos de 80% do total de tributos arrecadado pelas administrações públicas brasileiras.

TABELA 1
Tributos: percentual em relação ao total

Imposto	%	Imposto	%
ICMS	26,52	II	2,10
Previdência Social	20,17	IRPF	1,30
Cofins	12,77	IPI – vinculados à importação	1,27
IRPJ – demais empresas	6,08	IOF	1,19
IRRF – rendimento de trabalho	5,91	IRPJ – entidades financeiras	1,15
Outros tributos	4,64	IRRF – remessa para o exterior	1,09
Contribuição sobre Lucro de Pessoas Jurídicas (CLPJ)	3,68	IRRF – outros rendimentos	0,74
IRRF – rendimento de capital	3,66	IPI-Fumo	0,64
PIS/Pasep	3,58	IPI-Automóveis	0,60
IPI – outros	2,36	IPI-Bebidas	0,56

Elaboração dos autores.

10. Sistema S é o nome pelo qual ficou convencionado chamar o conjunto de nove instituições de interesse de categorias profissionais, estabelecidas pela Constituição brasileira. Entre as mais conhecidas podemos citar o Serviço Nacional de Aprendizagem Industrial (Senai), o Serviço Nacional de Aprendizagem Comercial (Senac), o Serviço Social da Indústria (Sesi) etc.

3 O MODELO FATORIAL

O MFD pertence à classe de modelos de variáveis latentes, que são ferramentas importantes para a análise de dados multivariados. Uma razão dessa importância é a redução de dimensionalidade que proporcionam. Num conjunto grande de variáveis é natural a possibilidade de existir alta colinearidade entre elas, pois existem apenas algumas fontes de informação comuns contidas nos dados. Se, de algum modo, a informação presente nas inter-relações de muitas variáveis puder ser passada, com boa aproximação, para um conjunto muito menor, a habilidade de ver a estrutura nos dados será melhor. O objeto da análise fatorial é reduzir as dimensões de modo que sejam apenas usadas as fontes de informação mais importantes. A ideia é que a informação contida nas variáveis que são mais correlacionadas pode ser mapeada por um subconjunto menor de variáveis latentes denominado fator. Os fatores devem ser não correlacionados, pois representam fontes independentes de informação. A redução de dimensionalidade aumenta a eficiência computacional resultante da ausência de correlação e da redução de dimensionalidade. Dado que os fatores são ortogonais, sua matriz de covariância não condicional é diagonal.

Além disso, essa redução de dimensionalidade faz com que os modelos fatoriais sejam parcimoniosos quando comparados com outros que analisam dados multivariados, sem proporcionar essa redução de dimensionalidade, como os modelos VAR (Hamilton, 1993), que são muito usados na literatura econométrica e lidam melhor quando o conjunto restrito de variáveis é restrito. Mais detalhes a respeito das vantagens do uso de modelos fatoriais podem ser encontrados em Bartholomew (1987). Como foi dito, o objeto da análise fatorial é descrever a variabilidade original do vetor aleatório y_t em um número menor de variáveis aleatórias (fatores) não correlacionadas. Assim um modelo fatorial y_t pode ser representado da seguinte forma:

$$y_t = \beta f_t + e_t, \;\; e_t \sim N(0, \Sigma),$$

onde y_t, f_t e e_t são vetores de dimensão m, k, m. Os fatores f_t são independentes de modo que o vetor de fatores f_t tem distribuição $f_t \sim N(0, I_k)$ onde $\Sigma = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_m^2)$. Os vetores e_t e f_s são independentes para todo t e s; e β é a matriz de carga $m \times k$ dos fatores. Assim, sendo a variância de y_t condicionada a β e Σ

$$V(y_t \mid \beta, \Sigma) = \beta\beta' + \Sigma.$$

3.1 Modelo fatorial dinâmico

Suponha *m* variáveis relacionadas que obedecem à distribuição normal multivariada denotada por $N(0, \Omega)$ onde Ω denota uma matriz $m \times m$ não singular de variância-covariância. Uma amostra aleatória de tamanho T é denotada por $\{\underset{\sim}{y}_t,\ t = 1,...,T\}$. O modelo utilizado neste estudo é dado por

$$\begin{aligned} \underset{\sim}{y}_t &= \beta \underset{\sim}{f}_t + \underset{\sim}{S}_t + \underset{\sim}{e}_t, \quad \underset{\sim}{e}_t \sim N(0, \Sigma) \\ \underset{\sim}{f}_t &= \Gamma_1 \underset{\sim}{f}_{t-1} + \Gamma_2 \underset{\sim}{f}_{t-2} + ... + \Gamma_p \underset{\sim}{f}_{t-p} + \underset{\sim}{w}_t, \quad \underset{\sim}{w}_t \sim N(0, \Lambda) \end{aligned} \tag{1)-(2)}$$

onde $\underset{\sim}{y}_t$, $\underset{\sim}{f}_t$, $\underset{\sim}{e}_t$ e $\underset{\sim}{w}_t$ são vetores de dimensão *m*, *k*, *m* e *k*, respectivamente, β é uma matriz $m \times k$, $\Gamma_i = diag(\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ik})$, $i = 1, ..., p$, $\Sigma = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_m^2)$ e $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_k)$.

Para algum inteiro positivo específico $k \leq m$, o modelo (1) relaciona cada $\underset{\sim}{y}_t$ com um vetor *k*-dimensional de variáveis aleatórias $\underset{\sim}{f}_t$, os fatores comuns, que obedecem a uma estrutura autorregressiva *AR*(*p*), e um componente sazonal, $\underset{\sim}{S}_t = (S_{1t} \quad S_{2t} \quad \cdots \quad S_{mt})^T$ (West e Harrison, 1997), definida por:

$$S_{it} = \sum_{j=1}^{6} \left[a_{ij} \text{sen}\left(\frac{2\pi jt}{12}\right) + b_{ij} \cos\left(\frac{2\pi jt}{12}\right) \right] = \underset{\sim}{\varphi}_i \underset{\sim}{X}_t,$$

onde:

$$\underset{\sim}{X}_t = \left(\text{sen}\left(\frac{2\pi t}{12}\right) \text{sen}\left(\frac{2\pi 2t}{12}\right) ... \text{sen}\left(\frac{2\pi 6t}{12}\right) \cos\left(\frac{2\pi t}{12}\right) \cos\left(\frac{2\pi 2t}{12}\right) ... \cos\left(\frac{2\pi 6t}{12}\right) \right)^T$$

e

$$\underset{\sim}{\varphi}_i = (a_{i1} \quad a_{i2} \quad \ldots \quad a_{i6} \quad b_{i1} \quad b_{i2} \quad \cdots \quad b_{i6}), \text{ para } i = 1, ..., m.$$

A ideia que motiva a aplicação da forma de Fourier para tratar sazonalidade está ligada ao fato de que qualquer padrão cíclico pode ser representado em termos da combinação linear de funções periódicas.[11] A motivação para o uso desta abordagem, além da flexibilidade, é a facilidade de se obter interpretação, já que as trajetórias em forma

11. Uma função $g(t)$ é dita periódica se para algum inteiro $p > 1$, e todo n, $t > 0$ $g(t + np) = g(t)$.

de ondas podem ser associadas aos variados padrões observados na prática. Portanto, distintamente do que se verifica em grande parte dos trabalhos empíricos, a sazonalidade é modelada endogenamente diminuindo assim mais uma fonte de erro no modelo.

Em problemas práticos, especialmente para valores grandes de *m*, o número de fatores *k* será pequeno quando comparado com *m*, de tal modo que muito da estrutura de variância-covariância é explicada pelos fatores comuns. As variâncias idiossincráticas, σ_i^2 medem a variabilidade residual em cada uma das variáveis dos dados uma vez que a contribuição dos fatores já foi considerada.

3.1.1 Restrições no modelo fatorial

Um modelo com *k* fatores deve apresentar alguma restrição para ser definido como um modelo único, não apresentando problemas de identificação. Há várias maneiras de identificar o modelo adicionando restrições no $\boldsymbol{\beta}$. Optou-se por seguir, aqui, a alternativa usada em Geweke e Zhou (1996) e Aguilar e West (2000), entre outros, ao restringir $\boldsymbol{\beta}$ de modo que seja uma matriz de bloco triangular inferior, assumindo ser de posto completo, com elementos da diagonal principal iguais a um, como apresentado adiante:

$$\boldsymbol{\beta} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \beta_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ \beta_{k,1} & \beta_{k,2} & \beta_{k,3} & \cdots & 1 \\ \beta_{k+1,1} & \beta_{k+1,2} & \beta_{k+1,3} & \cdots & \beta_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \beta_{m,3} & \cdots & \beta_{m,k} \end{pmatrix}.$$

A restrição exposta anteriormente garante que $\boldsymbol{\beta}$ tenha posto completo *k*, resolvendo os problemas de identificação que surgem pela invariância do modelo sob mudanças de locação da matriz $\boldsymbol{\beta}$. Ao mesmo tempo, garante que o número de parâmetros livres no tempo *t* na representação fatorial não exceda os $m(m + 1)/2$ parâmetros de $\boldsymbol{\Omega}$ não restrito. Além de resolver os problemas de identificação, essa estrutura de $\boldsymbol{\beta}$ influencia a interpretação dos fatores e pode impactar no ajuste do modelo e na escolha do *k* em particular, mas não tem impacto nas previsões, pois as variâncias e covariâncias entre as séries são independentes da decisão de modelagem.

3.1.2 Especificação das prioris

Para completar a especificação do nosso modelo, devemos atribuir classes de prioris para os parâmetros $\underset{\sim}{f}_0$, $\boldsymbol{\beta}$, Γ_i, $i = 1, ..., p$, Σ, Λ e φ_i, $i = 1, ..., m$. As prioris utilizadas são listadas a seguir.

1) Vetor de fatores comuns inicial

$$\underset{\sim}{f}_0 \sim N\left(\underset{\sim}{m}_0, C_0\right).$$

2) Elementos não nulos da matriz de cargas $\boldsymbol{\beta}$

$$\beta_{ij} \sim N\left(\mu_0, V_0\right) \text{ para } i > j.$$

3) Elementos da matriz diagonal de variância-covariância Σ

$$\sigma_i^2 \sim IG\left(\frac{n_\sigma}{2}, \frac{n_\sigma S_\sigma}{2}\right), \text{ para } i = 1, ..., m.$$

4) Elementos das matrizes diagonais dos coeficientes autorregressivos Γ_1, Γ_2, ..., Γ_p

$$\gamma_{ij} \sim N(m_r, S_r), \text{ para } i = 1, ..., p, j = 1, ..., k.$$

5) Elementos da matriz diagonal de variância-covariância Λ

$$\lambda_j \sim IG\left(\frac{n_\lambda}{2}, \frac{n_\lambda S_\lambda}{2}\right), j = 1, ..., k.$$

6) Elementos dos vetores de coeficientes sazonais φ_1, φ_2, ..., φ_m

$$a_{ij} \sim N(m_s, C_s) \text{ e } b_{ij} \sim N(m_s, C_s), \text{ para } i = 1, ..., m \text{ e } j = 1, ..., 6.$$

3.1.3 Método MCMC para o modelo fatorial com *k* fatores

Nosso objetivo é obter as posterioris dos parâmetros do modelo, assim como fazer inferência sobre o processo fatorial $\{\underset{\sim}{f}_t : t = 1,2,...,T\}$. Como uma solução exata requer o cálculo de integrais intratáveis do ponto de vista analítico, vamos usar o método do amostrador de Gibbs para obter uma aproximação das posterioris de interesse. As posterioris completas necessárias para implementar o algoritmo MCMC se encontram no anexo A. Para mais informações sobre como calcular essas condicionais completas no contexto de modelos fatoriais indicamos Kim e Nelson (1999).

3.1.4 Previsão do MFD com componente sazonal

Nesta subseção, será descrita a metodologia para previsão *h* passos à frente, para um modelo de ordem de defasagem *p* = 1 (que será o valor adotado nas análises da seção 4), a extensão para *p* > 1 ocorre de forma simples e se encontra descrita no anexo A.

O interesse aqui é na densidade preditiva *h* passos à frente:

$$p(\underset{\sim}{y}_{t+h} \mid \boldsymbol{y}) = \int p(\underset{\sim}{y}_{t+h} \mid \underset{\sim}{f}_{t+h}, \boldsymbol{\beta}, \varphi, \underset{\sim}{\Theta}) p(\underset{\sim}{f}_{t+h} \mid \underset{\sim}{f}_t, \boldsymbol{\beta}, \varphi, \underset{\sim}{\Theta}) p(\underset{\sim}{f}_t, \boldsymbol{\beta}, \varphi, \underset{\sim}{\Theta} \mid \boldsymbol{y}) d\underset{\sim}{f}_{t+h} d\underset{\sim}{f}_t d\boldsymbol{\beta} d\varphi d\underset{\sim}{\Theta},$$

onde:

$$\underset{\sim}{\Theta} = (\underset{\sim}{\sigma}, \underset{\sim}{\lambda}, \underset{\sim}{\gamma}), \underset{\sim}{\sigma} = (\sigma_1, ..., \sigma_m), \underset{\sim}{\gamma} = (\gamma_{11}, ..., \gamma_{1k}), \underset{\sim}{\lambda} = (\lambda_1, ..., \lambda_k),$$

$$(\underset{\sim}{y}_{t+h} \mid \underset{\sim}{f}_{t+h}, \boldsymbol{\beta}, \varphi, \underset{\sim}{\Theta}) \sim N(\boldsymbol{\beta} \underset{\sim}{f}_{t+h} + \varphi \underset{\sim}{X}_{t+h}, \Sigma) \quad \text{e} \quad (\underset{\sim}{f}_{t+h} \mid \underset{\sim}{f}_t, \boldsymbol{\beta}, \underset{\sim}{\sigma}) \sim N(\underset{\sim}{\mu}_h, V_h),$$

com $\underset{\sim}{\mu}_h = \Gamma_1^h \underset{\sim}{f}_t$ e $V_h = \sum_{k=1}^{h} \Gamma_1^{k-1} \Lambda (\Gamma_1^{k-1})^T$, para *h* > 0.

Portanto, se $\{(\boldsymbol{\beta}^{(1)}, \underset{\sim}{f}_t^{(1)}, \varphi^{(1)}, \Theta^{(1)}), ..., (\boldsymbol{\beta}^{(M)}, \underset{\sim}{f}_t^{(M)}, \varphi^{(M)}, \Theta^{(M)})\}$ é uma amostra de $p(\underset{\sim}{f}_t, \boldsymbol{\beta}, \varphi, \Theta \mid \boldsymbol{y})$, basta gerar uma amostra de tamanho *M* de $(\underset{\sim}{f}_{t+h} \mid \underset{\sim}{f}_t, \boldsymbol{\beta}, \varphi, \Theta)$, isto é, gerar $\{(\underset{\sim}{f}_{t+h}^{(1)}), ..., (\underset{\sim}{f}_{t+h}^{(M)})\}$, de modo que:

$$\hat{p}(\underset{\sim}{y}_{t+h} \mid \boldsymbol{y}) = M^{-1} \sum_{J=1}^{M} p(\underset{\sim}{y}_{t+h} \mid \underset{\sim}{f}_{t+h}^{(j)}, \boldsymbol{\beta}^{(j)}, \varphi^{(j)}, \Theta^{(j)})$$

seja uma aproximação de Monte Carlo para $p(\underset{\sim}{y}_{t+h} \mid \boldsymbol{y})$. Analogamente, uma amostra $\{\underset{\sim}{y}_{t+h}^{(1)},...,\underset{\sim}{y}_{t+h}^{(M)}\}$ de $p(\underset{\sim}{y}_{t+h} \mid \boldsymbol{y})$ é obtida apresentando $\underset{\sim}{y}_{t+h}^{(j)}$ de $p(\underset{\sim}{y}_{t+h} \mid \underset{\sim}{f}_{t+h}^{(j)}, \boldsymbol{\beta}^{(j)}, \varphi^{(j)}, \Theta^{(j)})$, para $j = 1, ..., M$.

3.2 MFD com fatores modelados

Neste tipo de MFD, os fatores são modelados pelas variáveis exógenas, sendo que o modelo é definido da seguinte forma:

$$\begin{aligned} y_t &= \beta f_t + S_t + e_t', \\ f_t &= X_{1t}\theta_{1t} + X_{2t}\theta_{2t} \cdots + X_{Kt}\theta_{Kt} + v_t', \\ \theta_{it} &= \Gamma_{i1t}\theta_{it-1} + \Gamma_{i2t}\theta_{it-2} \cdots \Gamma_{iNt}\theta_{it-p} + w_{it}, \end{aligned} \tag{3)-(5}$$

onde $e_t \sim N(0, \Sigma)$, $v_t \sim N(0, \Delta)$ e $w_{it} \sim N(0, W_i)$ são definidos para cada imposto *i*. O restante das variáveis é definido do mesmo modo como aparece na subseção 3.1. Conforme pode ser visto pela equação dos fatores f_t (4), os fatores são explicados pelas variáveis exógenas.

4 MLD

De modo a testar a confiabilidade do modelo fatorial, confrontamos os resultados deste com os obtidos pelo MLD (Mendonça, Medrano e Sachsida, 2013) descrito pelas equações (6) e (7), o qual associa cada tributo específico com seu possível fato gerador:

$$y_t = X_t\theta_t + S_t + \nu_t', \tag{6}$$

$$\theta_t = \Gamma_1\theta_{t-1} + \cdots + \Gamma_p\theta_{t-p} + \omega_t', \tag{7}$$

onde y_t é a variável endógena, ou seja, um tributo específico, e X_t é um vetor $k \times 1$ de variáveis exógenas, em outras palavras, as *proxies* para os geradores de arrecadação. Temos $w_{it} \sim N(0, W_i)$, para $i = 0, ..., 3$, enquanto φ_i é o componente autorregressivo da equação de estado para cada b_t. Este modelo permite, ainda, a presença de um componente sazonal, S_t, que pode ser modelado segundo a representação do tipo de Fourier (West e Harrison, 1997), tal como definido na subseção 3.1. Neste modelo, o termo constante aparece

também em X_t. A racionalidade por trás da introdução da constante não é ingênua. Pode-se supor que uma parcela da arrecadação do tributo é ser autônoma. A descrição acerca do método de inferência deste modelo é descrito no anexo B.

5 BASE DE DADOS

A amostra apresentada é composta de tributos federais cuja base de dados é gerenciada pela Secretaria de Política Econômica (SPE). Os dados de impostos federais utilizados neste estudo estão relacionados aos dados brutos ajustados. Estes dez tributos representam aproximadamente 50% da receita bruta do total de tributos arrecadados pelas administrações públicas brasileiras. Os dados são mensais e se referem ao período de janeiro de 2001 a dezembro de 2013.

Os tributos considerados estão discriminados da seguinte maneira: três são tributos incidentes sobre a renda, o patrimônio e o capital, a saber, o IRPF, o IRPJ e o IRRF. Quatro tributos são incidentes sobre produtos: o II, a Cofins, o IOF e o IPI. Estes últimos se compõem da soma de impostos incidentes sobre três produtos (bebidas, veículos, fumo). Integram ainda a amostra as seguintes contribuições previdenciárias: CSLL e a contribuição para o PIS/Pasep (CPIS). Há ainda o item outras receitas (OR) que se compõe de outras receitas tributárias de menor expressão.[12]

6 RESULTADOS

Antes de apresentar os resultados obtidos, é importante tecer algumas considerações. Além das receitas tributárias já listadas, foram incluídas no conjunto de variáveis dos modelos fatoriais algumas variáveis macroeconômicas que julgamos serem correlacionadas com os impostos e, portanto, capazes de adicionar capacidade preditiva ao modelo. Consideramos que a inclusão de certas variáveis macroeconômicas seria importante na medida em que elas estão necessariamente associadas na geração desses impostos.

12. Aqui integram a Contribuição para a Fundação do Desenvolvimento Administrativo (Fundap), Cide-Combustíveis, a Contribuição sobre Movimentação Financeira (CPMF), o Imposto Territorial Rural (ITR) além de outras receitas administradas.

Desse modo, após uma seleção parcimoniosa, consideramos incluir o conjunto de dez variáveis macroeconômicas apresentadas adiante.

1) Produto interno bruto (PIB) a preços correntes. Fonte: Banco Central do Brasil (BCB).

2) Taxa de câmbio R$/US$. Fonte: BCB.

3) Índice geral de preços do mercado (IGPM). Fonte: Fundação Getulio Vargas (FGV).

4) Índice do produto industrial (PIND). Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE).

5) Índice do produto industrial-Bebidas (PINDB). Fonte: IBGE.

6) Índice do produto industrial-Automóveis (PINDA). Fonte: IBGE.

7) Índice do produto industrial-Fumo (PINDF). Fonte: IBGE.

8) Concessões de crédito com recursos livres (CRED). Fonte: BCB.

9) Índice de faturamento nominal-indústria (FATN). Fonte: IBGE.

10) Rendimento médio do trabalho (RMT). Fonte: IBGE.

No modelo fatorial no qual os fatores são modelados (M1), as variáveis listadas entram como variáveis explicativas dos fatores. No modelo fatorial básico (M2), as variáveis macroeconômicas entram com os tributos no conjunto de variáveis endógenas. Conforme foi dito, este modelo tem a vantagem de não necessitar predefinir a trajetória de nenhuma variável para o horizonte temporal em que a previsão será feita no caso do modelo linear dinâmico (M3), usamos apenas o PIB como *proxy* para o fato gerador. Embora possa parecer simplório que apenas o PIB sirva de variável explicativa para modelar os diversos impostos, devemos ressaltar que Mendonça, Medrano e Sachsida (2013) testaram diversas variáveis explicativas para os impostos com base no MLD e

constatam que o PIB foi, na maior parte dos casos, a única variável que demonstrou capacidade significativa de previsão. Na parte autorregressiva de cada modelo, fizemos sempre uso de apenas uma defasagem.

Para avaliar a capacidade preditiva do modelo *p* passos à frente fora da amostra, usamos diversos tipos de critérios tais como o erro quadrado médio (*mean squared error* – MSE),[13] o desvio absoluto médio (*mean absolute deviation* – MAD),[14] o erro percentual absoluto médio (*mean absolute percentage error* – MAPE),[15] o erro percentual médio (*mean percentage error* – MPE) e o erro acumulado percentual (Erro %). Quando o objetivo específico se trata de checar a *performance* do modelo fora da amostra, ou seja, quando o objetivo é previsão, a literatura econométrica não dispensa muita discussão. Diferentemente, existem diversos critérios concorrentes quando o propósito é avaliar o desempenho do modelo dentro da amostra.[16] Em geral, o critério erro quadrático médio (EQM) é comumente mencionado na literatura econométrica teórica (Stock e Watson, 2010; Lutkepohl, 2007; Hamilton, 1993) quando o objetivo específico é checar a capacidade preditiva do modelo. A utilização do erro de previsão acumulado no período como critério de escolha entre modelos se deve ao fato de ter parecido a esses autores uma boa opção, pois este é um critério adotado pelos órgãos responsáveis em acompanhar a arrecadação tributária no Brasil, a exemplo da SPE. Isso se explica pelo fato de estarem interessados na previsão de arrecadação acumulada para o ano inteiro. Nesse caso, um erro de previsão positivo pode compensar outro negativo, de modo que eles se anulam ao final do período.

No modelo M1 o vetor y_t, conjunto de variáveis endógenas, é composto apenas pelos tributos. No modelo M2, o vetor y_t é composto de dezenove variáveis (nove tributos[17] e dez variáveis macroeconômicas). No modelo fatorial, as variáveis foram normalizadas[18]

13. $EQM = N^{-1}\sum_{i=1}^{N} e_i^2$.

14. $DAM = \frac{1}{n}\sum_{j=T+1}^{n} |e_j|$.

15. $MAPE = \left[\frac{1}{n}\sum_{j=1}^{n}\frac{|e_j|}{y_j}\right]\times 100$.

16. Os critérios de informação de Akaike, teste de razão de verossimilhança, Hannan-Quinn e Schwarz bayesiano são exemplos de critérios de escolha de modelos dentro da amostra (Lutkepohl, 2007).

17. Neste modelo, o IPI não entra desagregado.

18. $\hat{y}_t = \frac{y_t - \overline{y}}{\sigma_y}$ onde $\overline{y}$ é a média e σ_y o desvio-padrão amostral.

de modo a evitar problemas de escala entre elas, já que a matriz de variância-covariância se mantém inalterada. Isso se configura numa prática comum nesse tipo de modelo. O modelo fatorial M1 também foi estimado de modo a tratar o IPI desagregado (bebidas, automóveis e fumo), sendo que o resultado final de previsão, para fim de comparação com os demais modelos, foi tomado no agregado, ou seja, pela soma das receitas de IPI derivadas desses três produtos. Estimamos estes dois modelos com três (M4) e quatro fatores (M5). Com exceção do modelo M2, em que não se necessita saber de antemão o valor de nenhuma variável para se efetuar a previsão, nos demais modelos fez-se uso do valor observado das variáveis exógenas do ano de 2012. Conforme foi dito, a base de dados mensais é de janeiro de 2001 a dezembro de 2012. A previsão para o ano de 2013 é feita fora da amostra.

No anexo C são apresentados os resultados do modelo. O gráfico de tributo é apresentado com o valor obtido para cada critério usado de previsão. Observando os resultados do anexo C, percebe-se inicialmente que alguns tributos têm uma forte estrutura sazonal que não pode ser ignorada na análise. O que é o caso dos impostos incidentes sobre a renda (IRPF, IRPJ, IRRF), Cofins e a CSLL. Com efeito, o método usado para tratar a sazonalidade parece se adequar a estrutura sazonal desses tributos. Cumpre notar que não se verificou, de acordo com os critérios adotados, a preferência por um determinado modelo. No entanto, deve-se ter em mente que, com exceção do modelo M2, a arrecadação predita pelos demais modelos somente pode ser obtida mediante informação *a priori* das variáveis exógenas e que, num contexto prático, necessitam, portanto, ser estimadas. Assim, o modelo M2 traz ganhos substanciais de eficiência na medida em que, embora superado na maioria dos casos pelos demais modelos, seu erro de previsão é menor que 10%, de acordo com o critério de erro percentual (Erro %) critério de uso corrente da STN.

De modo a testar a robustez do modelo fatorial, estimamos o modelo com dados até 2011, fazendo previsão para o ano de 2012. Para todos os critérios de previsão adotados, o modelo fatorial obteve desempenho superior ao modelo linear univariado na qualidade da previsão para todos os tributos.

7 COMENTÁRIOS FINAIS

Este trabalho teve como objetivo a aplicação do MFD, estimado com base na proposta bayesiana para previsão de arrecadação de uma amostra substancial da CTBB, composta por onze importantes tributos, o que representa perto de 50% da arrecadação. A base de dados é mensal no período de janeiro de 2001 a dezembro de 2012, sendo que a previsão foi feita fora da amostra para o ano de 2013. A vantagem do modelo fatorial é que ele permite que a informação contida nas relações entre as variáveis seja levada em consideração na hora da modelagem, fazendo com que apenas informações genuínas[19] sejam levadas em conta.

A previsão da arrecadação tributária é de vital importância para a gestão da política econômica. Nesse sentido, o MFD proposto aqui pode se tornar uma ferramenta fundamental para o gestor de política por pelo menos três motivos fundamentais. Primeiro, diferentemente das ferramentas para previsão de arrecadação até então empregadas, o modelo fatorial trata um conjunto de tributos simultaneamente. Em segundo lugar, ele não necessita que se conheça antecipadamente o comportamento de alguma variável *proxy* para o fato gerador de determinado tributo. Por fim, ao reduzir a dimensionalidade do conjunto de tributos por meio da introdução dos fatores, o modelo fatorial permite, além da vantagem já mencionada, obter relevante ganho de eficiência computacional.

De modo a testar a qualidade da previsão, fez-se uso de uma gama de critérios estatísticos que evidenciaram que as previsões do modelo fatorial foram bastante razoáveis. Baseado em estudos preliminares, acreditamos que as previsões fornecidas por essa modelagem são promissoras, numa comparação formal com outros modelos geralmente usados para modelar a arrecadação tributária no Brasil.

19. Informação não correlacionada.

REFERÊNCIAS

AGUILAR, O.; WEST, M. Bayesian dynamic factor models and portfolio allocation. **Journal of Business and Economic Statistics**, v. 18, n. 3, p. 338-357, 2000.

BARTHOLOMEW, D. J. **Latent variable models and factor analysis**. London: Charles Griffin, 1987.

CARTER, C. K.; KOHN, R. On gibbs sampling for state space models. **Biometrika**, v. 81, n. 3, p. 541-553, 1994.

CARVALHO, J. C. **Tendência de longo prazo das finanças públicas no Brasil**. Brasília: Ipea, 2001. (Texto para Discussão, n. 795).

FRÜHWIRTH -SCHNATTER, S. Data augmentation and dynamic linear models. **Journal of Time Series Analysis**, n. 15, p. 183-202, 1994.

GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo**: stochastic simulation for Bayesian inference. Chapman & Hall, 2006.

GELMAN, A. *et al.* **Bayesian Data Analysis**, 2003.

GEWEKE, J.; ZHOU, G. Measuring the pricing error of the arbitrage pricing theory. **The Review of Financial Studies**, v. 9, n. 2, p. 557-587, 1996.

HAMILTON, J. **Time series analysis**. Princeton University Press, 1993.

HERNÁNDEZ, B. **Um modelo econométrico da conta corrente do governo no Brasil 1951/95**. Rio de Janeiro: Ipea, 1998. (Texto para Discussão, n. 543).

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistema de Contas Nacionais (SCN)**. 2004 (Série Relatórios Metodológicos, v. 24).

KIM, C. J.; NELSON, C. R. **State-space models with regime switching**. Massachusetts: The MIT Press, 1999.

LUTKEPOHL, H. **New introduction to multiple time series analysis**. Springer, 2007.

MELO, B. Modelo de previsão para arrecadação tributária. *In*: PRÊMIO TESOURO NACIONAL, 6., 2001, **Anais**... , ESAF-Escola de Administração Fazendária, 2001.

MENDONÇA, M. J.; MEDRANO, L. A. T; SACHSIDA, A. Um modelo econométrico com parâmetros variáveis para carga tributária bruta trimestral. **Pesquisa e Planejamento Econômico**, v. 41, n. 1, p. 133-162, 2011.

_____. Um modelo econométrico para previsão de receita tributária no Brasil. **Economia Aplicada**, v. 17, n. 2, p. 295-329, 2013.

MENDONÇA, M. J., SANTOS, C. H.; MARTINS, T. G. **Aplicação de um modelo fatorial dinâmico para previsão da arrecadação tributária no Brasil**. Ipea, 2009. (Texto para Discussão, n. 1453).

PORTUGAL, C.; PORTUGAL, M. Os efeitos da inflação sobre o orçamento do governo: uma análise empírica. **Estudos Econômicos**, v. 31, n. 2, p. 239-283, 2001.

REZENDE, F.; OLIVEIRA, F. A.; ARAÚJO, E. A. **O dilema fiscal**: remendar ou reformar? Editora FGV, 2008.

SÁFADI, T.; PENA, D. Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil. **Environmetrics**, v. 19, n. 6, p. 582-601, 2007.

SANTOS, C. H. M. dos; RIBEIRO, M. B.; GOBETTI, S. W. A evolução da carga tributária bruta brasileira no período 1995-2007: **tamanho, composição e especificações econométricas agregadas**. Ipea, 2008. (Texto para Discussão, n. 1350).

SIQUEIRA, M. **Melhorando a previsão de arrecadação tributária federal através da utilização de modelos de séries temporais**. *In*: PRÊMIO TESOURO NACIONAL, 6., 2002, **Anais**... ESAF-Escola de Administração Fazendária, 2002.

STOCK, J. H.; WATSON, M. W. **Introduction to econometrics**. Pearson, 2010.

WEST, M.; HARRISON, J. **Bayesian forecasting and dynamic models**. 2nd ed. Springer, 1997.

APÊNDICE A

DESCRIÇÃO DO ALGORITMO MCMC PARA O MFD

Descreveremos a seguir o procedimento bayesiano para estimação e previsão de um modelo fatorial dinâmico do tipo que foi introduzido na subseção 3.1. Tal modelo pode ser representado em uma estrutura de espaço de estados do seguinte modo:

$$y_t = F^E f_t + e_t^E,$$
$$f_t^E = G^E f_{t-1}^E + w_t^E,$$

onde $e_t^E \sim N(0,\Sigma)$ e $e_t^E \sim N(0,Q^E)$

$$G_{pkxpk}^E = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \ldots & \Gamma_{p-1} & \Gamma_p \\ I_{kxk} & \underset{\sim}{0}_{kxk} & \ldots & \underset{\sim}{0}_{kxk} & \underset{\sim}{0}_{kxk} \\ \underset{\sim}{0}_{kxk} & I_{kxk} & \ldots & \underset{\sim}{0}_{kxk} & \underset{\sim}{0}_{kxk} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \underset{\sim}{0}_{kxk} & \underset{\sim}{0}_{kxk} & \ldots & I_{kxk} & \underset{\sim}{0}_{kxk} \end{bmatrix}, \quad \underset{\sim}{f}_t^E = \begin{bmatrix} \underset{\sim}{f}_t \\ \underset{\sim}{f}_{t-1} \\ \vdots \\ \underset{\sim}{f}_{t-p-1} \end{bmatrix}, \quad \underset{\sim}{w}_t^E = \begin{bmatrix} \underset{\sim}{w}_t \\ \underset{\sim}{0}_{kx1} \\ \vdots \\ \underset{\sim}{0}_{kx1} \end{bmatrix},$$

$$Q_{pkxpk}^E = \begin{bmatrix} \Lambda_{kxk} & \underset{\sim}{0}_{kxk(p-1)} \\ \underset{\sim}{0}_{k(p-1)xk} & \underset{\sim}{0}_{k(p-1)xk(p-1)} \end{bmatrix}, \quad F_{mxpk}^E = \begin{bmatrix} \boldsymbol{\beta}_{mxk} & \underset{\sim}{0}_{mxk} & \underset{\sim}{0}_{mxk} & \ldots & \underset{\sim}{0}_{mxk} \end{bmatrix}.$$

POSTERIORIS COMPLETAS NECESSÁRIAS PARA A IMPLEMENTAÇÃO DO AMOSTRADOR DE GIBBS

Para $i = 1, \ldots, k$, escreva $\underset{\sim}{\beta}_i = (\beta_{i1},\ldots,\beta_{ii})^T$ que são os elementos diferentes de zero. Para os restantes das linhas $i = k + 1, \ldots, m$, escreva $\underset{\sim}{\beta}_i = (\beta_{i1},\ldots,\beta_{ik})$. Similarmente, para $i = 1, \ldots, k$ denote $\boldsymbol{F}_i$ a matriz $T \times i$ contendo as primeiras i colunas de $\boldsymbol{F}$ e para todo i seja $\boldsymbol{y}_i$ a i-ésima coluna de $\boldsymbol{y}$, $\boldsymbol{S}_i$ a i-ésima coluna de $(\boldsymbol{X}\varphi^T)$ com $\boldsymbol{X} = (\underset{\sim}{X}_1^T,\ldots,\underset{\sim}{X}_T^T)^T$ e $\boldsymbol{y}_i^* = \boldsymbol{y}_i - \boldsymbol{S}_i$. Seja $\underset{\sim}{\gamma}_j = (\gamma_{1j},\gamma_{2j},\ldots,\gamma_{pj})^T$ o vetor contendo o j-ésimo elemento da diagonal para cada uma das p matrizes $\Gamma_1,\Gamma_2,\ldots,\Gamma_p$. Defina também para $j = 1, \ldots, k$

$$\mathfrak{I}_j = \begin{pmatrix} f_{p,j} & f_{p-1,j} & \cdots & f_{1,j} \\ f_{p+1,j} & f_{p,j} & \cdots & f_{2,j} \\ \vdots & \vdots & \vdots & \vdots \\ f_{T-1,j} & f_{T-2,j} & \cdots & f_{T-p,j} \end{pmatrix} \quad \text{e} \quad \mathrm{H}_j = \begin{pmatrix} f_{p+1,j} \\ f_{p+2,j} \\ \vdots \\ f_{T,j} \end{pmatrix}.$$

- Para amostrar da condicional completa de $\underset{\sim}{f}_t$:

1) Aplicar filtro de Kalman para calcular $m_t = E[\underset{\sim}{f}_t^E \mid D_t]$ e $C_t = V[\underset{\sim}{f}_t^E \mid D_t]$ para $t = 1, ..., T$, do seguinte modo:

$$a_t = G^E m_{t-1}, \ldots R_t = G^E C_{t-1} G^{E^T} + Q^E,$$

$$f_t = F^E a_t, \ldots Q_t = F^E R_t F^{E^T} + \Sigma,$$

$$A_t = R_t F^{E^T} Q_t^{-1}, \ldots e_t = (y_t - S_t) - f_t$$

e armazenar m_t e C_t, $t = 1, ..., T$. No final use m_t e C_t para gerar $\underset{\sim}{f}_T^E \mid \cdot \sim N(m_T, C_T)$.

2) Definir $\underset{\sim}{f}_t^{E^*}$ como sendo os primeiros k elementos de $\underset{\sim}{f}_t^E$; G^{E^*} como sendo as primeiras k linhas de G^E; e Q^{E^*} como sendo o primeiro bloco $k \times k$ da matriz Q^E. Para $t = T-1, T-2, ..., 1$ faça:

$$m_t^* = m_t + C_t G^{E^{*T}} (G^{E^*} C_t G^{E^{*T}} + Q^{E^*})^{-1} (f_{t+1}^{E^*} - G^{E^*} m_t),$$

$$C_t^* = C_t - C_t G^{E^{*T}} (G^{E^*} C_t G^{E^{*T}} + Q^{E^*})^{-1} G^{E^*} C_t,$$

gere $\underset{\sim}{f}_t^E \mid ., f_{t+1}^{E^*} \sim N(m_t^*, C_t^*)$ e mantenha somente os k primeiros elementos gerados dessa forma. Esses primeiros k elementos formam uma amostra de f_t da condicional completa.

- *A posteriori* condicional completa de Σ se reduz a um conjunto de m gammas inversas independentes, com $\sigma_i^2 \sim IG\left(\frac{n_\sigma^*}{2}, \frac{n_\sigma^* S_\sigma^*}{2}\right)$, onde

$$n_\sigma^* = n_\sigma + T,$$

$$n_\sigma^* S_\sigma^* = n_\sigma S_\sigma + (\boldsymbol{y}_i^* - \boldsymbol{F}\underset{\sim}{\beta}_i^T)^T (\boldsymbol{y}_i^* - \boldsymbol{F}\underset{\sim}{\beta}_i^T).$$

- *A posteriori* condicional completa de β divid2-se em:

  para $i = 1, ..., k$, $\underset{\sim}{\beta}_i \sim N(\underset{\sim}{\mu}_i, \underset{\sim}{V}_i)1(\beta_{ii} = 1)$, onde $\underset{\sim}{\mu}_i = \underset{\sim}{V}_i(V_0^{-1}\mu_0 1_i + \sigma_i^{-2}\boldsymbol{F}_i^T \boldsymbol{y}_i^*)$ e $\underset{\sim}{V}_i^{-1} = V_0^{-1} I_i + \sigma_i^{-2}\boldsymbol{F}_i^T \boldsymbol{F}_i$;

  para $i = k + 1, ..., m$, $\underset{\sim}{\beta}_i \sim N(\underset{\sim}{\mu}_i, \underset{\sim}{V}_i)$, onde $\underset{\sim}{\mu}_i = \underset{\sim}{V}_i(V_0^{-1}\mu_0 1_k + \sigma_i^{-2}\boldsymbol{F}^T \boldsymbol{y}_i^*)$ e $\underset{\sim}{V}_i^{-1} = V_0^{-1} I_k + \sigma_i^{-2}\boldsymbol{F}^T \boldsymbol{F}$.

- *A posteriori* condicional completa de $\underset{\sim}{\gamma}_j$ é dada por:

  para $j = 1, ..., k$, $\underset{\sim}{\gamma}_j \sim N(A_j, B_j)$, onde $A_j = B_j(S_r^{-1} m_r 1_p + \lambda_j^{-1}\Im_j^T \mathrm{H}_j)$ e $B_j = S_r^{-1} I_{pxp} + \lambda_j^{-1}\Im_j^T \Im_j)$.

- *A posteriori* condicional completa de Λ se reduz a um conjunto de *k* gammas inversas independentes, com $\lambda_j \sim IG\left(\frac{n_\lambda^*}{2}, \frac{n_\lambda^* S_\lambda^*}{2}\right)$, onde

  $n_\lambda^* = n_\lambda + T$

  $n_\lambda^* S_\lambda^* = n_\lambda S_\lambda + (\mathrm{H}_j - \Im_j \underset{\sim}{\gamma}_j)^T (\mathrm{H}_j - \Im_j \underset{\sim}{\gamma}_j)$.

- *A posteriori* condicional completa de φ_i é dada por:

  para $i = 1, ..., m$, $\underset{\sim}{\varphi}_i \sim N(C_i, D_i)$, onde $C_i = D_i(C_s^{-1} m_s 1_{12} + \sigma_i^{-1}\boldsymbol{X}(\boldsymbol{y}_i - \boldsymbol{F}\beta_i))$ e $D_i = C_s^{-1} I_{12x12} + \sigma_i^{-2}\boldsymbol{X}^T \boldsymbol{X}$.

PREVISÃO MFD COM COMPONENTE SAZONAL COM *p* > 1

Passe o modelo (1) para a estrutura de espaço de estados (2) e utilize a mesma metodologia apresentada na subseção 2.6, substituindo $\underset{\sim}{f}_t$, Λ, β e Γ_1 por $\underset{\sim}{f}_t^E$, Q^E, F^E e G^E, respectivamente.

APÊNDICE B

PROCEDIMENTO DE ESTIMAÇÃO DO MLD

Neste anexo mostramos o método de estimação e previsão do MLD com base na inferência bayesina. A forma geral do modelo MLD é mostrada pelas equações (6) e (7) da seção 4 replicadas a seguir:

$$y_t = X_t\theta_t + S_t + \nu'_t, \tag{B.6}$$

$$\theta_t = \Gamma_1\theta_{t-1} + \cdots + \Gamma_p\theta_{t-p} + \omega'_t, \tag{B.7}$$

onde y_t é a variável endógena e X_t é um vetor $k\times 1$ de variáveis exógenas. Este modelo permite ainda a presença de um componente sazonal, S_t, que pode ser modelado segundo a representação do tipo de Fourier (West e Harrison, 1997), tal como definido na subseção 3.1.

$$S_t = \sum_{j=1}^{2}\left[a_j \text{sen}\left(\frac{2\pi jt}{4}\right) + b_j \cos\left(\frac{2\pi jt}{4}\right)\right] = Z_t\varphi \text{, onde}$$

$$Z_t = \left(\text{sen}\left(\frac{2\pi t}{4}\right)\ \text{sen}\left(\frac{2\pi 2t}{4}\right)\ \cos\left(\frac{2\pi t}{4}\right)\ \cos\left(\frac{2\pi 2t}{4}\right)\right)^T \quad \text{e} \quad \varphi = (a_1\ a_2\ b_1\ b_2),$$

sendo θ_t o vetor $k\times 1$ de parâmetros que segue um processo autorregressivo de ordem p, em que os coeficientes estão contidos na matriz diagonal Γ_i, $k\times k$, para i = 1, ..., p. Supõe-se que $\omega'_t \sim N(0_{k\times 1}, W_{k\times k})$ e $v'_t \sim N(0,V)$, admitindo-se ainda que ω'_t e v'_t não são correlacionados. Por simplicidade, admite-se também que $W = diag(w_1,\ldots,w_k)$.

A ideia da aplicação da forma de Fourier para tratar sazonalidade está ligada ao fato de que qualquer padrão cíclico pode ser representado em termos da combinação linear de funções periódicas.[1] Outra motivação para o uso desta abordagem, além da flexibilidade, é a facilidade de se obter interpretação, já que as trajetórias em forma de

1. Uma função $g(t)$ é dita periódica se, para algum inteiro $p > 1$, e todo n, $t > 0$, tem-se que $g(t + np) = g(t)$.

ondas podem ser associadas aos variados padrões observados na prática. Mais uma forma de tratar a sazonalidade seria introduzi-la diretamente na equação de estado. Neste caso, a equação (7) assume a seguinte forma:

$$\theta_t = \Gamma_1\theta_{t-1} + \cdots + \Gamma_p\theta_{t-p} + S_t + \omega'_t. \qquad \text{(B.7')}$$

Por fim, a sazonalidade pode tomar uma forma dinâmica. Neste caso, seria possível ter-se um modelo no qual a equação da medida é dada da seguinte forma: $y_t = F_t\theta^*_t + \nu'_t$, onde $F = [X_t, Z_t]$ e $\theta^*_t = [\theta_t, \varphi_t]$.

Será feito agora uso da seguinte notação para encaixar as equações numa forma mais compacta, que reduz o sistema expresso pelas equações (B.6 e B.7) numa forma de espaço-estado de ordem um:

$$\underline{\theta}_t = \begin{bmatrix} \theta_t \\ \theta_{t-1} \\ \vdots \\ \theta_{t-p+1} \end{bmatrix}_{kp\times 1}, G = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_{p-1} & \Gamma_p \\ I_k & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_k & 0 \end{bmatrix}_{kp\times kp}, \underline{\omega}_t = \begin{bmatrix} \omega'_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{kp\times 1} \text{ e } F'_t = \begin{bmatrix} X'_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{kp\times 1}.$$

Visto isso, temos que:

$$y_t = F_t\underline{\theta}_t + z_t\varphi + v'_t, \qquad \text{(B. 8)}$$

$$\underline{\theta}_t = G\underline{\theta}_{t-1} + \underline{\omega}_t, \qquad \text{(B.9)}$$

onde (B.8) representa a equação da medida ou do movimento, enquanto (B.9) define a equação de estado.

A função de verossimilhança de (θ, G, φ, V) é dada por $p(y \mid \theta, G, \varphi, V) = (2\pi)^{-T/2}V^{T/2}\exp\left(-\frac{1}{2V}\sum_{t=1}^{T}(y_t - F_t\underline{\theta}_t - z_t\varphi)^2\right)$, onde y = (y_1 ... y_T), θ =(θ_1 ... θ_T) e Γ = (Γ_1 ... Γ_p).

PROCEDIMENTO DE INFERÊNCIA

Nesta seção, é realizado um procedimento de inferência baseado no paradigma bayesiano. Inicialmente, apresentam-se distribuições *a priori* para todos os parâmetros. Em seguida, o algoritmo MCMC é utilizado para obter amostras *a posteriori* de todos os parâmetros do modelo.

Distribuição a *priori*

Por simplicidade, distribuições *a priori* conjugadas são utilizadas para todos os parâmetros. A distribuição *a priori* para θ_t é especificada na equação (3.2) e completada pela informação inicial θ_0 tal que $\theta_0 \sim N(m_0, C_0)$, onde m_0 e C_0 são hiperparâmetros conhecidos. As distribuições *a priori* para os parâmetros V, $W\Gamma$ e φ são as listadas a seguir:

1) $V \sim GI(n_V / 2, n_V S_V / 2)$.

2) $w_j \sim GI(n_W / 2, n_W S_W / 2)$, $j = 1, ..., k$.

3) $\Gamma_j \sim N(m_\Gamma, S_\Gamma)$, $j = 1, ..., p$.

4) $\varphi \sim N(m_\varphi, S_\varphi)$.

Onde n_V, S_V, n_W, S_W, m_Γ, S_Γ, m_φ e S_φ são hiperparâmetros conhecidos. Os valores dos hiperparâmetros utilizados foram os seguintes:

$$m_0 = 0,\ C_0 = 100,\ n_V = 2, n_V S_V = 0.1,\ n_W = 2, n_W S_W = 0.1,$$
$$m_\Gamma = 0,\ S_\Gamma = 100,\ m_\varphi = 0, S_\varphi = 100.$$

Inferência a *posteriori*

A distribuição *a posteriori* conjunta de (θ, Γ, φ, V, W) é dada por:

$$p(\theta, G, \varphi, V, W \mid y) \propto \prod_{t=1}^{T} p(y_t \mid \underline{\theta}_t, \varphi, V) \prod_{t=1}^{T} p(\underline{\theta}_t \mid \underline{\theta}_{t-1}, W, G) p(\underline{\theta}_0 \mid m_0, C_o) P(\varphi) p(V) p(W) p(\Gamma),$$

a qual é analiticamente intratável. Por conseguinte, a inferência *a posteriori* exata é feita utilizando-se o esquema MCMC. Neste caso, o vetor da variável de estado é amostrado

conjuntamente, utilizando-se o algoritmo FFBS (*forward filtering backward sampling*) de Carter e Kohn (1994) e Frühwirth-Schnatter (1994). Para o resto de parâmetros Γ, φ, V e W, todas as distribuições condicionais completas são distribuições normais ou gamas inversas. Todas as distribuições condicionais completas estão listadas a seguir.

1) A distribuição condicional de V é $(V \mid \theta, W, \varphi, \Gamma) \sim GI(n_V^* / 2, n_V^* S_V^* / 2)$, onde $n_V^* = n_V + T$ e $n_V^* S_V^* = n_V S_V + \sum_{t=1}^{T} (y_t - F_t \underline{\theta}_t - z_t \varphi)^2$.

2) A distribuição condicional de $W = diag(w_1, \ldots, w_k)$ é $(w_j \mid \theta, V, \varphi, \Gamma) \sim GI(n_W^* / 2, n_W^* S_W^* / 2)$ para j = 1, ..., k, onde $n_W^* = n_W + T - 1$ e $n_W^* S_W^* = n_W S_W + \sum_{t=2}^{T} (\underline{\theta}_t - G\underline{\theta}_{t-1})^T (\underline{\theta}_t - G\underline{\theta}_{t-1})$.

3) A distribuição condicional de $\Gamma = (\Gamma_1 \ \ldots \ \Gamma_p)$ é $(\Gamma_j \mid \theta, V, W\varphi) \sim N(m_\Gamma^*, S_\Gamma^*)$ para $j = 1, \ldots, k$, onde $S_\Gamma^* = (S_\Gamma^{-1} I_{pxp} + w_j^{-1} H_j^T H_j)^{-1}$ $m_\Gamma^* = S_T^{-1}(S_\Gamma^{-1} I_{pxp} 1_p + w_j^{-1} B_j^T H_j)$ para $H_j = \begin{pmatrix} \theta_{p,j} & \theta_{p-1,j} & \cdots & \theta_{1,j} \\ \theta_{p+1,j} & \theta_{p,j} & \cdots & \theta_{2,j} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{T-1,j} & \theta_{T-2,j} & \cdots & \theta_{T-p,j} \end{pmatrix}$ e $B_j = \begin{pmatrix} \theta_{p+1,j} \\ \theta_{p+2,j} \\ \vdots \\ \theta_{T,j} \end{pmatrix}$.

4) A distribuição condicional de φ é $(\varphi \mid \theta, V, W, \Gamma) \sim N(m_\varphi^*, S_\varphi^*)$, onde $S_\varphi^* = (S_\varphi^{-1} I_{4x4} + V^{-1} Z^T Z)^{-1}$ e $m_\varphi^* = S_\varphi^{-1}(S_\varphi^{-1} I_{4\times4} 1_4 + V^{-1} Z^T (y - F\theta))$.

Os estados θ_1, ..., θ_T são amostrados conjuntamente, utilizando-se o algoritmo FFBS, condicional sobre Γ, φ, V e W. A distribuição condicional completa conjunta segue a decomposição retrospectiva $p(\theta \mid y) = p(\theta_T \mid D_T) \prod_{t=1}^{T-1} p(\theta_t \mid \theta_{t+1}, D_t)$, onde $D_T = \{y_1, y_2, \ldots, y_t\}$, t = 1, ..., T e D_0 representam a informação inicial. Iniciando com $\theta_0 \sim N(m_0, C_0)$, por meio do filtro de Kalman pode ser mostrado que:

$\theta_t \mid D_t \sim N(m_t, C_t)$ onde $m_t = a_{t-1} + A(y_t - f_t - z_t \varphi)$, $C_t = R_t - A_t Q_t A_t'$, $a_t = Gm_{t-1}$, $R_t = GC_{t-1}G' + W$, $f_t = F_t' a_t + z_t \varphi$, $Q_t = F_t' R_t F_t + V$ e $A_t = R_t F_t Q_t^{-1}$, para t = 1, ..., T;

θ_T é amostrado de $p(\theta_T \mid D_T)$ (passo denominado *forward filtering*). Para $t = T-1$, $T-2, \ldots 2,1$, θ_t é amostrado de $p(\theta_t \mid \theta_{t+1}, D_t) = N(\tilde{a}_t, \tilde{C}_t)$, onde $\tilde{a}_t = m_t + B_t(f_{t+1} - a_{t+1})$, $\tilde{C}_t = C_t - B_t R_{t+1} B_t'$ e $B_t = C_t G' R_{t+1}^{-1}$ (passo denominado *backward sampling*).

Previsão modelo de regressão linear dinâmico

Fazer previsão nesta classe de modelos é teoricamente direto. Frequentemente, surge o interesse de aprender a densidade preditiva h passos à frente, $p(y_{T+h} \mid y)$, ou seja:

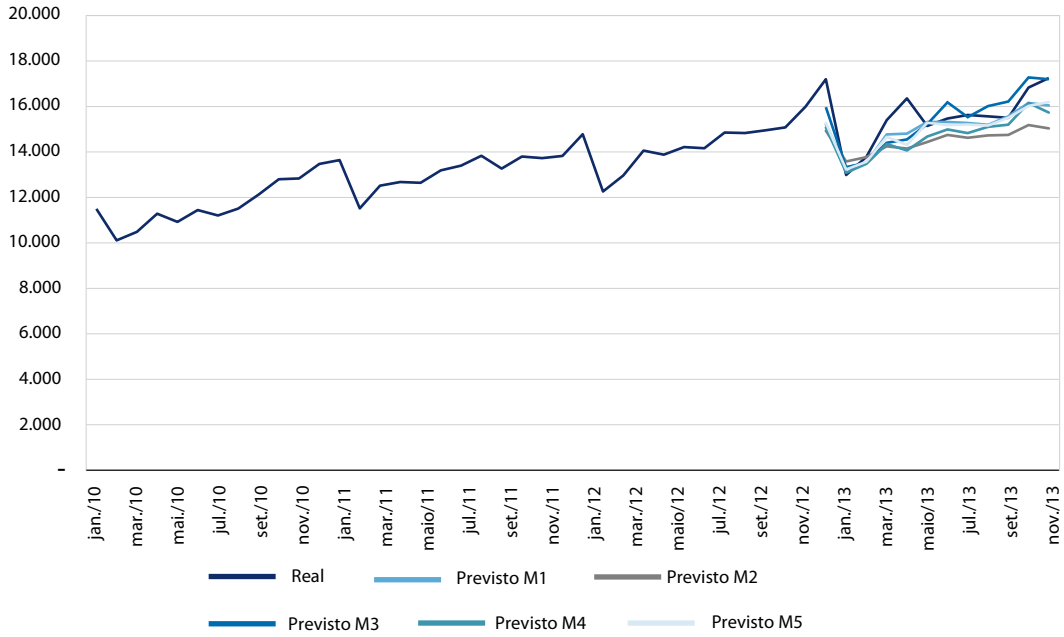
$$p(y_{T+h}|y)=\int p(y_{T+h}|\underline{\theta}_{T+h},\varphi,V)\,p(\underline{\theta}_{T+h}|\underline{\theta}_T,W,\Gamma)\,p(\underline{\theta}_T,W,\Gamma,V|y)\,d\underline{\theta}_{T+h}\,d\underline{\theta}_T\,d\varphi\,dV\,dW\,d\Gamma\,,$$

onde $(y_{T+h} \mid \underline{\theta}_{T+h}, \varphi, V) \sim N(F_{T+h}, V)$, $(\underline{\theta}_{T+h} \mid \underline{\theta}_T, W, \Gamma) \sim N(\mu_h, V_h)$, $\mu_h = G^h \underline{\theta}_t$ e $V_h = \sum_{j=1}^{h} G^{j-1} W (G^{j-1})'$, para $h > 0$. Portanto, se $\{(\underline{\theta}_T^{(1)}, \varphi^{(1)}, V^{(1)}, W^{(1)}, \Gamma^{(1)}), \ldots, (\underline{\theta}_T^{(L)}, \varphi^{(L)}, V^{(L)}, W^{(L)}, \Gamma^{(L)})$ é uma amostra de $p(\underline{\theta}_t, W, \Gamma, V \mid y)$, é fácil amostrar $\underline{\theta}_{T+h}^{(j)}$ de $p(\underline{\theta}_{T+h} \mid \underline{\theta}_T^{(j)}, W^{(j)}, \Gamma^{(j)})$, para todo $j = 1, \ldots, L$, tal que $p(y_{T+h} \mid y) = \frac{1}{L} \sum_{j=1}^{L} p(y_{T+h} \mid \underline{\theta}_{T+h}^{(j)}, \varphi^{(j)}, V^{(j)})$ é a aproximação via Monte Carlo de $p(y_{T+h} \mid y)$. Analogamente, se $y_{T+h}^{(j)}$ é amostrado de $p(y_{T+h} \mid \underline{\theta}_{T+h}^{(j)}, \varphi^{(j)}, V^{(j)})$, para $j = 1, \ldots, L$, então $\{y_{T+h}^{(1)}, \ldots, y_{T+h}^{(L)}\}$ representa uma amostra de $p(y_{T+h} \mid y)$.

APÊNDICE C

PREVISÃO

GRÁFICO C.1
Cofins



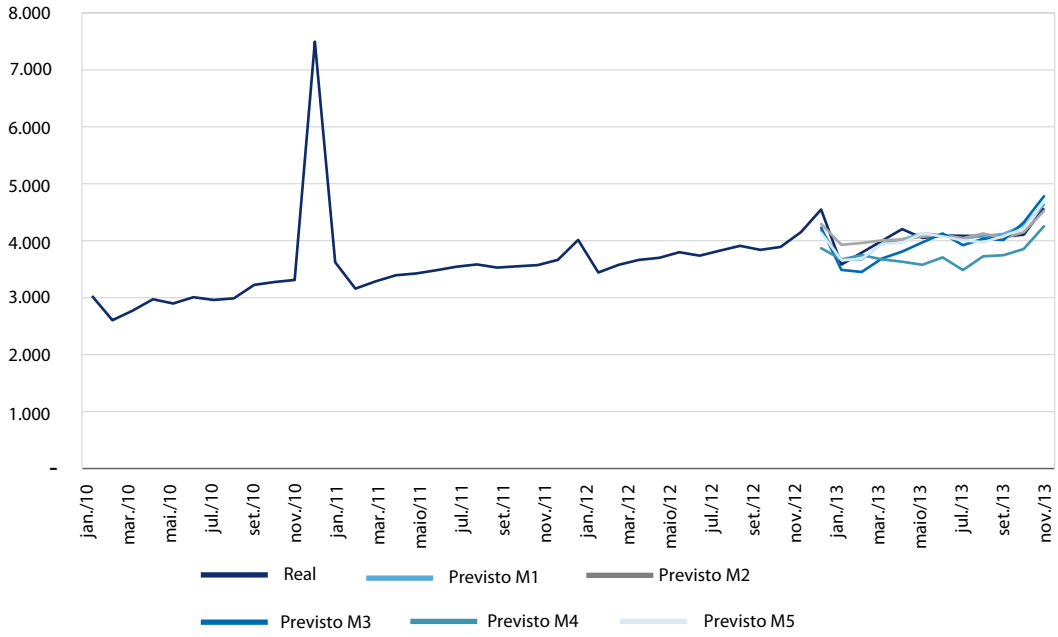
Elaboração dos autores.

TABELA C.1
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	3,53%	6,91%	-0,91%	5,54%	4,02%
RMSE[1]	R$ 872	R$ 1.372	R$ 784	R$ 1.122	R$ 948
MPE	3,29%	6,52%	-0.80%	5,30%	3,78%
MAPE	3,99%	7,29%	3,77%	5,43%	4,25%
MAD[1]	R$ 647	R$ 1.176	R$ 597	R$ 881	R$ 691

Elaboração dos autores.
Nota: [1] Em R$ milhões.

GRÁFICO C.2
CPIS



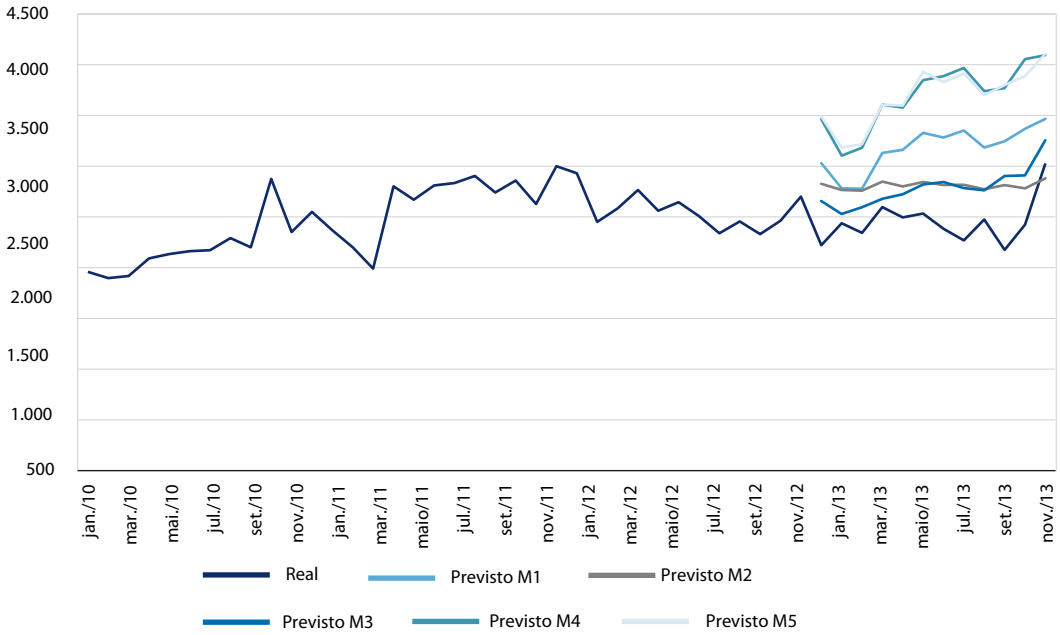
Elaboração dos autores.

TABELA C.2
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	0,67%	-0,20%	-2,75%	8,62%	1,26%
RMSE	R$ 140	R$ 146	R$ 225	R$ 412	R$ 157
MPE	0,59%	-0,40%	-2,81%	8,41%	1,19%
MAPE	2,40%	2,48%	4,58%	8,83%	2,80%
MAD	R$ 100	R$ 99	R$ 189	R$ 368	R$ 118

Elaboração dos autores.

GRÁFICO C.3
IOF



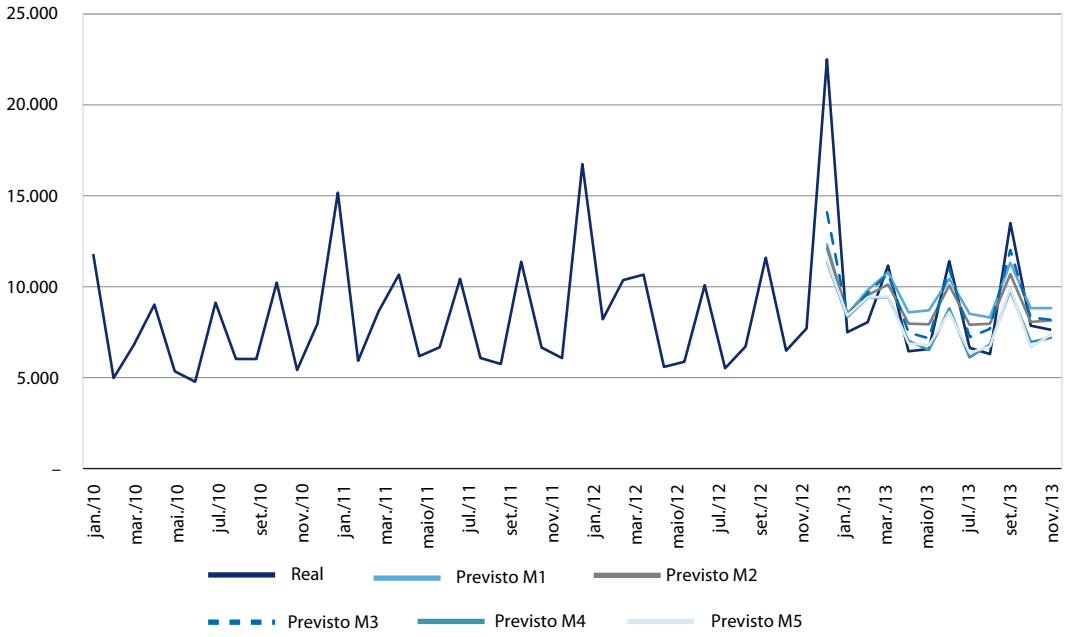
Elaboração dos autores.

TABELA C.3
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	-29,70%	-14,79%	13,89%	-50,77%	-50,64%
RMSE	R$ 766	R$ 412	R$ 385	R$ 1.281	R$ 1.270
MPE	-30,37%	-15,52%	14,37%	-51,52%	-51,38%
MAPE	30,37%	16,33%	14,37%	51,52%	51,38%
MAD	R$ 727	R$ 387	R$ 340	R$ 1.243	R$ 1.240

Elaboração dos autores.

GRÁFICO C.4
IRPJ



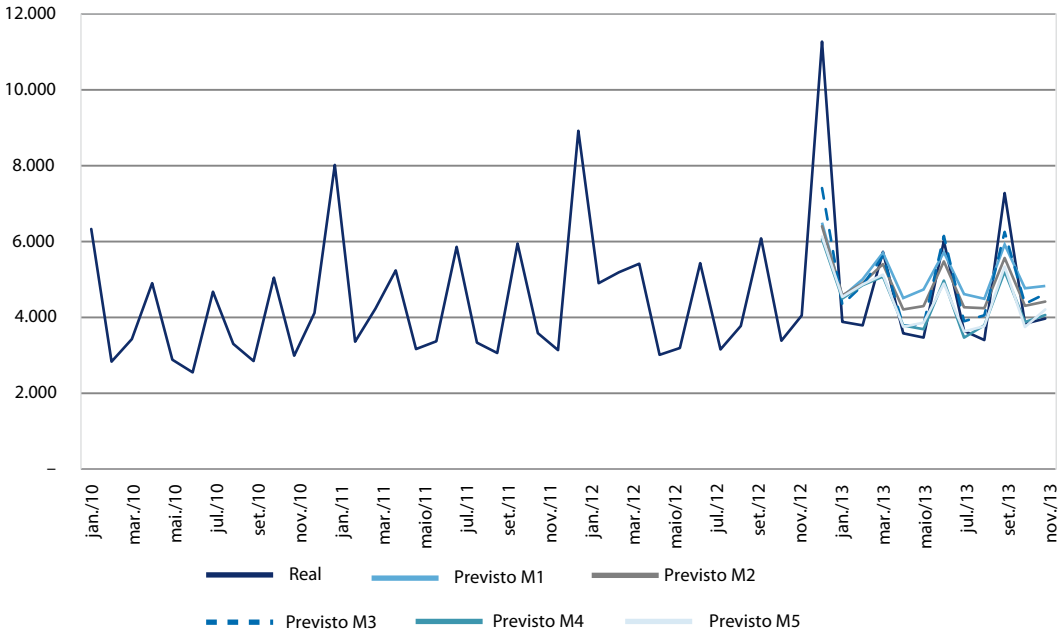
Elaboração dos autores.

TABELA C.4
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	0,53%	5,61%	-3,16%	15,41%	15,14%
RMSE	R$ 3.323	R$ 3.301	R$ 2.593	R$ 3.559	R$ 3.539
MPE	-9,65%	-3,65%	3,71%	7,99%	7,68%
MAPE	21,85%	18,30%	13,02%	15,53%	15,31%
MAD	R$ 2.232	R$ 2.053	R$ 1.484	R$ 2.032	R$ 2.016

Elaboração dos autores.

GRÁFICO C.5
CSLL



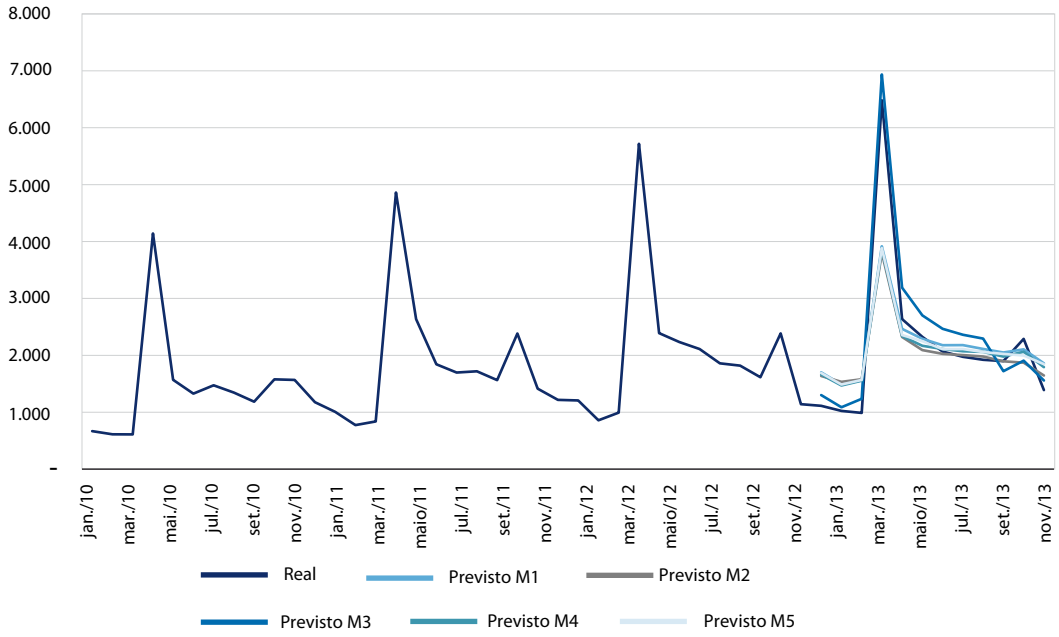
Elaboração dos autores.

TABELA C.5
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	-2,45%	3,00%	-1,23%	10,90%	10,03%
RMSE	R$ 1.654	R$ 1.616	R$ 1.247	R$ 1.700	R$ 1.678
MPE	-12,47%	-6,06%	5,25%	3,13%	2,16%
MAPE	23,54%	19,61%	13,62%	14,92%	15,21%
MAD	R$ 1.198	R$ 1.091	R$ 771	R$ 981	R$ 985

Elaboração dos autores.

GRÁFICO C.6
IRPF



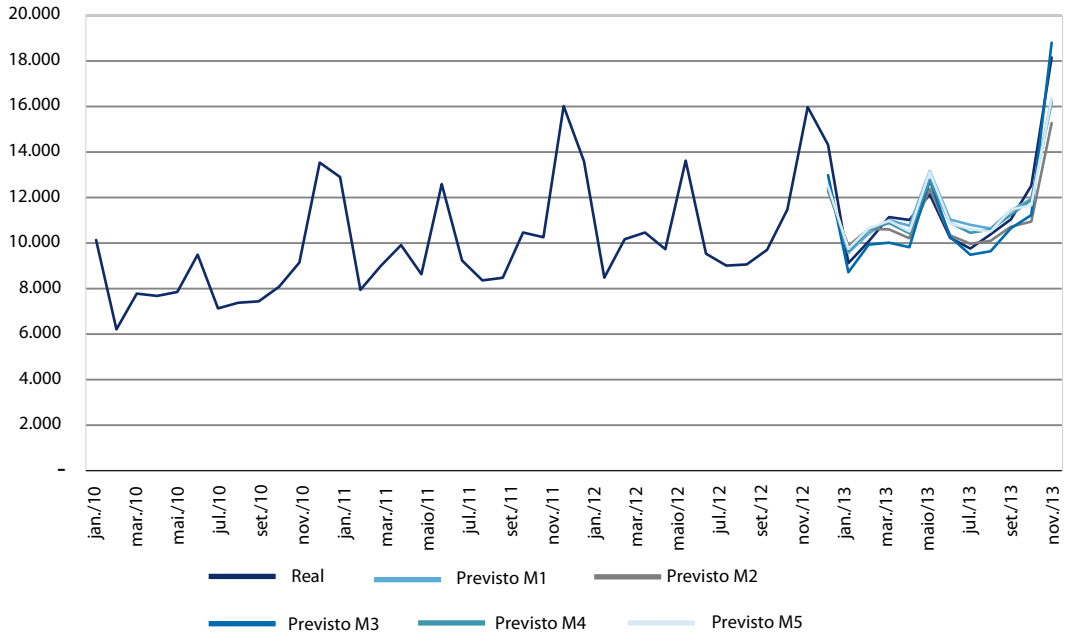
Elaboração dos autores.

TABELA C.6
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	0,76%	6,59%	10,17%	3,75%	2,80%
RMSE	R$ 810	R$ 840	R$ 343	R$ 819	R$ 820
MPE	-14,02%	-8,01%	11,42%	-10,78%	-12,05%
MAPE	23,28%	21,95%	15,76%	22,18%	23,15%
MAD	R$ 476	R$ 471	R$ 315	R$ 469	R$ 480

Elaboração dos autores.

GRÁFICO C.7
IRRF



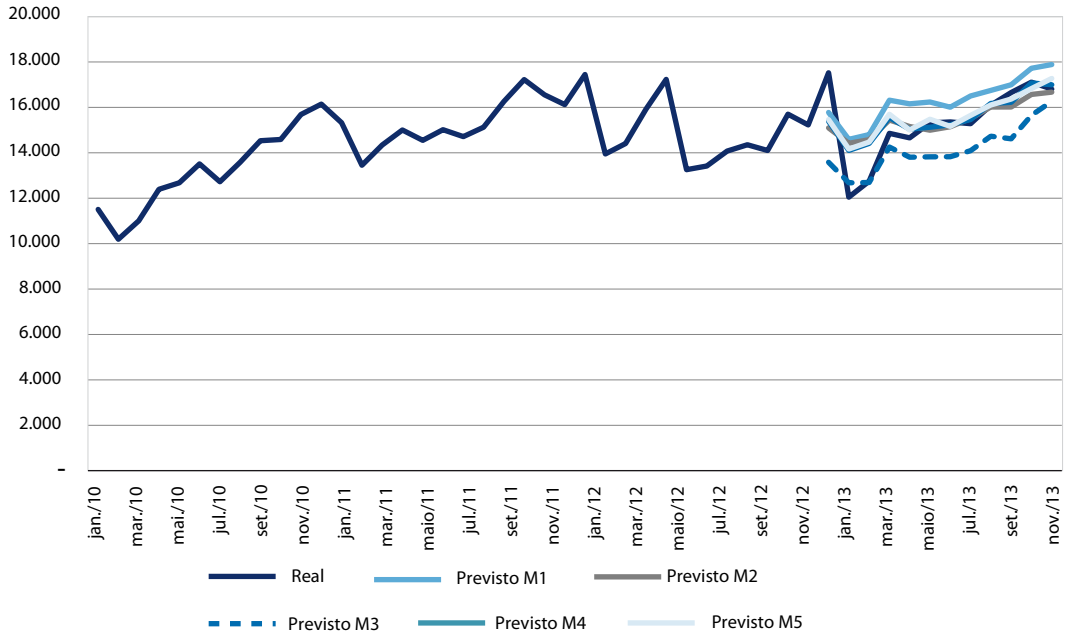
Elaboração dos autores.

TABELA C.7
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	0,25%	4,71%	-3,99%	1,37%	0,66%
RMSE	R$ 940	R$ 1.193	R$ 813	R$ 929	R$ 938
MPE	-0,84%	3,47%	-4,23%	0,28%	-0,42%
MAPE	6,05%	6,63%	5,76%	5,89%	6,34%
MAD	R$ 749	R$ 860	R$ 685	R$ 736	R$ 780

Elaboração dos autores.

GRÁFICO C.8
IPI



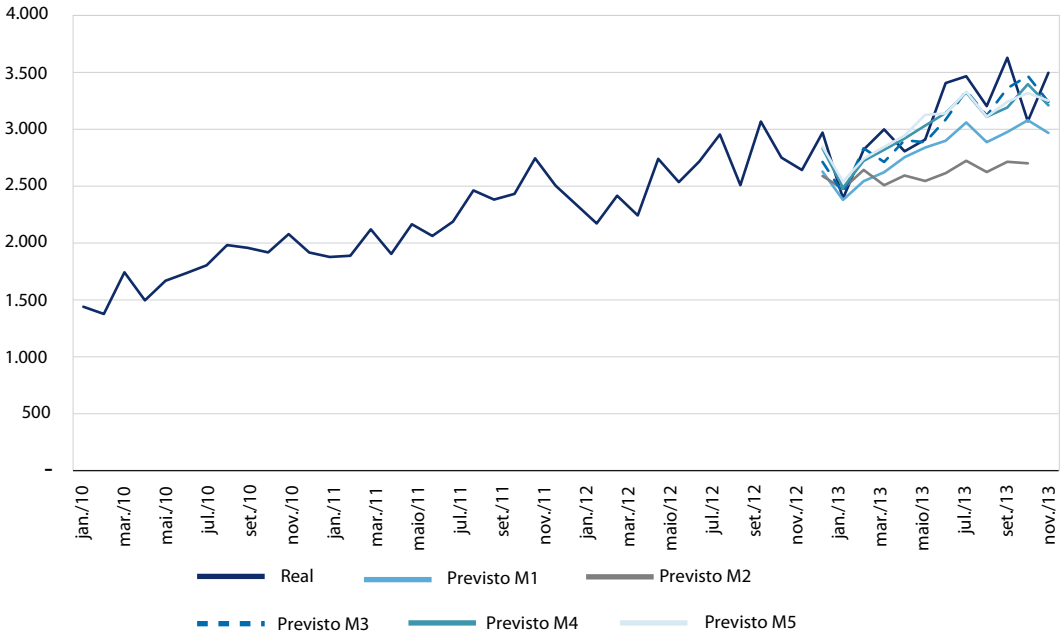
Elaboração dos autores.

TABELA C.8
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	-6,11%	-0,63%	-7,83%	-1,26%	-1,73%
RMSE	R$ 346	R$ 292	R$ 405	R$ 255	R$ 261
MPE	-6,79%	-1,43%	-7,30%	-1,93%	-2,41%
MAPE	8,46%	5,70%	8,18%	4,76%	5,14%
MAD	R$ 308	R$ 205	R$ 327	R$ 170	R$ 184

Elaboração dos autores.

GRÁFICO C.9
II



Elaboração dos autores.

TABELA C.9
Avaliação de desempenho da previsão

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Erro %	9,49%	15,78%	-2,87%	2,64%	1,97%
RMSE	R$ 362	R$ 573	R$ 220	R$ 219	R$ 205
MPE	8,99%	14,88%	-2,55%	2,19%	1,48%
MAPE	9,04%	15,46%	5,68%	6,00%	5,92%
MAD	R$ 296	R$ 503	R$ 181	R$ 191	R$ 186

Elaboração dos autores.

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Coordenação
Cláudio Passos de Oliveira

Supervisão
Andrea Bossle de Abreu

Revisão
Camilla de Miranda Mariath Gomes
Carlos Eduardo Gonçalves de Melo
Elaine Oliveira Couto
Elisabete de Carvalho Soares
Lucia Duarte Moreira
Luciana Bastos Dias
Luciana Nogueira Duarte
Míriam Nunes da Fonseca
Thais da Conceição Santos (estagiária)
Vivian Barros Volotão Santos (estagiária)

Editoração eletrônica
Roberto das Chagas Campos
Aeromilson Mesquita
Aline Cristine Torres da Silva Martins
Carlos Henrique Santos Vianna

Capa
Luís Cláudio Cardoso da Silva

Projeto Gráfico
Renato Rodrigues Bueno

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

Livraria Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo.
70076-900 – Brasília – DF
Fone: (61) 3315-5336

Correio eletrônico: livraria@ipea.gov.br

Composto em Adobe Garamond Pro 12/16 (texto)
Frutiger 67 Bold Condensed (títulos, gráficos e tabelas)
Rio de Janeiro-RJ

Logotipo/slogan de governo removido em observancia ao disposto no art. 40 da Lei n 9.504/1997

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

ISSN 1415-4765

ipea Instituto de Pesquisa Econômica Aplicada

Secretaria de
Assuntos Estratégicos