

Mário Jorge Mendonça et al.

Working Paper

Aplicação de modelos dinâmicos bayesianos para projeção de arrecadação tributária

Texto para Discussão, No. 2590

Provided in Cooperation with:

Institute of Applied Economic Research (ipea), Brasília

Suggested Citation: Mário Jorge Mendonça et al. (2020) : Aplicação de modelos dinâmicos bayesianos para projeção de arrecadação tributária, Texto para Discussão, No. 2590, Instituto de Pesquisa Econômica Aplicada (IPEA), Brasília, <https://doi.org/10.38116/td2590>

This Version is available at:

<https://hdl.handle.net/10419/240784>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

2590

APLICAÇÃO DE MODELOS DINÂMICOS BAYESIANOS PARA PROJEÇÃO DE ARRECADAÇÃO TRIBUTÁRIA

TEXTO PARA DISCUSSÃO

Mário Jorge Mendonça
Sérgio Ricardo B. Gadelha
Luís Alberto Medrano
Alex Felipe R. Lima
Marcus G. Lavagnole
Dermeson André Polli



APLICAÇÃO DE MODELOS DINÂMICOS BAYESIANOS PARA PROJEÇÃO DE ARRECADAÇÃO TRIBUTÁRIA¹

Mário Jorge Mendonça²
Sérgio Ricardo B. Gadelha³
Luis Alberto Medrano⁴
Alex Felipe R. Lima⁵
Marcus G. Lavagnole⁶
Dermeson André Polli⁷

1. Este estudo está inserido no âmbito do acordo de cooperação técnica entre a Secretaria de Política Econômica do Ministério da Economia (SPE/ME) e o Ipea.

2. Técnico de planejamento e pesquisa na Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) do Ipea. *E-mail*: <mario.mendonca@ipea.gov.br>.

3. Coordenador-geral de modelos e projeções econômico-fiscais da SPE/ME. *E-mail*: <sergio.gadelha@fazenda.gov.br>.

4. Assistente de pesquisa na Dirur/Ipea. *E-mail*: <luis.medrano@ipea.gov.br>.

5. Assessor da SPE/ME. *E-mail*: <alex.lima@fazenda.gov.br>.

6. Assistente de pesquisa na Dirur/Ipea. *E-mail*: <marcus.lavagnole@ipea.gov.br>.

7. Assessor da SPE/ME. *E-mail*: <demerson.polli@fazenda.gov.br>.

Governo Federal

Ministério da Economia

Ministro Paulo Guedes

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada ao Ministério da Economia, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Carlos von Doellinger

Diretor de Desenvolvimento Institucional

Manoel Rodrigues Junior

**Diretora de Estudos e Políticas do Estado,
das Instituições e da Democracia**

Flávia de Holanda Schmidt

**Diretor de Estudos e Políticas
Macroeconômicas**

José Ronaldo de Castro Souza Júnior

**Diretor de Estudos e Políticas Regionais,
Urbanas e Ambientais**

Nilo Luiz Saccaro Júnior

**Diretor de Estudos e Políticas Setoriais de Inovação
e Infraestrutura**

André Tortato Rauén

Diretora de Estudos e Políticas Sociais

Lenita Maria Turchi

**Diretor de Estudos e Relações Econômicas
e Políticas Internacionais**

Ivan Tiago Machado Oliveira

**Assessora-chefe de Imprensa
e Comunicação**

Mylena Fiori

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

Texto para Discussão

Publicação seriada que divulga resultados de estudos e pesquisas em desenvolvimento pelo Ipea com o objetivo de fomentar o debate e oferecer subsídios à formulação e avaliação de políticas públicas.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2019

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 publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos).
Acesse: <http://www.ipea.gov.br/portal/publicacoes>

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

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.

DOI:<http://dx.doi.org/10.38116/td2590>

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	7
2 METODOLOGIA.....	9
3 VALIDAÇÃO DA PROJEÇÃO	19
4 BASE DE DADOS E ESCOLHA DE REGRESSORES	20
5 RESULTADOS.....	21
6 CONSIDERAÇÕES FINAIS.....	30
REFERÊNCIAS	34
BIBLIOGRAFIA COMPLEMENTAR.....	36
APÊNDICE A	37

SINOPSE

Neste estudo, empreende-se uma análise econométrica, com vistas à projeção das séries desagregadas do Imposto sobre Operações Relativas à Circulação de Mercadorias e Prestação de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS), administradas pelo Conselho Nacional de Política Fazendária do Ministério da Economia (Confaz/ME). Três metodologias foram aplicadas: *i*) o modelo estrutural dinâmico (MED) – por meio da *bayesian structural time series* (BSTS); *ii*) o modelo linear dinâmico (MLD); e *iii*) o modelo fatorial dinâmico (MFD), todos estes estimados com base na prática bayesiana. Os exercícios econométricos objetivaram três tipos de resultados: *i*) a avaliação da projeção; *ii*) a elasticidade do tributo em relação ao fato gerador; e *iii*) a projeção sessenta meses à frente fora da amostra. Nossa base de dados é composta de dados no período entre janeiro de 2006 a dezembro de 2019. Tendo-se em vista a dificuldade para tratar as séries do Confaz devido à falta de regularidade, os exercícios feitos para validação da projeção apresentaram *performance* bastante razoável. De cerca de vinte séries para cada estado, aproximadamente 80% registram um erro percentual médio absoluto (Mape – em inglês, *mean absolute percentage error*) abaixo de 15%.

Palavras-chave: Confaz; modelos dinâmicos bayesianos; modelo fatorial dinâmico; Mape; projeção.

ABSTRACT

In this study, we undertook an econometric analysis aim to forecast the disaggregated series of ICMS administered by Confaz. Three methodologies were applied: *i*) the dynamic structural model (BSTS); *ii*) the dynamic linear model (MLD); and *iii*) the dynamic factorial model (MFD), all of them estimated based on Bayesian practice. The econometric exercises aimed at three types of results: *i*) the evaluation of the forecast; *ii*) the elasticity of the tax in relation to the relevant variable; and *iii*) the projection 60 months ahead outside the sample. Our database is composed of data from January 2006 to December 2019. Considering the difficulty to treat the Confaz series due to the lack of regularity, the exercises done to validate the forecast displayed a quite reasonable performance. From about 20 series for each state, approximately 80% register an MAPE below 15%.

Keywords: Confaz; Bayesian dynamic models; factorial dynamic model; MAPE; forecast.

1 INTRODUÇÃO

O objetivo deste estudo é apresentar uma proposta metodológica com vistas à projeção da arrecadação das séries de tributos no Brasil. Embora se tenha avançado nesse tema desde que se iniciaram as primeiras análises feitas para estimar a elasticidade da carga tributária bruta (Mendonça, Medrano e Sachsida, 2011), com o objetivo de depois evoluírem em direção ao tratamento das séries individuais de tributos, aumentando assim a capacidade de controle para o gestor de política. Exemplo disso é o estudo elaborado por Mendonça, Medrano e Sachsida (2015), em cuja amostra de tributos estão algumas das séries do imposto de renda e do Imposto sobre Produtos Industrializados (IPI), receita previdenciária, além da série agregada do Imposto sobre Operações Relativas à Circulação de Mercadorias e Prestação de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS). Mais recentemente, Lima, Polli e Gadelha (2019) aplicaram o modelo de combinação de previsão (Bates e Granger, 1969) em uma base de dados de arrecadação bruta de tributos federais administrados pela Secretaria Especial da Receita Federal do Brasil (RFB), obtendo-se projeções com grande grau de acurácia. No entanto, cabe notar que boa parte dessas séries de tributos federais é bem-comportada do ponto de vista estatístico, apresentando-se tendência e sazonalidade bem definidas de forma geral.

Infelizmente, essa não é situação comum com que se depara o gestor de política, de modo que a necessidade de extrair informação vai bem além daquela contida na arrecadação agregada ou nas séries de tributos com formato bem definido. Esse é o caso dos tributos de competência dos entes subnacionais, como os estados e o Distrito Federal (DF). Existe forte heterogeneidade entre esses entes no que diz respeito às suas economias, o que se reflete também na dinâmica de arrecadação. Além disso, tem-se de levar também em conta que, embora a arrecadação de um ente tenha como fator principal sua capacidade econômica, esta também deriva da eficiência de arrecadar desse ente. Portanto, quando se visa ao tratamento dos tributos dos entes subnacionais, é necessário propor instrumentos mais flexíveis capazes de lidar com séries de tempo que apresentam dinâmicas bastante complexas e díspares.

Em particular, estaremos focando na análise das séries de tributos administradas pelo Conselho Nacional de Política Fazendária do Ministério da Economia (Confaz/ME). Cabe ter em mente que as dinâmicas das séries de uma mesma categoria de tributo podem apresentar comportamentos bem distintos entre os estados, pois a estrutura econômica

de um estado muitas vezes é completamente distinta de outro. Por exemplo, a série do ICMS arrecadado em um setor difere sobremaneira conforme composição econômica dos estados. Naturalmente, um estado que tenha sua economia baseada na agricultura deve ter participação mais expressiva no setor primário que nos demais setores.

Com o intuito de contemplar todos esses fatores, os quais incorporam tendência, sazonalidade e diversidade de fatos geradores, propõe-se neste estudo uma metodologia para projeção de arrecadação de tributos. O método proposto baseia-se na utilização de três categorias de modelos dinâmicos estimados com base na abordagem bayesiana (Gamerman e Lopes, 1997; Kim e Nelson, 1999; West e Harrison, 1997; Durbin e Koopman, 2001), cujas vantagens serão externadas a frente. Não se trata de exercício de combinação de projeções, o que pode também ser implementado sem dificuldade, mas sim de propor metodologia capaz de contemplar diferentes aspectos das séries de tempo pouco provável de serem detectados pela aplicação de apenas um instrumento.

O primeiro modelo recai no emprego de conjunto composto de dez modelos estruturais bayesianos – por meio da *bayesian structural time series* (BSTS) regidos por um algoritmo que seleciona a melhor previsão, tendo-se em vista um critério predeterminado de avaliação da capacidade preditiva.¹ Cada modelo se diferencia por sua especificação garantindo grande flexibilidade quanto à especificação da forma funcional que pode variar com relação ao nível, à tendência e ao tratamento para a sazonalidade, assim como lidar com diferentes hipóteses acerca do distúrbio ou erro da regressão. Incorporam ainda métodos para lidar com tendência explosiva da variância; algo inerente a esse tipo de modelo e que implica o aumento da incerteza conforme o horizonte da projeção se expande.

A segunda abordagem faz uso do modelo linear dinâmico (MLD), com coeficientes variáveis que, entre outras vantagens, permitem obter as elasticidades dinâmicas de cada regressor. Isso pode ser justificado pela conjectura – que parece muito plausível *a priori* – de que as sucessivas mudanças na tributação estão associadas a mudanças recorrentes nas elasticidades relevantes

1. Como se verá adiante, o critério elegido para seleção do modelo quanto à sua performance preditiva é o erro percentual absoluto médio (Mape – em inglês, *mean absolute percentage error*).

Por fim, modelamos o conjunto de séries do Confaz por meio de modelo multivariado denominado de modelo fatorial dinâmico (MFD). Uma razão fundamental para o uso do MFD é que este permite obter ganho de informação importante, por meio da redução de dimensionalidade – isto é, de aumento no grau de agregação – das variáveis. Nesse caso, a informação contida nas inter-relações entre as variáveis é *passada para um conjunto menor de variáveis* composto pelos fatores. Os fatores possuem a habilidade de identificar a “verdadeira” estrutura dos dados. Do ponto de vista prático, como é modelo multivariado, o MFD tem a vantagem de projetar todas as séries simultaneamente.

Além desta introdução, o texto está estruturado da seguinte forma. A seção 2 tem como objetivo descrever os diferentes modelos que se usam para projeção, apresentando as especificações econométricas, suas motivações teóricas e métodos de correção. A seção 3 apresenta os critérios de avaliação da projeção e detalha o procedimento utilizado para levar a cabo a validação. A seção 4 descreve a base de dados e a escolha dos regressores. A seção 5 revela e analisa os resultados. Por fim, realizam-se as considerações finais. Detalhes sobre a metodologia econométrica utilizada foram acrescentados no apêndice.

2 METODOLOGIA

Os modelos utilizados neste trabalho são casos particulares de abordagem mais geral denominada de modelo de espaço de estado. Tais modelos têm como características básicas a linearidade e a normalidade para o erro. Nos modelos de espaço de estado, a estimação e a previsão podem ser obtidas recursivamente pelo conhecido filtro de Kalman. Esses modelos podem ser usados para modelar séries de tempo na presença de não estacionaridade, mudanças estruturais e padrões irregulares. Se a série se mostra bastante previsível, uma vez que suas características básicas se repetem regularmente, temos o caso de uma tendência e uma sazonalidade bastante regular com uma variabilidade ligeiramente crescente, o que facilita sua previsão futura.

O experimento torna-se complexo quando as séries temporais não assumem um padrão regular e não apresentam propriedades de estacionariedade, mas pode incluir pontos de mudança na tendência ou nas quebras estruturais. A análise de séries de tempo não estacionárias com modelos autorregressivos de médias móveis (*Arma* – em inglês, *autoregressive-moving-average*) requer, pelo menos, uma transformação preliminar dos

dados, de modo que estes adquiram aparência mais tratável. No entanto, é desejável ter meios para analisar os dados mais diretamente, sem desconsiderar a instabilidade na média e na variância, as quebras estruturais e os saltos. Os modelos de espaço de estado incluem os modelos Arma como um caso especial, mas pode ser aplicado em séries não estacionárias sem necessidade de transformação preliminar dos dados.

A estrutura mais geral dos modelos gaussianos lineares de espaço de estado nos quais estão incluídos modelos estruturais propostos neste estudo é especificada por uma distribuição *a priori* normal para o vetor θ de dimensão p no tempo $t=0$, de modo que $\theta(0) \sim N_p(m_0, C_0)$, com o par de equações em que para cada $t \geq 1$, tem-se:

$$Y(t) = F(t)\theta(t) + v(t), \quad (1)$$

$$\theta(t) = G(t)\theta(t-1) + w(t). \quad (2)$$

Os distúrbios $v(t)$ e $w(t)$ são independentes e gaussianos, de modo que:

$$v(t) \sim N_m(0, V(t)), \quad (3)$$

$$w(t) \sim N_p(0, W(t)). \quad (4)$$

As matrizes $G(t)$ e $F(t)$ são de ordem, respectivamente, $p \times p$ e $m \times p$. $v(0) \sim N_p(m_0, C_0)$ $\theta(0) \sim N_p(m_0, C_0)$. A equação (1) é conhecida como equação de observação, enquanto a equação (2), como equação do estado.

Diferentemente do modelo definido pelas equações (1) e (2), o modelo de espaço de estado mais geral pode ser especificado a partir de distribuição *a priori* para $\theta(0)$, com as equações de observação e estado:

$$Y(t) = h(t)[\theta(t), v(t)], \quad (5)$$

$$\theta(t) = g(t)[\theta(t-1), w(t)]. \quad (6)$$

para funções arbitrárias $g(t)$ e $h(t)$. Os modelos de espaço de estado linear especificam essas funções como lineares, e os modelos lineares gaussianos acrescentam as suposições

de normalidade para os distúrbios $v(t)$ e $w(t)$. A hipótese de normalidade é plausível em muitas aplicações. No entanto, existem muitas extensões importantes, tais como erros de cauda pesadas para modelagem de *outliers*, bem como o MLD generalizado para o tratamento de séries temporais discretas. Ao retirar-se a hipótese de normalidade, tem-se como consequência o surgimento de maiores dificuldades computacionais.

A abordagem básica proposta parte do princípio de que qualquer série temporal econômica $y(t)$ pode ser decomposta da seguinte forma:

$$y(t) = f(t) + \varepsilon(t), \quad (7)$$

em que $f(t)$ é a parcela sistemática ou regular da série, e $\varepsilon(t)$ é o termo estocástico e independente de $f(t)$. O termo regular, por sua vez, admite a seguinte decomposição:

$$f(t) = \mu(t) + S(t) + Z(t), \quad (8)$$

na qual $\mu(t)$ é a componente autônoma da série; $S(t)$ são as flutuações cíclicas e sazonais; e $Z(t)$ é o efeito que determinadas variáveis, ditas regressores, exercem sobre $y(t)$. Cada um dos métodos que serão usados para modelar as séries de tributos do Confaz admite variantes para cada termo da parte sistemática da série. Observe agora como os três modelos se enquadram na estrutura definida pela equação (7), tendo-se em vista a análise das séries de tributos do Confaz.

2.1 Modelos estruturais bayesianos (BSTS)²

Nessa classe de modelos, a dinâmica é imposta a partir do termo autônomo, o qual se denomina como nível. Neste estudo, trabalha-se com duas especificações básicas para tratamento do nível que estão representadas, respectivamente, pelas equações (8) e (9). Na equação (8), o nível assume uma tendência estocástica imposta pela trajetória do parâmetro $\delta_i(t)$; modelo que se denomina nível e tendência. Na equação (9), a trajetória do nível segue um passeio aleatório. Também nas regressões geradas por meio dos modelos estruturais, os parâmetros associados aos regressores são invariantes no tempo. Na seção 3, será discutida a questão dos regressores.

2. Usa-se a rotina BSTS (Scott e Varian, 2014) disponível no *software* R para estimar os modelos estruturais.

modelo estrutural com nível e tendência estocástica:

$$\begin{aligned}
 ICMS_i(t) &= \mu(t)_i + \sum_k^K \beta_{ik} VAR_{ik}(t) + S_i(t) + v_i(t) \\
 \mu_i(t) &= \mu_i(t-1) + \delta_i(t) + w_{1i}(t) \\
 \delta_i(t) &= \delta_i(t-1) + w_{2i}(t) \\
 k &= 1, \dots, K
 \end{aligned} \tag{9-11}$$

modelo estrutural no nível:

$$ICMS_i(t) = \mu(t)_i + \sum_k^K \beta_{ik} VAR_{ik}(t) + S_i(t) + v_i(t), \tag{12}$$

$$\mu_i(t) = \mu_i(t-1) + w_{1i}(t). \tag{13}$$

Neste modelo, $\mu_i(t)$ e $\delta_i(t)$ são parâmetros que evoluem temporalmente e representam, respectivamente, o nível e a tendência estocástica; $w_{1i}(t)$ e $w_{2i}(t)$ são os distúrbios associados às equações que regem os processos geradores $\mu_i(t)$ e $\delta_i(t)$. O índice i refere-se ao estado da Federação; k é o índice referente à variável explicativa; β_{ik} é o parâmetro associado à variável exógena VAR_{ik} ; e $S_i(t)$ é a componente sazonal. Os distúrbios assumem as seguintes distribuições $w_1(t) \sim N(0, W_1)$ e $w_2(t) \sim N(0, W_2)$.

A sazonalidade $S_i(t)$ é uma característica idiossincrática de cada série de tributos, o que faz com que a aplicação de um método puramente artificial e *ad hoc* contribua para diminuir a qualidade do ajustamento e a previsão. O método adotado neste estudo permite mitigar essas dificuldades na medida em que a sazonalidade é tratada como sendo um conjunto de parâmetros a mais no modelo. Além disso, são empregados dois métodos distintos para modelar a sazonalidade.

1) Forma trigonométrica

Nessa forma, tem-se a presença de um componente sazonal, S_p , que pode ser modelado segundo a representação do tipo de Fourier (West e Harrison, 1997), de modo que:

$$S_i(t) = \sum_{j=1}^2 \left[a_{ij} \operatorname{sen}\left(\frac{2\pi j t}{4}\right) + b_{ij} \cos\left(\frac{2\pi j t}{4}\right) \right] = Z_i \gamma_i, \tag{14}$$

em que:

$$Z_t = \left(\text{sen}\left(\frac{2\pi t}{4}\right) \quad \text{sen}\left(\frac{2\pi 2t}{4}\right) \quad \cos\left(\frac{2\pi t}{4}\right) \quad \cos\left(\frac{2\pi 2t}{4}\right) \right)^T, \quad (15)$$

e

$$\gamma_i = (\gamma_{i1}, \gamma_{i2}, \gamma_{i3}, \gamma_{i4}). \quad (16)$$

Cada componente da soma entre os colchetes se denomina harmônico. A vantagem dessa representação é a economia que se pode fornecer, pois, se algum harmônico for pouco relevante, este pode ser excluído. Nem sempre se necessita de todos estes. Por exemplo, com dados mensais, pode-se modelar com o primeiro e o quarto harmônicos. O primeiro representaria o ciclo anual, e o quarto harmônico corresponderia ao ciclo trimestral. 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.³ Outra motivação para o uso dessa abordagem, além da flexibilidade, é a facilidade de obter-se interpretação, uma vez que as trajetórias em forma de ciclos podem ser associadas aos variados padrões observados na prática.

2) *Dummies* sazonais

O modelo de *dummies* sazonais pode ser pensado como uma regressão em variáveis *dummy* das “periodicidades” em um período. Se houver N “periodicidades”, então o vetor estado é de dimensão $N - 1$. O primeiro elemento do vetor de estado obedece à seguinte equação:

$$S_{1i}(t) = -\sum_{n=2}^N S_{in}(t) + \eta_i(t). \quad (17)$$

O somatório dos coeficientes deve necessariamente ser igual a 1.

2.1.1 Problemas na previsão de longo prazo

No modelo de nível local, o estado evolui de acordo com um passeio aleatório:

$$\mu_i(t) = \mu_i(t-1) + w_i(t). \quad (18)$$

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

Com $w_i(0) \sim N(0, W_i)$, temos então que $\mu_i(t) \sim N(0, tW_i)$ para $t > 0$. Verifica-se, portanto, que a variância cresce com t . A tendência linear local é ainda mais volátil. À medida que a projeção se estende, a flexibilidade proporcionada por esses modelos torna-se uma espada de duas pontas, uma vez que a flexibilidade local de curto prazo se traduz em variação extrema a longo prazo. Uma alternativa é substituir o passeio aleatório por um processo autorregressivo (AR) estacionário, de modo que:

$$\mu_i(t) = \rho\mu_i(t-1) + w_i(t). \quad (19)$$

Com $|\rho| < 1$, tem-se que $\mu_i(t) \sim N(0, W_i(1-\rho))$. Isso significa que a incerteza cresce para uma assíntota finita.

Outro modo de lidar com o crescimento da variância se faz pelo emprego do modelo “híbrido”. Esse modelo modifica a tendência linear local, ao substituir o passeio aleatório por um processo AR(1) estacionário, revertendo a trajetória da tendência para o nível do processo, de modo que:

$$\begin{aligned} \mu_i(t) &= \mu_i(t-1) + \delta_i(t) + w_{1i}(t) \\ \delta_i(t) &= D + \rho(\delta_i(t-1) - D) + w_{2i}(t). \end{aligned} \quad (20)$$

O parâmetro D é a inclinação de longo prazo da componente tendencial, para a qual $\delta(t)$ eventualmente reverterá. No entanto, $\delta(t)$ pode ter desvios autorregressivos de curto prazo da tendência de longo prazo, com memória determinada por ρ . Quanto menor for o valor de ρ , maior será a rapidez com que a tendência retorne à média diante de um choque D . Neste estudo, refere-se a esse modelo pelo termo *tendência linear semilocal*.

2.1.2 Tratamento de *outliers*

Como modo de considerar a presença de observações aberrantes ou *outliers*, usa-se uma versão do modelo de tendência linear local, o qual assume que o distúrbio tem distribuição *t-Student*, em vez de considerar erros gaussianos. Esse é um modelo útil para previsões de curto prazo, quando a média da série temporal exhibe saltos ocasionais. O erro *t-Student* pode ser introduzido na equação de observação. Permitir erros com cauda pesada na equação de observação torna o modelo robusto contra *outliers* individuais, enquanto introduzir erro com cauda pesada no modelo de estado fornece robustez contra mudanças súbitas e persistentes de nível ou na tendência. Isso pode levar a limites de

previsão mais apertados do que os modelos gaussianos, quando se modelam dados que foram poluídos por *outliers*.

2.1.3 Conjunto de modelos e validação da previsão

Tendo-se detalhado as diversas variantes que as componentes de uma regressão podem assumir incorporando ainda os tratamentos disponíveis para correção da incerteza e presença de *outliers*, lista-se o conjunto de modelos que serão utilizados para projetar a trajetória futura de um tributo, com base na melhor projeção de acordo com um critério de validação da previsão.

- Arima + sazonalidade (Sarima);
- nível + tendência + AR;
- nível + AR;
- nível + tendência + AR + sazonalidade;
- nível + AR + sazonalidade;
- nível + tendência semilocal + AR + trigonométrica;
- nível + AR + trigonométrica;
- nível + tendência T + AR;
- nível + tendência T + AR + sazonalidade; e
- nível + tendência T + AR + trigonométrica.

Cada um desses modelos projeta o tributo com base nos valores predeterminados das variáveis exógenas. Decorre disso que, para projetar valores futuros do tributo, se deve dispor igualmente dos valores projetados para os regressores. Portanto, é necessário que se tenha também meios para também projetar as variáveis exógenas. Essa é uma situação mais realística, pois não se sabe os valores dessas variáveis quando a projeção para o tributo se fará fora da amostra. Da mesma forma, usa-se o mesmo conjunto de modelos para projetar as variáveis exógenas. Para cada variável exógena, obtém-se a projeção gerada pelo melhor modelo, tal como indicado pelo valor do indicador Mape para um horizonte de sessenta meses; e, condicionado à melhor previsão das variáveis exógenas entre os dez modelos, consegue-se a melhor projeção do ICMS entre aquelas geradas pelo mesmo conjunto de modelo e mesmo horizonte temporal.

2.2 Modelo linear dinâmico

O MLD definido pela equação (10) tem como diferença básica em relação ao modelo estrutural o fato de que os parâmetros dos regressores variam no tempo, o que implica que a tendência da variável dependente deverá ser imposta pelo regressor:

$$\begin{aligned} ICMS_i(t) &= \mu_{i0}(t) + \sum_k^K \beta_{ik}(t) VAR_{ik}(t) + S_i(t) + v_i(t) \\ \beta_{ik}(t) &= \varphi_k \beta_{ik}(t-1) + w_{ki}(t), \\ \mu_i(t) &= \mu_i(t-1) + w_{2i}(t) \\ k &= 1, \dots, K \end{aligned} \quad , \quad (21-23)$$

em que o índice i se refere ao estado ou à região, φ_k é o parâmetro autorregressivo das equações de estado para cada parâmetro, β_{ik} é o parâmetro associado à variável exógena VAR_{ik} e $S_i(t)$ é o componente sazonal. Os distúrbios assumem as seguintes distribuições: $w_{ki}(t) \sim N(0, W_{ki})$ e $w_i(t) \sim N(0, W_i)$ e $v_i(t) \sim N(0, V_i)$.

Ambos os modelos (estrutural e linear dinâmico) podem ser vistos como casos particulares da forma geral do modelo de espaço de estado representado pelas equações (1) e (2). Os procedimentos para estimação e projeção desses dois modelos tendo por base a abordagem bayesiana estão descritos no apêndice.

2.3 Modelo fatorial dinâmico

O modelo fatorial 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. 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.⁴ Mais detalhes a respeito das vantagens do uso de modelos fatoriais podem ser encontrados em Bartholomew (1987).

Conforme é conhecido, o MFD é capaz de lidar com um vasto conjunto de séries simultaneamente. Essa é uma característica do modelo que é útil àquele que trabalha com grande quantidade de dados e que tem necessidade de tomar decisão de forma rápida. Também pode ser de boa valia quando se tem pouca informação com respeito ao fato gerador pelo fato do modelo fatorial fazer uso eficiente das correlações existentes entre as variáveis.

Para estimação do modelo, aplicam-se aqui os 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 “frequentistas” (Gelman *et al.*, 2003, p. 696).

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

$$\begin{aligned}
 y(t) &= \beta f(t) + S(t) + e(t) \\
 f_i(t) &= \sum_j^J \beta_{ij}(t) VAR_{ij}(t) + v(t) \\
 \beta_{ij}(t) &= \varphi_{ij} \beta_{ij}(t-1) + w_j(t), \\
 j &= 1, \dots, J \\
 i &= 1, \dots, k
 \end{aligned}
 \tag{24-26}$$

em que y_t , f_t , e_t e w_t são vetores de dimensão m , k , m e k , respectivamente; β é uma matriz $m \times k$, $\Gamma_i = \text{diag}(\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{ik})$, $i = 1, \dots, p$, $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$; e $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ com $e(t) \sim N(0, \Sigma)$ e $w \sim N(0, \Lambda)$.

4. Não há interpretação no que se refere ao que esses fatores representam, uma vez que são combinações lineares das variáveis originais.

Para algum inteiro positivo específico $k \leq m$, o modelo das equações (7)-(8) relaciona cada \mathcal{Y}_t com um vetor k – dimensional de fatores comuns f_t explicados por um conjunto de variáveis explicativas $\{VAR_{j=1}^J\}$; φ_j é o componente autorregressivo de cada parâmetro; β_j é associado à variável exógena VAR_j e que tal como no modelo linear dinâmico varia no tempo; e $S_j(t)$ é o componente sazonal. Os distúrbios assumem a seguinte distribuição: $w_j(t) \sim N(0, W_j)$. Esse modelo tem uma peculiaridade, pois permite avaliar a dinâmica dos parâmetros associados às variáveis exógenas. Conforme mostra a equação (8), cada fator f_t é modelado por um conjunto de variáveis exógenas. Tais variáveis podem ser consideradas como preditores para as séries do Confaz. Detalhes metodológicos com respeito à estimação e à previsão do MDF com base na inferência bayesiana podem ser encontrados em Mendonça, Santos e Martins (2009) e Mendonça e Medrano (2015).

2.4 Comentários sobre a inferência bayesiana

Na estimação econométrica, muitas vezes, surge a questão relacionada ao problema de ordem de integração devido as séries serem não estacionárias. Quanto a esse ponto, deve-se ressaltar que a abordagem bayesiana minimiza o problema da ordem de integração das séries, bem como diminui a relevância do tamanho da amostra. Diversos autores (DeJong e Whiteman, 1991; Koop, 1992; Sims e Uhlig, 1991) advogam vigorosamente em favor da alternativa bayesiana sobre a abordagem clássica mais tradicional usada no que se refere à questão da raiz unitária. Por exemplo, sabe-se que os testes do tipo Dickey-Fuller aumentado (ADF) são de baixa potência ante alternativas plausíveis, especialmente em relação à hipótese alternativa de tendência estacionária. A abordagem bayesiana, por sua vez, revelaria que as hipóteses de raiz unitária e de tendência estacionária apresentariam probabilidades bastante similares quanto às suas funções *a posteriori*. Assim, a abordagem bayesiana fornece tratamento mais fidedigno da informação amostral que a abordagem clássica.

Outro problema com os testes clássicos de raiz unitária é a descontinuidade gerada na teoria assintótica (Sims, 1988). A abordagem bayesiana, por ser baseada na função de distribuição *a posteriori*, não apresenta o problema da descontinuidade. Koop (1992) assinala que na abordagem clássica os valores críticos gerados a partir de pequenas amostras podem diferir substancialmente dos valores críticos assintóticos. A abordagem bayesiana, visto que é condicional à amostra observada, fornece resultados para pequenas amostras mais exatos. Em resumo, o problema da raiz unitária não é um ponto crítico na estatística bayesiana. Por fim, diferentemente da abordagem clássica, a

prática bayesiana não é dependente do teorema central do limite, o que elimina o uso de propriedades assintóticas (Gelman *et al.*, 2003, p. 696). Isto traz uma importante vantagem em termos da habilidade de obter estimativas mais confiáveis em contexto de pequenas amostras e alta dimensionalidade paramétrica.

3 VALIDAÇÃO DA PROJEÇÃO

Nesta seção, apresentam-se os resultados referentes à avaliação da previsão na amostra com variáveis exógenas projetadas (validação), por meio dos modelos descritos na seção 2. Foram usados diferentes critérios concorrentes para avaliar a capacidade preditiva do modelo. Entre estes, há a raiz do erro quadrático médio (RMSE),⁵ que penaliza os *outliers*, o desvio absoluto médio (MAD – em inglês, *mean absolute deviation*),⁶ o erro percentual médio absoluto,⁷ o erro percentual médio (MPE – em inglês, *mean percentage error*)⁸ e o erro acumulado percentual absoluto (Eapa).⁹

Costuma-se utilizar o Mape para avaliar a qualidade da previsão por sua facilidade de interpretação. No entanto, este é sensível ao volume de dados, não sendo aconselhável seu emprego quando se trabalha com poucos dados históricos. Cabe ressaltar que, embora pouco considerado do ponto de vista teórico, o critério do erro de Eapa é importante em termos de praticidade quando se trata da arrecadação de tributos, pois esse é um critério adotado pelos órgãos responsáveis em acompanhar a arrecadação tributária no Brasil. Isso se explica pelo fato de estarem interessados na previsão de arrecadação acumulada para determinado horizonte temporal. Nesse caso, um erro de previsão negativo em dado mês pode ser compensado por um erro positivo. Assim, embora um modelo possa

5. $RSME = N^{-1} \sum_{t=1}^N e_t^2$ em que $e_t = Y_t - \hat{F}_t$, em que Y_t é o valor observado e \hat{F}_t o valor previsto.

6. $MAD = N^{-1} \sum_{t=1}^N |e_t|$.

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

8. $MPE = \left[N^{-1} \sum_{t=1}^N \frac{e_t}{y_t} \right] \times 100$.

9. $EAPA = \sum_{t=1}^N e_t \sum_{t=1}^N y_t \times 100$.

apresentar erros discrepantes pontualmente, este pode ser útil em termos de projetar a arrecadação na média que, no cômputo geral, é capaz de gerar um resultado satisfatório.

Usa-se ainda o coeficiente U de Theil, que avalia o desempenho da previsão em relação à previsão ingênua ou trivial representada por um passeio aleatório. Previsão ingênua ou trivial significa que a estimativa do valor futuro é igual ao valor atual. O coeficiente U de Theil ou Theil U^{10} analisa a qualidade de uma previsão por meio dos seguintes valores: $U > 1$ significa que o erro do modelo é maior do que da previsão ingênua; $U < 1$ quer dizer que o erro do modelo é menor do que da previsão ingênua. O coeficiente U de Theil menor do que 1 já indica uma previsão melhor que a previsão ingênua; quanto mais próximo este for de 0, melhor será o resultado da previsão.

4 BASE DE DADOS E ESCOLHA DE REGRESSORES

O Conselho Nacional de Política Fazendária disponibiliza informações mensais sobre o comportamento do comércio entre as Unidades Federativas (UFs). A disponibilização do boletim comercial é com base nos dados das notas fiscais eletrônicas (NFe), que se encontram depositadas no “ambiente nacional”. O Confaz disponibiliza dados mensais de tributos dos entes subnacionais, entre os quais se situam as diversas séries do ICMS, do Imposto sobre a Propriedade de Veículos Automotores (IPVA), do Imposto sobre Transmissão Causa Mortis e Doação de Quaisquer Bens ou Direitos (ITCD) etc. O ICMS aparece para diferentes níveis de desagregação. Por exemplo, tem-se a versão desse imposto que é arrecadada nos setores primário e terciário, a versão que é desagregada do setor terciário, além de outras. Para cada uma das 26 UFs, tem-se dezessete séries genuínas,¹¹ o que perfaz um total de 442 series. No entanto, aumenta-se quando se incluem os tributos agregados. As séries disponibilizadas estão listadas no apêndice.

A abordagem parte da hipótese que o ICMS pode ser previsto com base no comportamento de determinadas variáveis, ditas predictoras, e que são tomadas como

$$10 \text{ Theil } U = \frac{1}{T} \frac{(f_{t+1} - y_{t+1})}{y_t} \cdot \frac{1}{T} \frac{(y_{t+1} - y_t)}{y_t}$$

11 .

variáveis exógenas para modelar as séries individuais do Confaz. No entanto, com relação à escolha dos regressores, depara-se com certas dificuldades. Por experiência, sabe-se que variável fundamental usada como fato gerador para explicar a arrecadação é o produto interno bruto (PIB). Contudo, essa variável não está em nível regional para uma frequência mensal. Como alternativa, emprega-se o Índice de Atividade Econômica (IBC-Br), disponibilizado pelo Banco Central do Brasil (BCB). Outras variáveis também foram incluídas entre os regressores. Entre estas, tem-se o índice Volume de Vendas – Varejo da Pesquisa Mensal do Comércio (IV/PMC) e o Índice Nacional de Preços ao Consumidor Amplo (IPCA), ambos do Instituto Brasileiro de Geografia e Estatística (IBGE).

Naturalmente, cada uma dessas variáveis é de âmbito regional, assim não necessariamente se usa o mesmo conjunto de regressores para todas as séries do ICMS, na medida em que uma mesma variável pode não estar disponível para todos os entes subnacionais, como é o caso do IPCA, que não é calculado para todas as capitais do Brasil, do IBC-Br, que não está disponível para o DF, etc. Assim, a base reduz os entes subnacionais para os quais as variáveis explicativas estão disponibilizadas:

- Norte: Amazonas (AM) e Pará (PA);
- Nordeste: Bahia (BA), Ceará (CE) e Pernambuco (PE);
- Centro-Oeste: Distrito Federal e Goiás (GO);
- Sudeste: Minas Gerais (MG), Rio de Janeiro (RJ), Espírito Santo (ES) e São Paulo (SP); e
- Sul: Paraná (PR), Rio Grande do Sul (RS) e Santa Catarina (SC).

Para os entes subnacionais que não dispõem do IPCA (ES, SC e PA), usa-se a mesma variável desse ente da mesma região. Quando não se dispõe do IBC-Br, como é o caso do DF, emprega-se o Índice de Emprego Formal (IEF). Por fim, a amostra é composta de dados que abrangem o período de janeiro de 2006 a dezembro de 2019.

5 RESULTADOS

Nesta seção, apresentam-se os resultados obtidos a partir dos modelos descritos ao longo da seção 2. Conforme foi dito, estimam-se o modelo estrutural dinâmico (MED), o modelo linear dinâmico e o modelo fatorial dinâmico. Os exercícios levados a cabo foram os seguintes: a avaliação da capacidade preditiva em amostra para o período entre

janeiro de 2018 a dezembro de 2019; as elasticidades em relação à renda (IBC-Br); e a projeção sessenta passos à frente para o período de janeiro de 2020 a dezembro de 2024. Para o MED, conforme já apontado, apresenta-se a melhor previsão do conjunto dos dez modelos estruturais listados na subseção 2.1.3.

Conforme dito, a amostra é composta de dados mensais em termos correntes e abrange o período de janeiro de 2006 a dezembro de 2019. Empregou-se a transformação logarítmica nos dados para estimação das elasticidades, sendo que a recuperação do dado original obedeceu à desigualdade de Jensen. A projeção deve ser feita conforme conhecimento prévio das variáveis exógenas no horizonte de projeção. Assim, deve-se, conforme visto na subseção 2.1.3, perfazer o mesmo procedimento de projeção para as variáveis exógenas, para, na etapa seguinte, realizar a projeção do tributo. Para os modelos linear e fatorial, usa-se para projeção das variáveis exógenas o MLD, enquanto para o modelo estrutural se emprega também a melhor projeção advinda dos dez modelos estruturais.

Por economia, mostramos os resultados para algumas séries do ICMS agregado para o estado de São Paulo. Os gráficos 1 a 4 apresentam a elasticidade da arrecadação com relação ao IBC-Br para três séries do Confaz, obtidas por meio do MLD. A elasticidade é dinâmica, pois este é um modelo com parâmetro variável.

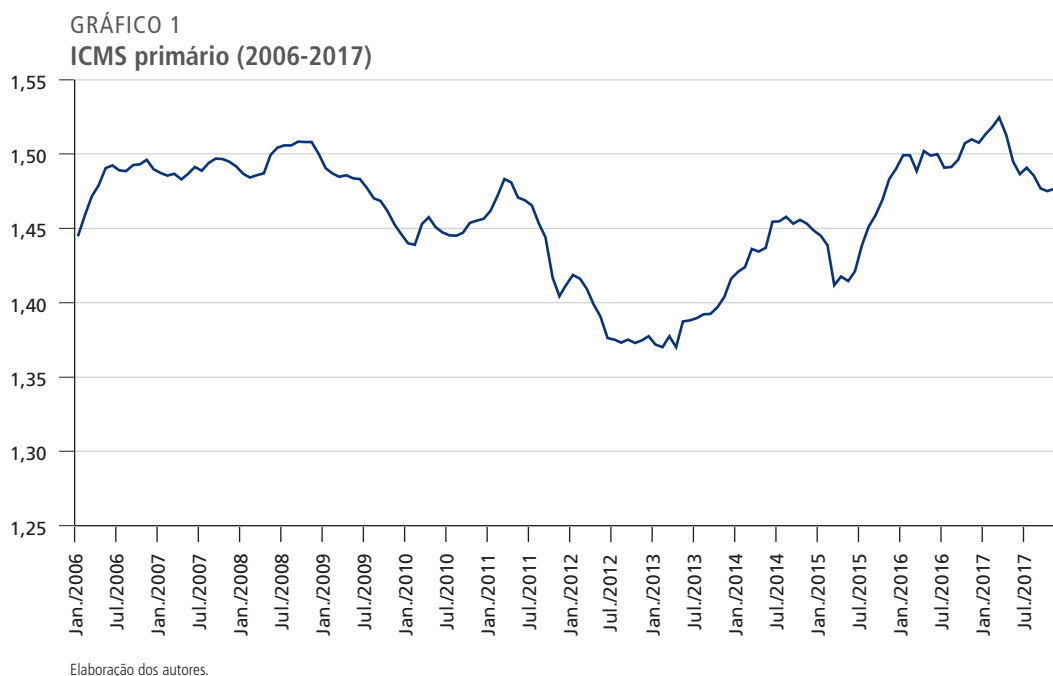
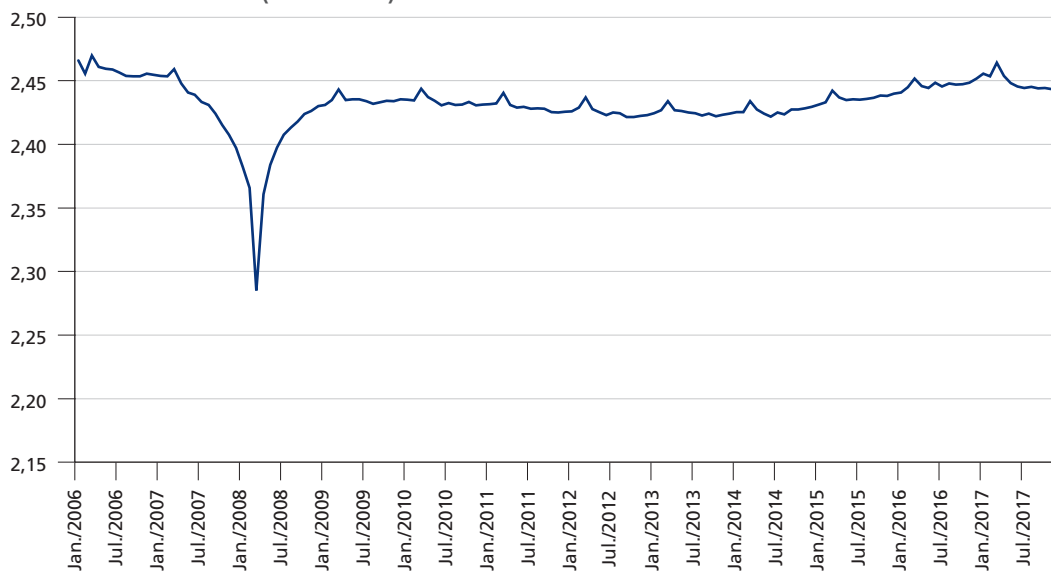
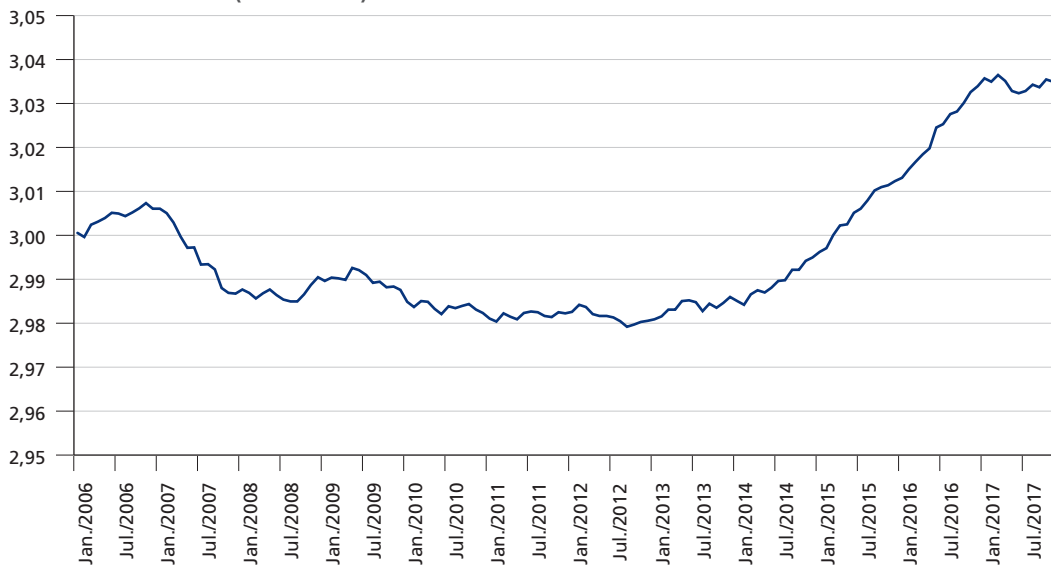


GRÁFICO 2
ICMS secundário (2006-2017)

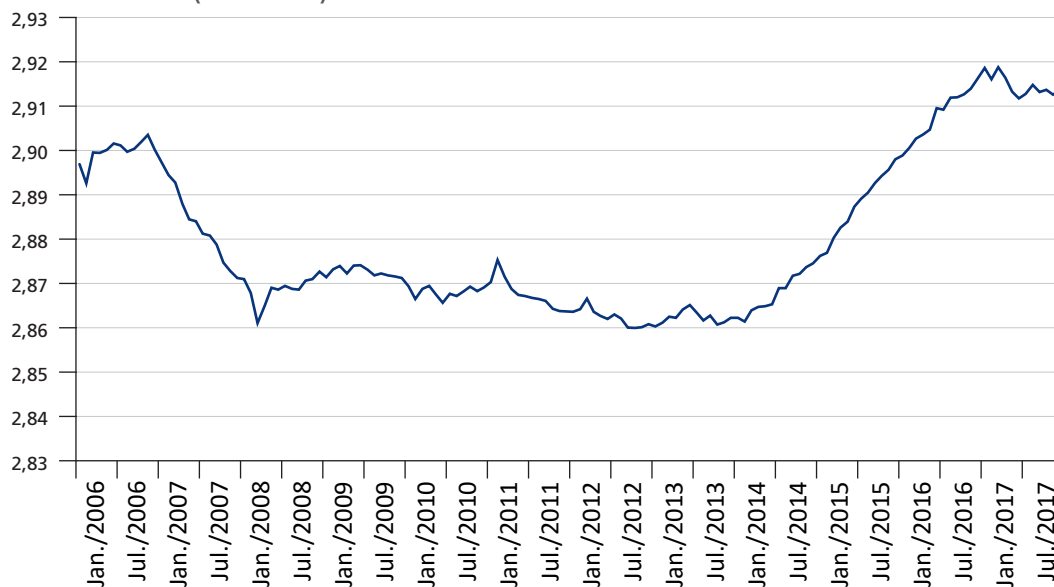


Elaboração dos autores.

GRÁFICO 3
ICMS terciário (2006-2017)



Elaboração dos autores.

GRÁFICO 4
ICMS total (2006-2017)

Elaboração dos autores.

A seguir, na figura 1, mostramos os resultados da validação em que revelamos os valores referentes ao Mape, Theil U e Eapa obtidos com as projeções feitas para todas as séries de São Paulo. Conforme pode ser visto e tomando-se por base somente o procedimento de validação, os modelos BSTS geram, de modo geral, a melhor projeção enquanto os modelos linear e fatorial originam resultados semelhantes. Cabe ressaltar que grande parte das projeções relacionadas ao modelo BSTS apresenta resultados bastante razoáveis, tendo-se como parâmetro o Mape igual ou menor que 15%. O motivo pelo qual existe diferença entre as projeções geradas pelo BSTS e os modelos fatorial e linear pode se dever ao fato de que no BSTS o efeito dos regressores é bem menos sentido na dinâmica do tributo, na medida em que nesse modelo o parâmetro do regressor é fixo. Isso faz com que as mudanças estruturais explicadas pelo regressor passem a ser indiretamente refletida no componente autônomo. De forma distinta, como nos modelos linear e fatorial, os coeficientes dos regressores variam no tempo; tem-se que as alterações na dinâmica do tributo devidas a estes sejam de fato diretamente incorporadas.

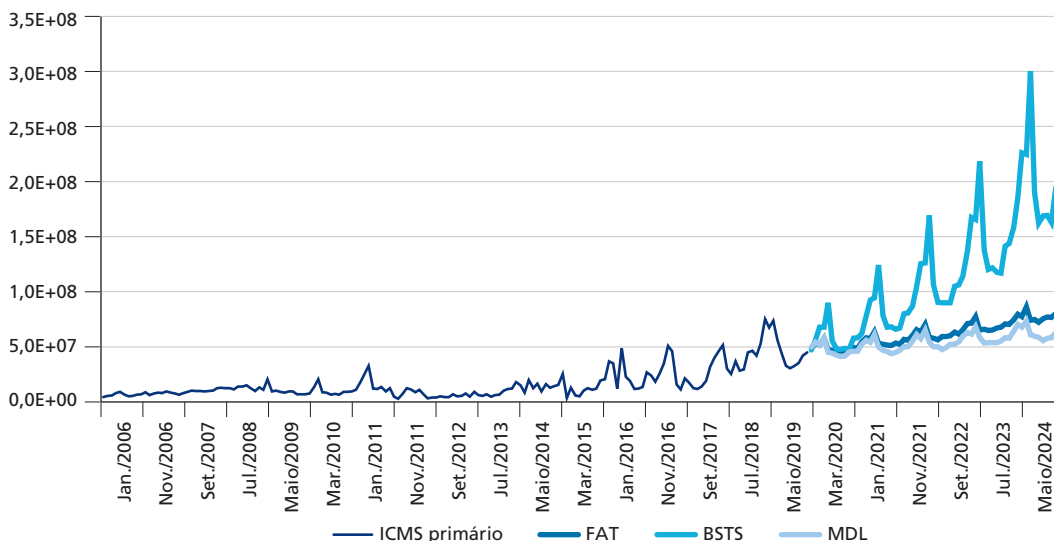
FIGURA 1
Validação da projeção (São Paulo)

ICMS primário				ICMS secundário				Comércio atacadista			
	BSTS	MLD	FAT		BSTS	MLD	FAT		BSTS	MLD	FAT
Mape	41,915	49,374	69,528	Mape	4,830	11,715	7,027	Mape	4,304	12,898	5,394
EAP	44,993	51,963	71,297	EAP	4,697	11,502	7,180	EAP	4,282	12,621	5,426
RMSE	2,254E+07	2,513E+07	3,316E+07	RMSE	2,288E+08	6,144E+08	3,261E+08	RMSE	1,518E+08	3,336E+08	1,672E+08
THEIL U	1,895	0,379	0,573	THEIL U	0,755	0,078	0,042	THEIL U	0,927	0,065	0,035
Comércio varejista				Serviço de transporte				Serviço de comunicação			
	BSTS	MLD	FAT		BSTS	MLD	FAT		BSTS	MLD	FAT
Mape	7,395	8,578	12,816	Mape	6,724	7,858	10,907	Mape	8,005	20,645	21,337
EAP	8,056	8,984	13,571	EAP	6,889	7,690	11,353	EAP	8,009	20,528	21,025
RMSE	1,435E+08	1,282E+08	1,870E+08	RMSE	2,709E+07	3,255E+07	4,285E+07	RMSE	7,514E+07	1,816E+08	1,816E+08
THEIL U	0,919	0,057	0,088	THEIL U	0,859	0,055	0,076	THEIL U	1,757	0,098	0,098
Outros ICMS				Energia elétrica secundária				Energia elétrica terciária			
	BSTS	MLD	FAT		BSTS	MLD	FAT		BSTS	MLD	FAT
Mape	6,929	12,453	13,898	Mape	48,643	57,716	134,081	Mape	5,983	13,300	14,376
EAP	6,835	12,012	13,325	EAP	62,634	70,036	117,083	EAP	5,842	12,761	13,525
RMSE	1,803E+07	3,159E+07	3,368E+07	RMSE	3,861E+06	4,156E+06	6,614E+06	RMSE	6,361E+07	1,517E+08	1,534E+08
THEIL U	1,081	0,069	0,073	THEIL U	1,219	0,606	0,701	THEIL U	0,718	0,079	0,077
Energia elétrica total				Petróleo combustível secundário				Petróleo combustível terciário			
	BSTS	MLD	FAT		BSTS	MLD	FAT		BSTS	MLD	FAT
Mape	7,280	13,469	14,101	Mape	10,350	25,856	14,900	Mape	17,501	17,827	21,256
EAP	7,298	12,929	13,277	EAP	9,313	24,810	14,713	EAP	19,648	19,915	23,888
RMSE	8,195E+07	1,541E+08	1,508E+08	RMSE	1,350E+08	3,035E+08	1,963E+08	RMSE	8,303E+07	7,963E+07	9,196E+07
THEIL U	0,883	0,080	0,076	THEIL U	0,962	0,126	0,091	THEIL U	0,956	0,141	0,168
Petróleo combustível total				Dívida ativa				Outras fontes			
	BSTS	MLD	FAT		BSTS	MLD	FAT		BSTS	MLD	FAT
Mape	6,323	13,998	13,223	Mape	21,091	26,362	23,094	Mape	8,151	15,361	10,420
EAP	6,146	13,416	13,337	EAP	22,159	26,957	20,580	EAP	15,097	20,434	17,067
RMSE	1,108E+08	2,219E+08	2,110E+08	RMSE	1,036E+07	1,189E+07	1,059E+07	RMSE	4,061E+08	3,699E+08	4,056E+08
THEIL U	0,593	0,076	0,078	THEIL U	1,031	0,173	0,135	THEIL U	1,062	0,224	0,260
ICMS total				IPVA				ITCD			
	BSTS	MLD	FAT		BSTS	MLD	FAT		BSTS	MLD	FAT
Mape	3,361	9,379	5,035	Mape	31,387	12,216	42,155	Mape	31,704	27,307	29,322
EAP	3,569	9,397	5,240	EAP	62,721	12,352	28,178	EAP	26,425	23,831	26,308
RMSE	6,925E+08	1,386E+09	9,099E+08	RMSE	1,840E+09	2,890E+08	5,410E+08	RMSE	7,718E+07	6,936E+07	7,416E+07
THEIL U	0,848	0,056	0,038	THEIL U	1,000	0,064	0,137	THEIL U	0,506	0,128	0,147
Taxas				Outros impostos							
	BSTS	MLD	FAT		BSTS	MLD	FAT				
Mape	8,267	9,667	7,418	Mape	399,416	1085,177	555,220				
EAP	7,855	9,108	7,700	EAP	85,155	77,455	86,900				
RMSE	5,582E+07	5,976E+07	5,259E+07	RMSE	1,480E+08	1,253E+08	1,427E+08				
THEIL U	0,696	0,053	0,049	THEIL U	10,204	0,476	0,660				

Elaboração dos autores.

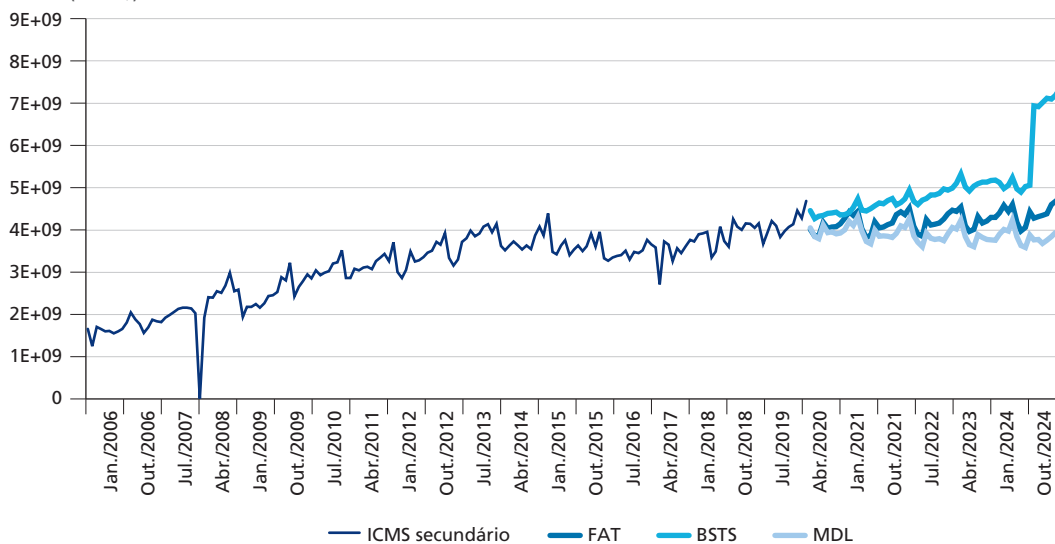
Por fim, apresentam-se os gráficos também para alguns tributos das projeções geradas pelos três modelos. Importante comentar o fato de que um modelo pode apresentar valor alto para o Mape, o que não significa que este esteja projetando mal. Isso pode acontecer devido a dois fatores. O primeiro é que a série pode ser fortemente errática, fazendo com que seja difícil mesmo para um modelo com coeficientes dinâmicos acompanhá-la; entretanto, é possível que na média o modelo esteja acertando. O segundo fator ocorre em situações em que o dado é poluído por choques; isso contribui para que o comportamento da série se altere bruscamente. No entanto, passado esse período, a série retoma a tendência que tinha anteriormente. Tais períodos de turbulência nas séries de tempo corroem os resultados da validação, muito embora a série possa estar sendo bem projetada, tendo-se em vista que os modelos dinâmicos se ajustam rapidamente às mudanças. Isso é bem ilustrado analisando o gráfico referente da série *Outros tributos*; último gráfico da seção 4. Conforme pode ser visto, as três projeções são compatíveis com os valores finais da série, embora isso tenha sido melhor captado pelos modelos BSTS e linear dinâmico.

GRÁFICO 5
Projeção de arrecadação¹ (2020-2024)
 (Em R\$)



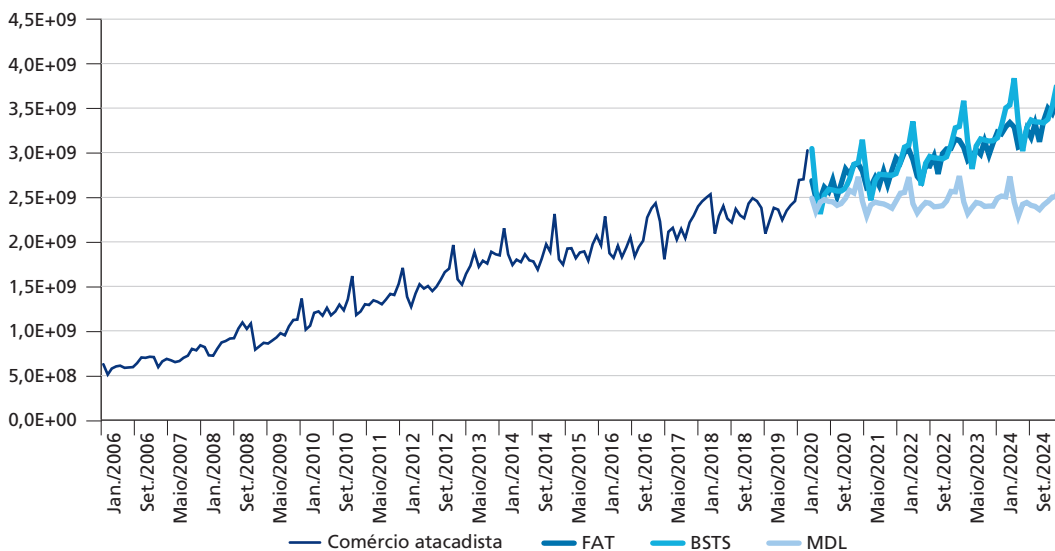
Elaboração dos autores.
 Nota ¹ Horizonte de projeção: janeiro de 2020 a dezembro de 2024.

GRÁFICO 6
Projeção de arrecadação (2020-2024)
(Em R\$)



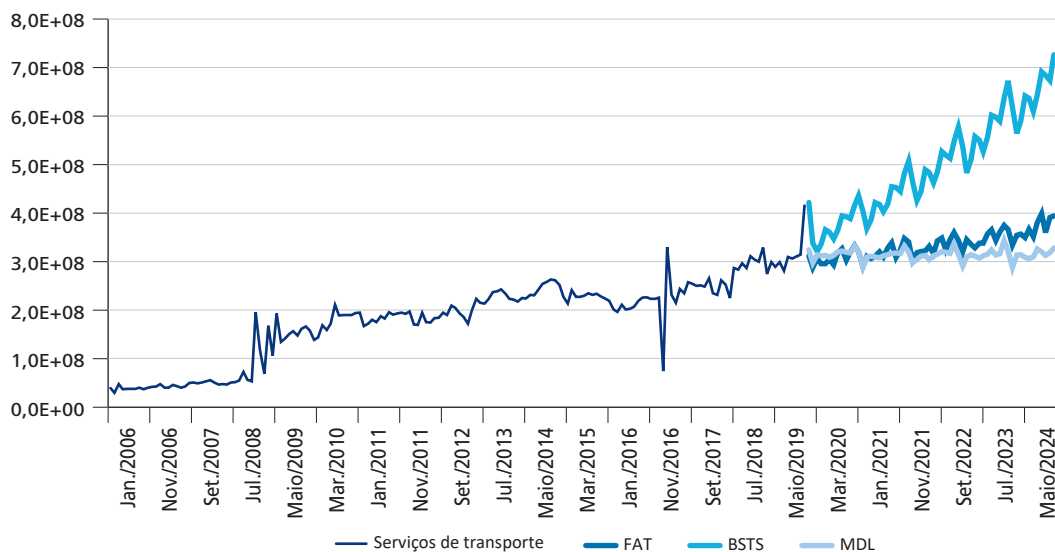
Elaboração dos autores.

GRÁFICO 7
Projeção de arrecadação (2020-2024)
(Em R\$)



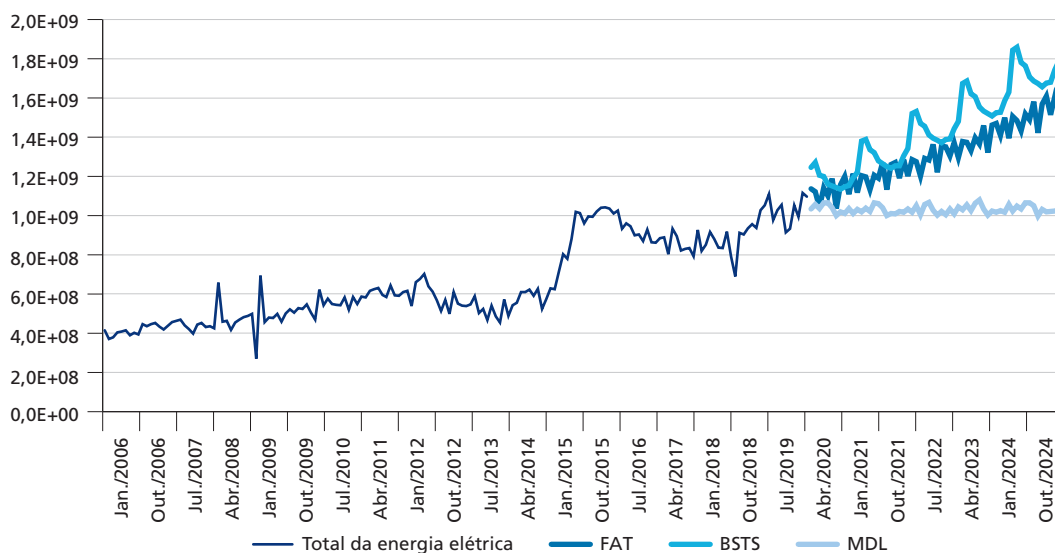
Elaboração dos autores.

GRÁFICO 8
Projeção de arrecadação (2020-2024)
 (Em R\$)



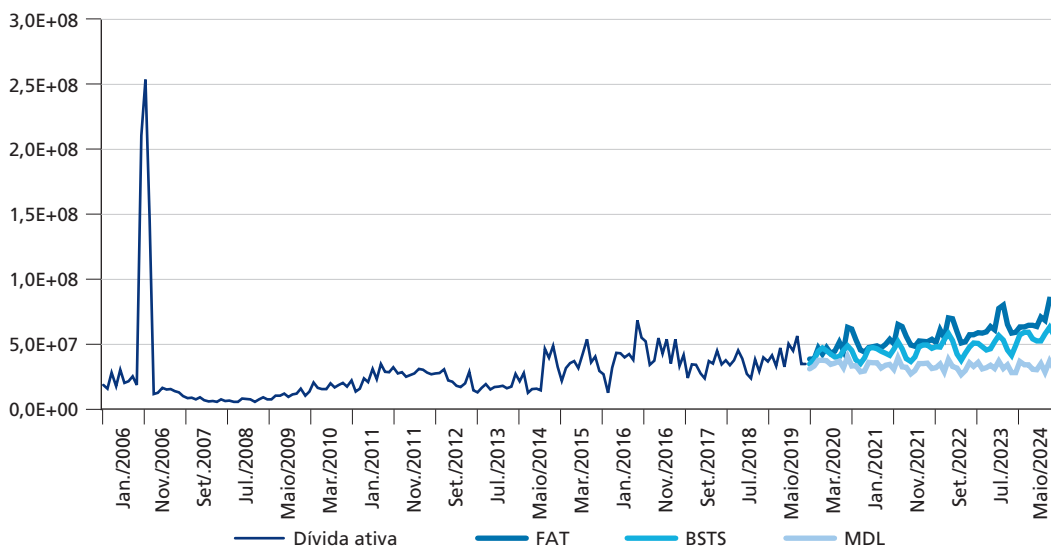
Elaboração dos autores.

GRÁFICO 9
Projeção de arrecadação (2020-2024)
 (Em R\$)



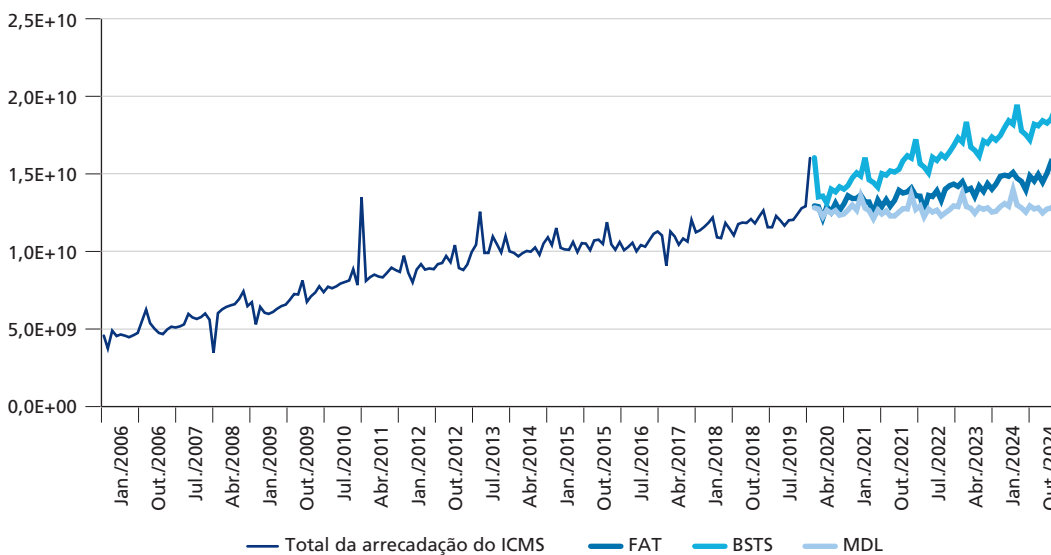
Elaboração dos autores.

GRÁFICO 10
Projeção de arrecadação (2020-2024)
(Em R\$)



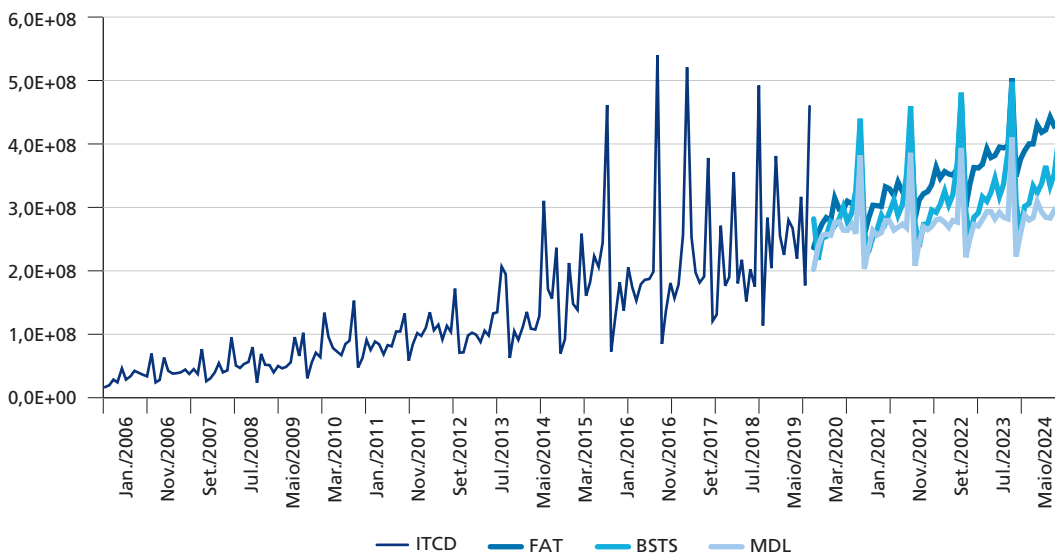
Elaboração dos autores.

GRÁFICO 11
Projeção de arrecadação (2020-2024)
(Em R\$)



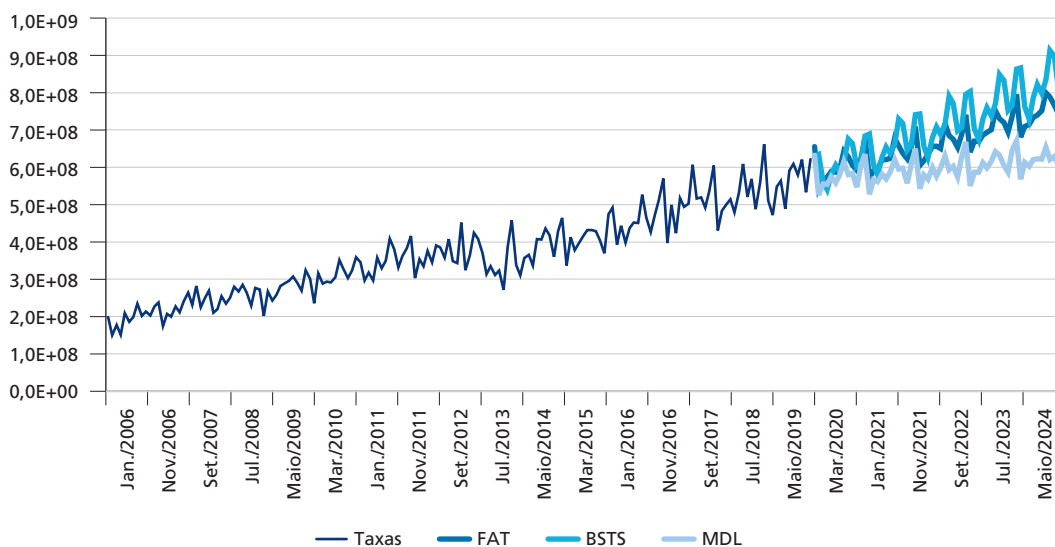
Elaboração dos autores.

GRÁFICO 12
Projeção de arrecadação (2020-2024)
 (Em R\$)



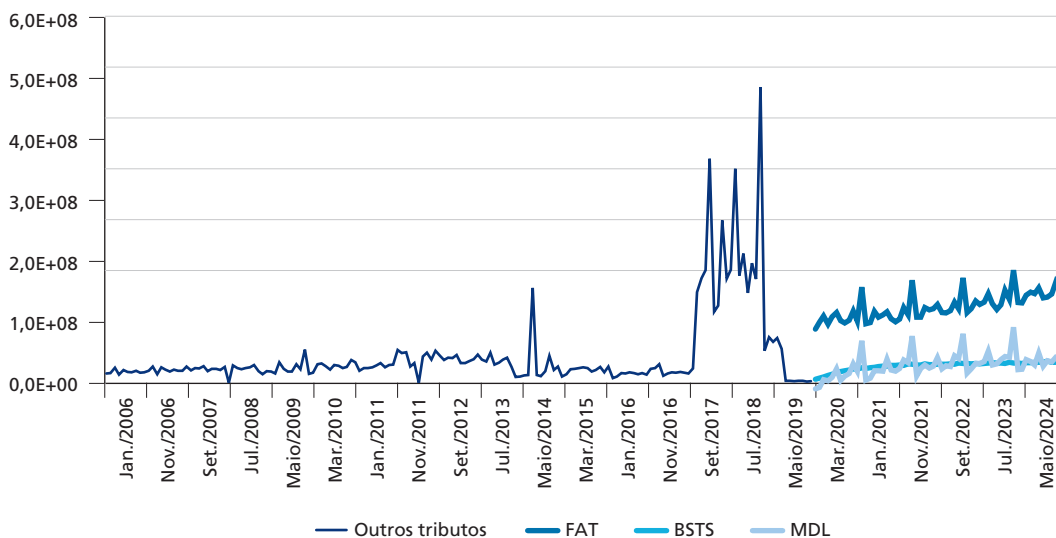
Elaboração dos autores.

GRÁFICO 13
Projeção de arrecadação (2020-2024)
 (Em R\$)



Elaboração dos autores.

GRÁFICO 14
Projeção de arrecadação BSTS com e sem variáveis exógenas (2020-2024)
(Em R\$)



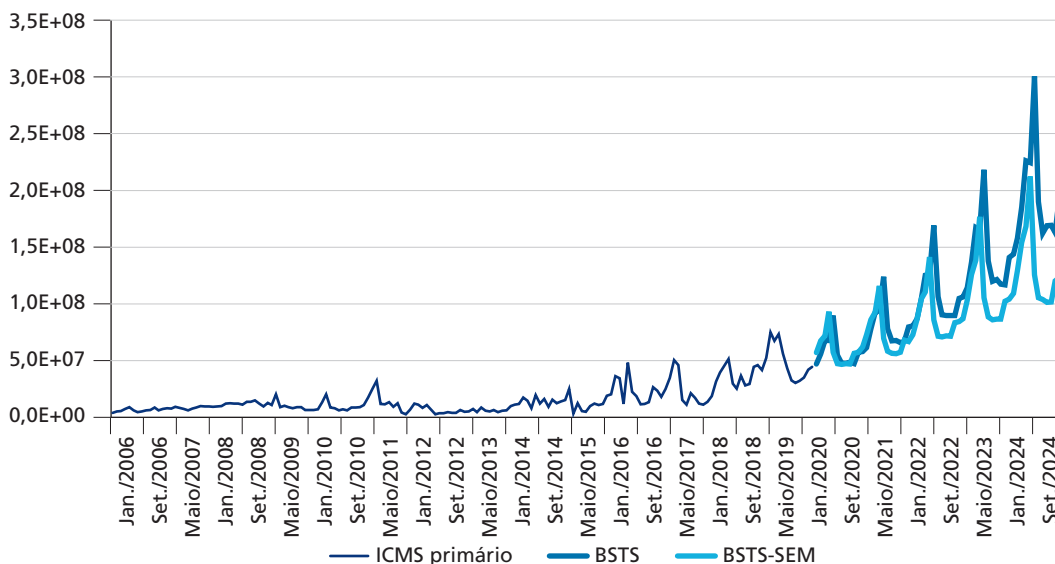
Elaboração dos autores.

6 CONSIDERAÇÕES FINAIS

Diferente das séries de tributos fortemente agregadas, as séries do Confaz apresentam diversos níveis de desagregação, tornando difícil encontrar variáveis que sirvam como fato gerador. Nesse caso, como foi visto, usamos o índice de atividade regional como *proxy* para o PIB estadual, na hipótese deste último ser o fato gerador do tributo. Considerando-se a existência de diversas fontes de incerteza quanto à identificação dos fatos geradores para as séries do conselho, é possível que, para algumas dessas séries, os regressores podem não estar contribuindo para a melhoria no desempenho preditivo do modelo. Nos gráficos e nas tabelas a seguir, comparam-se as projeções com e sem regressores geradas pelo modelo BSTS para as séries do ICMS primário, do IPVA e do ITCD.¹²

12. Esse procedimento foi feito para as demais séries do Confaz/SP, indicando que não houve melhoria no desempenho preditivo.

GRÁFICO 15
Projeção de arrecadação BSTS com e sem variáveis exógenas (2020-2024)
 (Em R\$)



Elaboração dos autores.

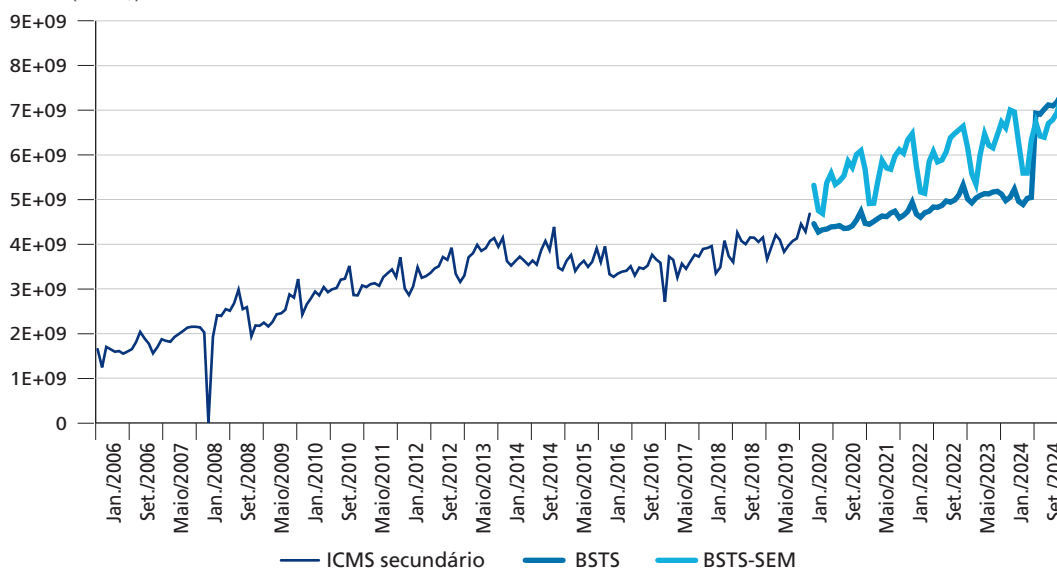
TABELA 1
Avaliação da capacidade preditiva

COM		SEM
		ModeloY9
MAPE	41.91493	15.1699
EAP	44.99324	17.24771
RMSE	22542138	10348769
THEIL U	1.894825	0.795337

Elaboração dos autores.

Obs.: Figura cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO 16
Projeção de arrecadação BSTS com e sem variáveis exógenas (2020-2024)
(Em R\$)



Elaboração dos autores.

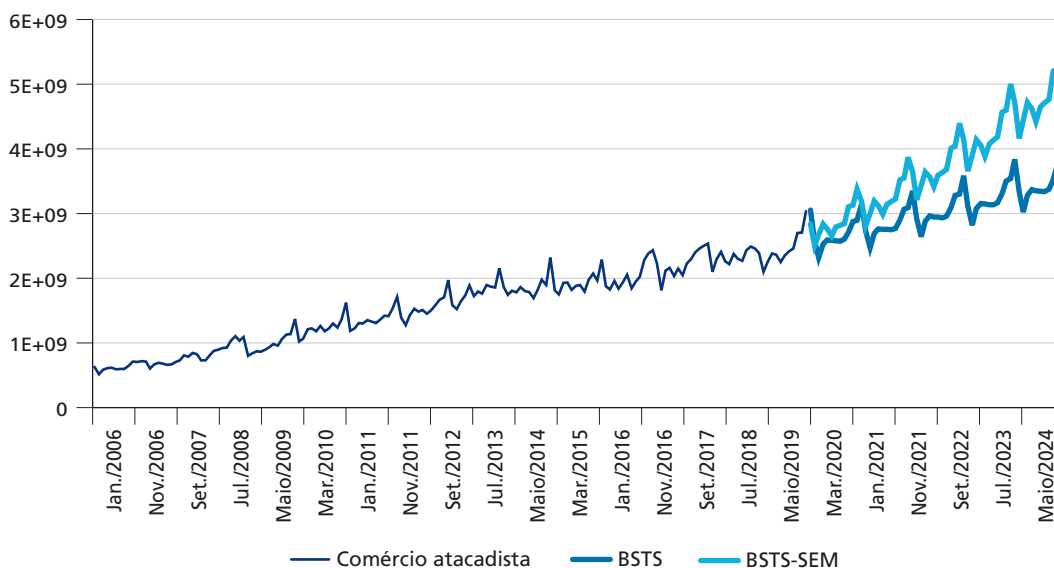
TABELA 2
Avaliação da previsão

COM		SEM
MAPE	31.38657	6.2253
EAP	62.72103	4.36385
RMSE	1.84E+09	90153296
THEIL U	1.00049	0.023797

Elaboração dos autores.

Obs.: Figura cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO 17
Projeção de arrecadação BSTS com e sem variáveis exógenas (2020-2024)
 (Em R\$)



Elaboração dos autores.

TABELA 3
Avaliação da previsão

COM		SEM
MAPE	31.70381	21.85002
EAP	26.42534	24.63522
RMSE	77182521	79192419
THEIL U	0.505601	0.600959

Elaboração dos autores.

Obs.: Figura cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Note-se que, nos três casos, o Mape se reduz fortemente para o modelo sem variáveis exógenas. Portanto, isso se deve possivelmente ao fato de os regressores não estarem contribuindo com informação relevante. Realmente, como os tributos do Confaz são fortemente desagregados, é difícil vender a ideia que se possa encontrar variáveis que os determinem com a dimensão temporal necessária. Na verdade, estes estão contribuindo, mas piorando o ajuste em vez de melhorar – os regressores apontam para sentido oposto do movimento da série. Além do que já foi dito, isso pode ocorrer, pois: *i*) existe algum outro regressor que não está no modelo cujo efeito deveria dominar os demais corrigindo

o ajuste; ou *ii*) a série não é longa o suficiente para se “calibrar” os efeitos dos regressores. Caso os regressores não estivessem contribuindo para o modelo, o efeito destes deveria ser nulo, e o Mape dos modelos com e sem regressores deveria ser muito semelhante.

REFERÊNCIAS

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

BATES, J. M.; GRANGER, C. W. J. The combination of forecasts. **Operations Research Quaterly**, v. 20, p. 319-325, 1969.

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

DURBIN, J.; KOOMAN, S. J. **Time series analysis by state space methods**. New York: Oxford University Press, 2001.

GAMERMAN, D.; LOPES, H. **Markov chain Monte Carlo**. London: Chapman & Hall, 1997.

GELMAN, A. *et al.* **Bayesian data analysis**. London: Chapman & Hall; CRC Press, 2003.

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

LIMA, A. F. R.; POLLI, D. A.; GADELHA, S. R. B. Uso da metodologia de combinação de previsões para arrecadação de receitas brutas primárias de tributos federais. **Cadernos de Finanças Públicas**, v. 1, n. 1, 2019.

MENDONÇA, M. J.; MEDRANO, L. A. T. **Aplicação do modelo fatorial dinâmico para previsão da receita tributária no Brasil**. Brasília: Ipea, 2015. (Texto para Discussão, n. 2064).

MENDONÇA, M. J.; MEDRANO, L. A.; 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**. Brasília: Ipea, 2009. (Texto para Discussão, n. 1453).

SCOTT, S. L.; VARIAN, H. R. Predicting the present with Bayesian structural time series. **International Journal of Mathematical Modelling and Numerical Optimisation**, v. 5, p. 4-23, 2014.

BIBLIOGRAFIA COMPLEMENTAR

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.

FRÜHWIRTH-SCHNATTER, S. Data augmentation and dynamic linear models. **Journal of Time Series Analysis**, v. 15, n. 2, p. 183-202, 1994.

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

APÊNDICE A

BASE DE DADOS

Total geral da receita tributária: 1 + 2

1. Total de arrecadação do Imposto sobre Operações Relativas à Circulação de Mercadorias e Prestação de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS)
 - 1.1 ICMS – setor primário (119 – ICMS primário)
 - 1.2 ICMS – setor secundário (129 – ICMS secundário)
 - 1.3 ICMS – setor terciário total
 - 1.3.1 ICMS – setor terciário – comércio atacadista (131 -Comércio Atacadista)
 - 1.3.2 ICMS – setor terciário – comércio varejista (132 - comércio varejista)
 - 1.3.3 ICMS – setor terciário – serviço de transporte (133 - serviço de transporte)
 - 1.3.4 ICMS – setor terciário – serviço de comunicação (134 - serviço de comunicação)
 - 1.3.5 ICMS – setor terciário – outros (135 - outros ICMS)
 - 1.4 ICMS – energia elétrica total
 - 1.4.1 ICMS – energia elétrica – setor secundário (211 - energia elétrica secundária)
 - 1.4.2 ICMS – energia elétrica – setor terciário (212 - elétrica terciária)
 - 1.5 ICMS – petróleo, combustíveis e lubrificantes – total
 - 1.5.1 ICMS – petróleo, combustíveis e lubrificantes – setor secundário (221 - petróleo-combustível-lubrificantes)
 - 1.5.2 ICMS – petróleo, combustíveis e lubrificantes – setor terciário (222 - petróleo-combustível-lubrificantes)
 - 1.6 ICMS – dívida ativa (239 - dívida ativa)
 - 1.7 ICMS – outras fontes de receitas (249 - outras fontes de receitas)

2. Total de arrecadação de outros tributos

2.1 Total de arrecadação do Imposto sobre a Propriedade de Veículos Automotores (IPVA) (409 - IPVA)

2.2 Total de arrecadação do Imposto sobre Transmissão Causa Mortis e Doação de Quaisquer Bens ou Direitos (ITCD) (419 - ITCD)

2.3 Total de arrecadação de taxas (429 - taxas)

2.4 Total de arrecadação de outros (439 - outros tributos)

Ipea – Instituto de Pesquisa Econômica Aplicada

Assessoria de Imprensa e Comunicação

EDITORIAL

EDITORIAL

Coordenação

Reginaldo da Silva Domingos

Assistente de Coordenação

Rafael Augusto Ferreira Cardoso

Supervisão

Camilla de Miranda Mariath Gomes

Everson da Silva Moura

Editoração

Aeromilson Trajano de Mesquita

Cristiano Ferreira de Araújo

Danilo Leite de Macedo Tavares

Herllyson da Silva Souza

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Capa

Danielle de Oliveira Ayres

Flaviane Dias de Sant'ana

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

Tel.: (61) 2026-5336

Correio eletrônico: livraria@ipea.gov.br

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.

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DA
ECONOMIA



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

ISSN 1415-4765

