

# UM MODELO DE COMBINAÇÃO DE PREVISÕES PARA ARRECAÇÃO DE RECEITA TRIBUTÁRIA NO BRASIL

## A FORECASTING COMBINATION MODEL FOR TAX REVENUE COLLECTION IN BRAZIL

### **Mário Jorge Mendonça**

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

### **Luis Alberto Medrano**

Assistente de Pesquisa da Dimac do Ipea  
Professor da Universidade Federal Rural do Rio de Janeiro (UFRRJ).  
lmedrano10@msn.com

### **RESUMO**

Este trabalho teve como objetivo a aplicação do modelo de combinação de previsões para prever a arrecadação tributária no Brasil. Combinamos aqui as previsões obtidas a partir de três modelos: modelo fatorial dinâmico (MDF), modelo SARIMA e modelo de suavização de Holt-Winters sendo que os critérios de combinação adotados foram o método de combinação ótima, performance, regressão simples, média simples e mediana. Trabalhamos com dados mensais para um total de nove tributos federais para o período de janeiro de 2001 a dezembro de 2013. A previsão para o ano de 2014 é feita fora da amostra. De acordo os resultados, pode-se observar que as previsões combinadas se mostraram de modo geral superiores às derivadas dos modelos genuínos com exceção de poucos impostos. Importante mencionar que embora o desempenho dos modelos de previsões combinadas tenha sido grosso modo superior ao dos modelos originais, não se verificou que algum deles tenha apresentado performance superior aos demais.

**Palavras-chave:** carga tributária; modelo fatorial dinâmico, SARIMA, combinação de previsões.

### **ABSTRACT**

This study aimed to the application of forecasts combination of model to predict tax revenues in Brazil. Here we combine the predictions obtained from three models: dynamic factor model (MDF), SARIMA model and model of Holt-Winters smoothing. We adopted five criteria for combine predictions: optimal combination, performance, simple regression, simple average and median. We work with monthly data for a total of nine federal taxes for the period from January 2001 to December 2013. The out of sample forecast are done for the year 2014. Considering the results, it can be seen that the combined predictions proved generally superior to those derived from genuine models with the exception of a few taxes. Notwithstanding there is no specific method of combination that provide better forecasts concern the others.

**Keywords:** tax burden; dynamic fatorial model, Bayesian methods, SARIMA, forecast combination.

## 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, via de regra,<sup>1</sup> 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,<sup>2</sup> 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 pré-existentes – e, vez por outra, também pela

---

<sup>1</sup>. 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.

<sup>2</sup>. Rezende, Oliveira e Araujo (2008, cap.2) faz um excelente resumo das inúmeras mudanças na legislação tributária brasileira no período 1998-2006.

eliminação, 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.<sup>3</sup>

Os estudos implementados anteriormente seguiram 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<sup>4</sup> que existem na economia brasileira e formam uma amostra fortemente representativa da CTBB no período de 1995 a 2010 (Mendonça, Medrano e Sachsida, 2013). 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 e Medrano (2014) estimaram o modelo fatorial dinâmico (MFD), utilizando dados mensais para o período de 1995 a 2013.

Um bom modelo de previsão de arrecadação de tributos é de importância crítica para o governo. Existe uma série de preditores para a variável de interesse. Em vez de se concentrar na seleção do melhor modelo de previsão, Bates e Granger (1969) sugeriram combinar diferentes previsões. Para justificar tal procedimento usam o fato de que combinar previsões a partir de diferentes modelos pode ser visto como uma forma de fazer a previsão mais robusta contra distintos erros de especificação. A proposta do presente estudo tem como objetivo a previsão de arrecadação de tributos com base no modelo de combinação de previsão. Combinamos aqui as previsões obtidas a partir de três modelos: modelo fatorial dinâmico (MDF), modelo SARIMA e modelo de suavização de Holt-Winters. Os critérios de combinação adotados foram os seguintes: combinação ótima, performance, regressão simples, média simples e mediana.

---

<sup>3</sup>. E mesmo antes disso, como bem apontam Rezende, Oliveira e Araujo (2008).

<sup>4</sup>. 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).

De acordo com os resultados obtidos, verificou-se que as previsões combinadas se mostraram de modo geral superiores às derivadas dos modelos genuínos com exceção de poucos impostos. Importante mencionar que embora o desempenho dos modelos de previsões combinadas tenha sido grosso modo superior ao dos modelos originais, não se verificou que algum deles tenha apresentado performance superior aos demais.

Este estudo está estruturado da seguinte forma. A seção 2 apresenta uma descrição acerca da composição da amostra de tributos que usaremos neste estudo. Em seguida, a seção 3 descreve de modo geral a teoria da combinação de previsões assim como os critérios que usamos neste trabalho para levar a cabo tal procedimento. Na seção 4 fazemos também uma breve descrição dos modelos adotados para previsão. Os resultados e a análise dos resultados são mostrados na seção 5. Por fim, a seção 6 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.<sup>5</sup> 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 (IIMP);
- Imposto sobre Operações de Crédito, Câmbio e Seguros (IOF); e

---

<sup>5</sup>. Para uma descrição sucinta da estrutura tributária no Brasil, ver Santos, Ribeiro e Gobetti (2008).

- Contribuição de Intervenção do Domínio Econômico (Cide).<sup>6</sup>

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,<sup>7</sup> 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”).

Por fim, o último e quantitativamente menos importante componente da CTBB é o item outros impostos ligados à produção, composto basicamente por taxas diversas

---

<sup>6</sup> 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.

<sup>7</sup> Ver o anexo 8 de IBGE (2004) para uma lista que exclui as contribuições para o PIS-Pasep, que foram reclassificadas como contribuições previdenciárias no Sistema de Contas Nacionais (SCN) 2005.

(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<sup>8</sup>).

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 Secretária 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**

<b>Imposto</b>	<b>%</b>	<b>Imposto</b>	<b>%</b>
ICMS	26,52%	Imposto de Importação	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 - Rend. Capital	3,66%	IPI - Fumo	0,64%
PIS/Pasep	3,58%	IPI - Automóveis	0,60%
IPI - Outros	2,36%	IPI - Bebidas	0,56%

Fonte: Elaborada pelos autores.

### 3. MODELO DE COMBINAÇÃO DE PREVISÕES

Muitas vezes, há uma série de preditores para uma variável de interesse. Em vez de se concentrar na seleção do melhor modelo de previsão, Bates e Granger (1969)

<sup>8</sup> Sistema S é o nome pelo qual ficou convencionado de se chamar ao conjunto de nove instituições de interesse de categorias profissionais, estabelecidas pela Constituição Brasileira. Entre as mais conhecidas podemos citar o Senai, Senac, Sesi, etc.

sugeriram combinar diferentes previsões. Entre os principais argumentos para o uso de combinação de previsões podemos temer que, primeiro o verdadeiro processo de geração de dados é desconhecido. Mesmo o modelo mais complicado é provável que esteja mal especificado, sendo na melhor das hipóteses uma razoável aproximação "local" do verdadeiro modelo. Assim é altamente improvável que um único modelo vai dominar uniformemente ao longo do tempo. Em segundo lugar, mesmo o melhor modelo pode mudar ao longo do tempo de uma forma que pode ser difícil modelar com base no passado pode conduzir a erros de especificação. Por fim, é possível que diversificação traga ganhos na medida que a combinação em um conjunto de modelos de previsão pode dominar a estratégia de usar apenas um único modelo de previsão. Desde então, numerosos métodos de combinação foram propostos (Timmermann, 2006).

Neste estudo usamos cinco métodos para combinar previsões: combinação ótima (C1), regressão simples (C2), desempenho ou performance (C3), média simples (C4) e mediana (C5). A seguir faremos uma breve descrição dos três métodos primeiros, pois os demais carecem de explicação por serem demais triviais. Sejam  $F_{1t}$  e  $F_{2t}$  duas previsões para a mesma variável de interesse  $Y_t$  uma combinação das previsões é representada por

$$F_{ct} = (1 - \lambda)F_{1t} + \lambda F_{2t} \text{ para } 0 \leq \lambda \leq 1 \quad (1).$$

assumindo que os erros das previsões individuais são não-viesados, o erro da previsão combinada

$$\varepsilon_{ct} = (1 - \lambda)\varepsilon_{1t} + \lambda\varepsilon_{2t}$$

onde  $\varepsilon_{ct} = Y_{ct} - F_{ct}$  é também é não-viesado e possui variância

$$Var(\varepsilon_{ct}) = E(\varepsilon_{ct}^2) = (1 - \lambda)^2 \sigma_1^2 + \lambda^2 \sigma_2^2 + 2\lambda(1 - \lambda)\rho\sigma_1\sigma_2$$

O valor de  $\lambda$  que minimiza a variância de  $\varepsilon_{ct}$  é dado pela seguinte expressão igual a

$$\hat{\lambda} = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (2)$$

Na prática, os parâmetros  $\sigma_1^2$ ,  $\sigma_2^2$  e  $\rho$ , portanto temos que usar seus usuais de modo que o valor ótimo de  $\lambda$  pode ser obtido a partir da seguinte expressão

$$\hat{\lambda} = \frac{\sum_{t=1}^T e_{1t}^2 - \sum_{t=1}^T e_{1t}^2 e_{2t}^2}{\sum_{t=1}^T e_{1t}^2 + \sum_{t=1}^T e_{2t}^2 - 2 \sum_{t=1}^T e_{1t}^2 e_{2t}^2} \quad (3)$$

No caso onde se tem  $N$  previsões  $\hat{\lambda}$  é definido da seguinte forma

$$\hat{\lambda} = \frac{\hat{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}^{-1} \mathbf{1}} \quad (4)$$

$\hat{\Sigma}$  é a matrix de covariância amostral dos erros e  $\mathbf{1}$  é o vetor coluna  $(1,1,\dots,1)'$ . Uma vez obtido o valor ótimo de  $\lambda$  podemos fazer a substituição no correlato amostral da equação (1)<sup>9</sup> e obter a combinação ótima, denotada aqui por C1. Conforme apontado por Newbold e Harvey (2004), a equação (3) pode ser interpretada como o estimador de mínimos quadrados de  $\lambda$  com certas restrições nos coeficientes. Portanto, um segundo procedimento de combinação que se denomina, por motivo explícito, regressão linear (C2) e sugerido por Granger e Ramanathan (1984), é generalizar obter a combinação a partir da estimação de uma regressão linear tal como expressa na equação (5) levando em consideração as seguintes restrições  $\beta_0 = 0$  de que  $\beta_0 = 0$  e  $\beta_1 + \beta_2 = 1$ .

$$F_{ct} = \beta_0 + \beta_1 F_{1t} + \beta_2 F_{2t} + \varepsilon_t \quad (5)$$

Um terceiro procedimento de combinação que aqui denominamos Performance (C3) é obtido ao se ignorar a covariância amostral em (3) e atribuir pesos inversamente proporcionais ao erro quadrado médio. Tal procedimento sugerido por Bates e Granger (1969) é levado a cabo quando temos que  $\hat{\lambda}$  é definido pela seguinte expressão

$$\hat{\lambda} = \frac{\sum_{t=1}^T e_{1t}^2}{\sum_{t=1}^T e_{1t}^2 + \sum_{t=1}^T e_{2t}^2} \quad (6)$$

A equação (6) pode ser generalizada por o caso de  $N$  previsões de modo que neste caso temos que

---

<sup>9</sup> Aqui temos os seguintes correlatos amostrais  $\hat{F}_{ct} = (1 - \lambda)\hat{F}_{1t} + \lambda\hat{F}_{2t}$  e  $e_{ct} = Y_{ct} - \hat{F}_{ct}$ .

$$\hat{\lambda}_i = \frac{\left( \sum_{t=1}^T e_{it}^2 \right)^{-1}}{\sum_{i=1}^N \left( \sum_{t=1}^T e_{it}^2 \right)^{-1}}$$

Onde  $\hat{\lambda}_i$  é o peso para a i-ésima previsão. Além dos três procedimentos expostos acima, também serão utilizados procedimentos simples de combinação, que empiricamente possuem um desempenho satisfatório (CLEMEN, 1989). Serão utilizados como forma de determinação dos pesos a média simples (C4) e a mediana (C5).

#### 4. MODELOS DE PREVISÃO

Nesta seção apresentamos uma descrição geral acerca dos modelos que usaremos para obter as previsões que serão combinadas de acordo com os critérios introduzidos na seção 3. Os modelos que usaremos para obter as previsões são os seguintes: modelo fatorial dinâmico (MDF), modelo SARIMA e modelo de suavização de Holt-Winters. O

##### 4.1. O MODELO FATORIAL

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. 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 pode ser representado da seguinte forma

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

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

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

#### 4.2.1 MODELO FATORIAL DINÂMICO (MFD)

Para estimação do modelo aplicamos aqui os procedimentos de inferência bayesiana – Kim e Nelson (1999); West e Harrison (1997); Carter e Kohn (1994) Gammerman 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, Carlin, Stern e Rubin, 2003, p. 696). Para maiores detalhes ver Mendonça e Medrano (2014).

Suponha  $m$  variáveis relacionadas que obedecem à distribuição normal multivariada denotada por  $N(0, \Omega)$  onde  $\Omega$  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 nesse estudo é dado por

$$\begin{aligned} y_t &= \beta f_t + S_t + e_t, \quad e_t \sim N(0, \Sigma) \\ f_t &= \Gamma_1 f_{t-1} + \Gamma_2 f_{t-2} + \dots + \Gamma_p f_{t-p} + w_t, \quad w_t \sim N(0, \Lambda) \end{aligned} \quad (1)-(2)$$

onde  $y_t$ ,  $f_t$ ,  $e_t$  e  $w_t$  são vetores de dimensão  $m$ ,  $k$ ,  $m$  e  $k$ , respectivamente,  $\beta$  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  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ .

Para algum inteiro positivo específico  $k \leq m$ , o modelo (1) relaciona cada  $y_t$  com um vetor  $k$ -dimensional de variáveis aleatórias  $f_t$ , os fatores comuns, que obedecem a uma estrutura auto-regressiva  $AR(p)$ , e um componente sazonal,  $S_t = (S_{1t} \ S_{2t} \ \dots \ S_{mt})^T$  (West e Harrison, 1997), definida por

$$S_{it} = \sum_{j=1}^6 \left[ a_{ij} \text{sen} \left( \frac{2\pi jt}{12} \right) + b_{ij} \cos \left( \frac{2\pi jt}{12} \right) \right] = \varphi_i X_t,$$

onde,

$$X_t = \left( \text{sen} \left( \frac{2\pi t}{12} \right) \ \text{sen} \left( \frac{2\pi 2t}{12} \right) \ \dots \ \text{sen} \left( \frac{2\pi 6t}{12} \right) \ \cos \left( \frac{2\pi t}{12} \right) \ \cos \left( \frac{2\pi 2t}{12} \right) \ \dots \ \cos \left( \frac{2\pi 6t}{12} \right) \right)^T$$

e

$$\varphi_i = (a_{i1} \ a_{i2} \ \dots \ a_{i6} \ b_{i1} \ b_{i2} \ \dots \ b_{i6}), \text{ para } i = 1, \dots, m.$$

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.<sup>10</sup> A motivação para o uso desta abordagem, além da flexibilidade, é a facilidade de se obter interpretação, já que as trajetórias em forma 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.

<sup>10</sup>. Uma função  $g(t)$  é dita periódica se para algum inteiro  $p > 1$ , e todo  $n, t > 0$   $g(t + np) = g(t)$ .

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,  $\sigma_i^2$  medem a variabilidade residual em cada uma das variáveis dos dados uma vez que a contribuição dos fatores já foi considerada.

#### 4.2.2 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  $\beta$ . Optou-se por seguir, aqui, a alternativa usada em Geweke e Zhou (1996) e Aguilar e West (2000), entre outros, ao restringir  $\beta$  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:

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

A restrição exposta anteriormente garante que  $\beta$  tenha posto completo  $k$ , resolvendo os problemas de identificação que surgem pela invariância do modelo sob mudanças de locação da matriz  $\beta$ . 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  $\Omega$  não restrito. Além de resolver os problemas de identificação, essa estrutura de  $\beta$  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.

#### 4.2.3 Especificação das prioris

Para completar a especificação do nosso modelo, deve-se atribuir classes de prioris para os parâmetros  $f_0$ ,  $\beta$ ,  $\Gamma_i$ ,  $i = 1, \dots, p$ ,  $\Sigma$ ,  $\Lambda$  e  $\varphi_i$ ,  $i = 1, \dots, m$ . As prioris utilizadas são listadas a seguir.

- Vetor de fatores comuns inicial

$$f_0 \sim N(\underline{m}_0, \mathbf{C}_0).$$

- Elementos não nulos da matriz de cargas  $\beta$

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

- Elementos da matriz diagonal de variância-covariância  $\Sigma$

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

- Elementos das matrizes diagonais dos coeficientes auto-regressivos  $\Gamma_1, \Gamma_2, \dots, \Gamma_p$

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

- Elementos da matriz diagonal de variância-covariância  $\Lambda$

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

- Elementos dos vetores de coeficientes sazonais  $\varphi_1, \varphi_2, \dots, \varphi_m$

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

#### 4.2.4 Método MCMC para o modelo fatorial com $k$ fatores

Nosso objetivo é obter as posteriores dos parâmetros do modelo, assim como fazer inferência sobre o processo fatorial  $\{f_t : t = 1, 2, \dots, T\}$ . 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 posteriores de interesse. As posteriores 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).

#### 4.2.5 Previsão do MFD com componente sazonal

Nesta seção, 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

$$p(\underline{y}_{t+h} | \mathbf{y}) = \int p(\underline{y}_{t+h} | \underline{f}_{t+h}, \boldsymbol{\beta}, \boldsymbol{\varphi}, \Theta) p(\underline{f}_{t+h} | \underline{f}_t, \boldsymbol{\beta}, \boldsymbol{\varphi}, \Theta) p(\underline{f}_t, \boldsymbol{\beta}, \boldsymbol{\varphi}, \Theta | \mathbf{y}) d\underline{f}_{t+h} d\underline{f}_t d\boldsymbol{\beta} d\boldsymbol{\varphi} d\Theta$$

onde

$$\Theta = (\boldsymbol{\sigma}, \boldsymbol{\lambda}, \boldsymbol{\gamma}), \quad \boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_m), \quad \boldsymbol{\gamma} = (\gamma_{11}, \dots, \gamma_{1k}), \quad \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k),$$

$$(\underline{y}_{t+h} | \underline{f}_{t+h}, \boldsymbol{\beta}, \boldsymbol{\varphi}, \Theta) \sim N(\boldsymbol{\beta} \underline{f}_{t+h} + \boldsymbol{\varphi} X_{t+h}, \boldsymbol{\Sigma}) \text{ e } (\underline{f}_{t+h} | \underline{f}_t, \boldsymbol{\beta}, \boldsymbol{\sigma}) \sim N(\underline{\mu}_h, \mathbf{V}_h),$$

com  $\underline{\mu}_h = \Gamma_1^h \underline{f}_t$  e  $\mathbf{V}_h = \sum_{k=1}^h \Gamma_1^{k-1} \boldsymbol{\Lambda} (\Gamma_1^{k-1})^T$ , para  $h > 0$ .

Portanto, se  $\{(\boldsymbol{\beta}^{(1)}, \underline{f}_t^{(1)}, \boldsymbol{\varphi}^{(1)}, \Theta^{(1)}), \dots, (\boldsymbol{\beta}^{(M)}, \underline{f}_t^{(M)}, \boldsymbol{\varphi}^{(M)}, \Theta^{(M)})\}$  é uma amostra de  $p(\underline{f}_t, \boldsymbol{\beta}, \boldsymbol{\varphi}, \Theta | \mathbf{y})$ , basta gerar uma amostra de tamanho  $M$  de  $(\underline{f}_{t+h} | \underline{f}_t, \boldsymbol{\beta}, \boldsymbol{\varphi}, \Theta)$ , isto é, gerar  $\{(\underline{f}_{t+h}^{(1)}), \dots, (\underline{f}_{t+h}^{(M)})\}$ , de modo que

$$\hat{p}(\underline{y}_{t+h} | \mathbf{y}) = M^{-1} \sum_{j=1}^M p(\underline{y}_{t+h} | \underline{f}_{t+h}^{(j)}, \boldsymbol{\beta}^{(j)}, \boldsymbol{\varphi}^{(j)}, \Theta^{(j)})$$

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

## 4.2. MODELO SAZONAL AUTOREGRESSIVO INTEGRADO COM MÉDIA MÓVEL (SARIMA)

O modelo SARIMA é uma variante do modelo *ARIMA* que visa tratar o problema da sazonalidade. Isso ocorre quando existe autocorrelação significativa entre defasagens sazonais, ou seja, múltiplos de um período qualquer  $s$ . Neste caso, tem-se um modelo *ARIMA* sazonal (*SARIMA*). Para expressarmos este tipo de modelo, precisamos estender a notação utilizada nos modelos *ARIMA*( $p, d, q$ ) para *SARIMA*( $p, d, q$ )( $P, D, Q$ ) $_s$ ,

onde  $(p, d, q)$  representa a parte não sazonal,  $(P, D, Q)$  a parte sazonal e  $s$  uma sazonalidade de  $s$  períodos.

### 4.3. MODELO DE SUAVIZAÇÃO EXPONENCIAL DE HOLT-WINTERS

Os métodos de suavização exponencial apresentados até o momento ainda não são adequados para previsão de séries temporais que possuem um comportamento sazonal. O modelo de suavização exponencial de Holt-Winters é uma extensão ao modelo de Holt para tratar do padrão sazonal. Novamente, essa característica é obtida por meio de uma nova constante de suavização,  $\gamma$ , utilizada para suavização da sazonalidade. As equações descritivas da suavização exponencial de Holt-Winters são

$$\begin{aligned} F_{t+h} &= (l_t + b_t h) s_{t-m+h} \\ l_t &= \alpha Y_t s_{t-m} + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta (l_t - l_{t-1}) + (1 - \beta) b_{t-1} \\ s_t &= \gamma Y_t l_t + (1 - \gamma) s_{t-m} \end{aligned}$$

onde  $F_{t+h}$  é a previsão para o período  $t+h$ ,  $l_t$  é o nível suavizado no período  $t$ ,  $b_t$  é a tendência suavizada no período  $t$ ,  $s_t$  é a componente sazonal no período  $t$  e  $m$  e apresenta o número de períodos da sazonalidade (e.g., número de meses em um ano). As equações expostas acima tratam a sazonalidade de forma multiplicativa, ou seja, a sazonalidade é proporcional ao nível atual da série. Apesar de ser a forma mais usual de sazonalidade, especialmente em séries econômicas e financeiras, também é possível modelar a sazonalidade aditiva, expressa como

$$\begin{aligned} F_{t+h} &= (l_t + b_t h + s_{t-m+h}) \\ l_t &= \alpha Y_t - s_{t-m} + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta (l_t - l_{t-1}) + (1 - \beta) b_{t-1} \\ s_t &= \gamma Y_t - l_t + (1 - \gamma) s_{t-m} \end{aligned}$$

Outra vez, os problemas se encontram na determinação dos valores ótimos para as constantes de suavização e no problema da inicialização. Para a determinação das constantes de suavização, podemos utilizar novamente a minimização da soma dos quadrados dos erros de previsão. Já para o problema da inicialização, Makridakis, Wheelwright e Hyndman (1998) recomendam o uso das seguintes fórmulas

$$l_m = \frac{1}{m} (Y_1 + Y_2 + \dots + Y_m)$$

$$b_m = \frac{1}{m} \left( \frac{Y_{m-1} - Y_1}{m} + \frac{Y_{m+2} - Y_2}{m} + \dots + \frac{Y_{m+m} - Y_m}{m} \right)$$

$$s_1 = \frac{Y_1}{l_m}, s_2 = \frac{Y_2}{l_m}, \dots, s_m = \frac{Y_m}{l_m}$$

## 5. RESULTADOS

Nesta seção apresentamos os resultados obtidos a partir do modelo de combinação de previsões. Conforme foi dito, combinamos aqui as previsões obtidas a partir de três modelos: modelo fatorial dinâmico (MDF), modelo SARIMA e modelo de suavização de Holt-Winters. Os critérios de combinação adotados foram os seguintes: combinação ótima, performance, regressão simples, média simples e mediana. De modo a avaliar a capacidade preditiva do modelo  $p$  passos à frente fora da amostra cinco critérios de avaliação que são a raiz do erro quadrado médio (RMSE),<sup>11</sup> o desvio absoluto médio (mean absolute deviation - MAD),<sup>12</sup> o erro percentual absoluto médio (mean absolute percentage error - MAPE),<sup>13</sup> o erro percentual médio (mean percentage error - MPE)<sup>14</sup> e o erro acumulado percentual. 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.<sup>15</sup> Em geral, o critério erro quadrático médio (EQM) é comumente mencionado na literatura econométrica teórica (Stock e Watson, 2010; Lutkpohl, 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

---


$$^{11}. RSME = \sqrt{N^{-1} \sum_{t=1}^N e_t^2} \text{ onde } e_t = Y_t - \hat{F}_t.$$

$$^{12}. MAD = N^{-1} \sum_{t=1}^N |e_t|.$$

$$^{13}. MAPE = \left[ \frac{1}{n} \sum_{j=1}^n \frac{|e_j|}{y_j} \right] \times 100.$$

$$^{14}. MPE = \left[ N^{-1} \sum_{t=1}^N \frac{e_t}{y_t} \right] \times 100$$

<sup>15</sup>. 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 (Lutkpohl, 2007).

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.

Nossa base de dados mensais abrange o período de janeiro de 2001 a dezembro de 2013. Os tributos analisados se restringem à esfera federal englobando os seguintes tributos, os impostos incidentes sobre a renda (IRPF, IRPJ, IRRF), Cofins, CSLL, IIMP, IPI, CPIS e o IOF. A previsão para o ano de 2014 é feita fora da amostra. No anexo 1 são apresentados os resultados do modelo. No anexo 1 são apresentados os valores relacionados para os erros de projeção não apenas das combinações de previsões assim como as previsões geradas pelos modelos originais. Também são apresentados os gráficos das previsões derivadas dos métodos de combinação ótima. A partir dos resultados do Anexo 1, podemos dizer que as previsões combinadas se mostraram de modo geral superiores às obtidas pelos modelos genuínos embora isso não tenha sido verificado para o IIMP e o IOF. No caso deste último já se verificou a dificuldade com relação à modelagem deste imposto. Deve-se ainda ressaltar que embora o desempenho dos modelos de previsões combinadas tenha sido grosso modo superior ao dos modelos originais, não se verificou que algum deles tenha apresentado performance superior aos demais.

## **6. CONCLUSÃO**

Dando sequência a uma série de estudos já desenvolvidos pela Diretoria de Macroeconomia do IPEA, este trabalho teve como objetivo a aplicação do modelo de combinação de previsões para prever a arrecadação tributária no Brasil. Combinamos aqui as previsões obtidas a partir de três modelos: modelo fatorial dinâmico (MDF), modelo SARIMA e modelo de suavização de Holt-Winters sendo que os critérios de combinação adotados foram o método de combinação ótima, performance, regressão simples, média simples e mediana. Trabalhamos com dados mensais para um total de 09 tributos federais para o período de janeiro de 2001 a dezembro de 2013. A previsão para

o ano de 2014 é feita fora da amostra. No anexo 1 são apresentados os resultados do modelo.

De acordo os resultados obtidos, pode-se observar que as previsões combinadas se mostraram de modo geral superiores às derivadas dos modelos genuínos com exceção de poucos impostos. Importante mencionar que embora o desempenho dos modelos de previsões combinadas tenha sido grosso modo superior ao dos modelos originais, não se verificou que algum deles tenha apresentado performance superior aos demais.

## 7. 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.

BATES, J. M.; GRANGER, C. W. J. The Combination of Forecasts. *Operational Research Quarterly*, v. 20, p. 451-468, 1969.

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).

CLEMEN, R. T. Combining forecasts : A review and annotated bibliography. *International Journal of Forecasting*, v. 5, p. 559-583, 1989.

DOS SANTOS, C. H. M.; 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. 1.350).

FRUWIRTH-SCHMATTER, S. Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15, 183-202, 1994.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. *Bayesian Data Analysis*, 2003.

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

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.

GRANGER, C.; RAMANATHAN, R. Improved Methods of Combining Forecasts. *Journal of Forecasting*, v. 3, p. 197–204, 1984.

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).

Instituto Brasileiro de Geografia e Estatística. Sistema de Contas Nacionais (SCN). Série Relatórios Metodológicos v. 24, 2004

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: VI PRÊMIO TESOURO NACIONAL*, 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*, 41(1), 133–162, 2011.

MENDONÇA, M. J.; MEDRANO, L. A. T; SACHSIDA, A. Um Modelo Econométrico para Previsão de Receita Tributária no Brasil. *Economia Aplicada*, 17(2), 295-329, 2013

MENDONÇA, M. J., MEDRANO, L. A. Aplicação de um modelo fatorial dinâmico para previsão da arrecadação tributária no Brasil. *Texto para discussão IPEA*, 2014.

NEWBOLD, P.; HARVEY, D. Forecast Combination and Encompassing. *In: CLEMENTS, M. P.; HENDRY, D. (Ed.). A Companion to Economic Forecasting*. 1. ed. Oxford: Blackwell Publishing, 2004. cap. 12, p. 620.

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., RIBEIRO, M. B.; GOBETTI, S. W. A evolução da carga tributária bruta brasileira no período 1995-2007: Tamanho, composição e especificações econométricas agregadas, *Texto para discussão 1350*, IPEA, 2008.

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

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

TIMMERMANN, A. (2006): .Forecast Combinations,.in Handbook of Forecasting Volume 1, ed. by G. E. et. al, pp. 135.196. Elsevier, Amsterdam.

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

# ANEXO 1

## 1. Imposto sobre Importação (IIMP)

Modelo	Erro acumulado	Raiz do erro quadrático	Mean Percentual	Mean Absolute Percentual	Mean Absolute	Pesos				
	percentual	médio (RMSE)	Error (MPE)	Error (MAPE)	Deviation (MAD)	M1	M2	M3		
Modelos Originais										
SARIMA - M1	-18.76%	R\$ 693	-19.20%	20.92%	R\$ 626	1				
Holt-Winters - M2	-11.92%	R\$ 538	-12.26%	15.75%	R\$ 477		1			
Fatorial Dinamico - M3	0.73%	R\$ 212	0.44%	5.55%	R\$ 171			1		
Combinações										
Combinação Ótima – C1	7.83%	R\$ 290	7.56%	7.64%	R\$ 241	0.026	-0.601	1.575		
Regressão Linear – C2	-4.47%	R\$ 213	-4.78%	6.22%	R\$ 184	0.054	-0.699	1.776		
Performance – C3	-2.31%	R\$ 250	-2.62%	6.87%	R\$ 209	0.075	0.124	0.8		
Média Simples – C4	-9.98%	R\$ 441	-10.34%	13.25%	R\$ 399	0.333	0.333	0.333		
Mediana – C5	-11.38%	R\$ 513	-11.73%	15.09%	R\$ 457					
	Real	M1	M2	M3	C1	C2	C3	C4	C5	
jan/14	3,424	3,504	3,086	2,939	2,865	3,251	3,000	3,176	3,086	
fev/14	3,083	2,764	2,743	2,799	2,831	3,202	2,789	2,768	2,764	
mar/14	3,057	3,523	3,242	2,852	2,634	2,988	2,951	3,206	3,242	
abr/14	2,955	3,371	3,015	2,902	2,846	3,228	2,951	3,096	3,015	
mai/14	2,953	3,433	3,221	3,019	2,908	3,295	3,075	3,224	3,221	
jun/14	2,669	3,593	3,200	2,864	2,681	3,044	2,961	3,219	3,200	
jul/14	2,992	3,840	3,498	3,035	2,778	3,153	3,154	3,458	3,498	
ago/14	2,915	4,051	3,824	3,182	2,819	3,197	3,327	3,685	3,824	
set/14	3,210	3,702	3,628	3,098	2,796	3,166	3,210	3,476	3,628	
out/14	3,371	4,150	3,959	3,352	3,009	3,410	3,488	3,820	3,959	
nov/14	3,082	3,682	3,902	3,263	2,890	3,266	3,374	3,615	3,682	
dez/14	2,960	3,940	3,725	3,098	2,743	3,110	3,239	3,587	3,725	

Month	Real	M3	C3
jan/14	3,424	2,939	2,865
fev/14	3,083	2,764	2,743
mar/14	3,057	3,523	3,242
abr/14	2,955	3,371	3,015
mai/14	2,953	3,433	3,221
jun/14	2,669	3,593	3,200
jul/14	2,992	3,840	3,498
ago/14	2,915	4,051	3,824
set/14	3,210	3,702	3,628
out/14	3,371	4,150	3,959
nov/14	3,082	3,682	3,902
dez/14	2,960	3,940	3,725

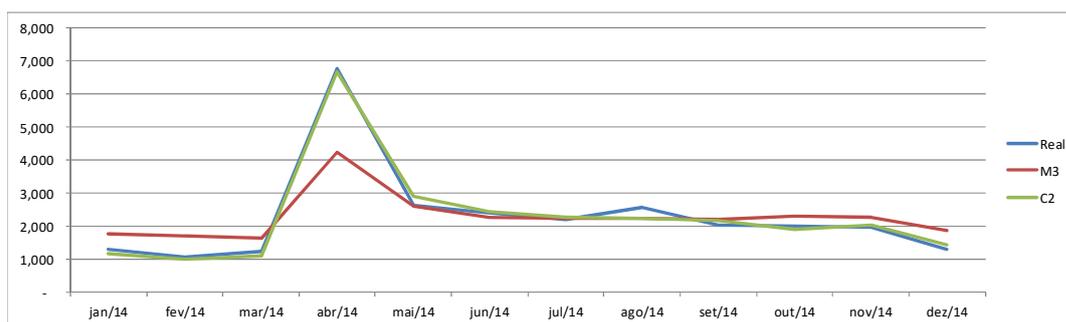
## 2. Imposto sobre produtos Industrializados (IPI)

Modelo	Erro acumulado percentual	Raiz do erro quadrático médio (RMSE)	Mean Percentual Error (MPE)	Mean Absolute Percentual Error (MAPE)	Mean Absolute Deviation (MAD)	Pesos					
						M1	M2	M3	C1	C2	C3
<b>Modelos Originais</b>											
SARIMA - M1	3.02%	R\$ 299	2.46%	3.90%	R\$ 183	1					
Holt-Winters - M2	7.06%	R\$ 422	6.54%	6.54%	R\$ 297		1				
Fatorial Dinamico - M3	-0.24%	R\$ 330	-0.98%	5.74%	R\$ 249				1		
<b>Combinações</b>											
Combinação Ótima - C1	15.43%	R\$ 686	15.28%	15.28%	R\$ 650	1.83	1.328	-2.158			
Regressão Linear - C2	-30.70%	R\$ 1,309	-31.26%	31.26%	R\$ 1,293	1.977	0.745	-1.3			
Performance - C3	2.75%	R\$ 315	2.13%	4.32%	R\$ 200	0.43	0.217	0.353			
Média Simples - C4	3.28%	R\$ 326	2.67%	4.37%	R\$ 203	0.333	0.333	0.333			
Mediana - C5	3.22%	R\$ 317	2.63%	4.07%	R\$ 192						
	Real	M1	M2	M3		C1	C2	C3	C4	C5	
jan/14	4,913	4,312	3,838	4,177		3,976	5,958	4,162	4,109	4,177	
fev/14	3,511	3,542	3,436	3,878		2,676	4,523	3,638	3,619	3,542	
mar/14	3,727	3,696	3,520	3,793		3,255	5,002	3,692	3,670	3,696	
abr/14	3,952	4,048	3,875	4,218		3,452	5,409	4,070	4,047	4,048	
mai/14	4,063	4,064	3,768	4,147		3,491	5,452	4,029	3,993	4,064	
jun/14	3,877	3,953	3,749	4,046		3,483	5,352	3,942	3,916	3,953	
jul/14	3,860	3,961	3,784	4,107		3,411	5,313	3,974	3,951	3,961	
ago/14	4,077	4,049	3,922	4,228		3,493	5,432	4,084	4,066	4,049	
set/14	4,196	4,228	4,136	4,399		3,738	5,724	4,268	4,254	4,228	
out/14	4,606	4,309	4,156	4,548		3,590	5,706	4,360	4,338	4,309	
nov/14	4,607	4,456	4,357	4,598		4,017	6,079	4,485	4,470	4,456	
dez/14	5,159	4,402	4,437	4,533		4,167	6,119	4,456	4,457	4,437	

### 3. Imposto Renda Sobre Pessoa Física (IRPF)

Modelo	Erro acumulado	Raiz do erro quadrático	Mean Percentual	Mean Absolute Percentual	Mean Absolute	Pesos		
	percentual	médio (RMSE)	Error (MPE)	Error (MAPE)	Deviation (MAD)	M1	M2	M3
<b>Modelos Originais</b>								
SARIMA - M1	-2.72%	R\$ 268	-2.58%	8.89%	R\$ 185	1		
Holt-Winters - M2	-4.52%	R\$ 196	-5.19%	7.27%	R\$ 148		1	
Fatorial Dinamico - M3	0.64%	R\$ 811	-12.74%	22.25%	R\$ 492			1
<b>Combinações</b>								
Combinação Ótima – C1	-4.23%	R\$ 191	-5.04%	7.24%	R\$ 146	0.11	0.873	0.018
Regressão Linear – C2	0.47%	R\$ 165	0.82%	7.40%	R\$ 142	0.12	0.901	-0.07
Performance – C3	-3.72%	R\$ 192	-4.59%	7.34%	R\$ 149	0.337	0.627	0.037
Média Simples – C4	-2.20%	R\$ 307	-6.84%	10.90%	R\$ 237	0.333	0.333	0.333
Mediana – C5	-3.25%	R\$ 176	-4.21%	6.75%	R\$ 134			

	Real	M1	M2	M3	C1	C2	C3	C4	C5
jan/14	1,310	1,296	1,274	1,778	1,285	1,179	1,300	1,449	1,296
fev/14	1,083	1,048	1,089	1,691	1,095	988	1,097	1,276	1,089
mar/14	1,234	1,096	1,216	1,650	1,211	1,112	1,192	1,321	1,216
abr/14	6,786	6,931	6,814	4,243	6,781	6,673	6,759	5,996	6,814
mai/14	2,641	2,792	3,071	2,605	3,032	2,920	2,960	2,823	2,792
jun/14	2,394	2,573	2,558	2,262	2,554	2,455	2,552	2,464	2,558
jul/14	2,219	2,344	2,396	2,243	2,387	2,283	2,373	2,328	2,344
ago/14	2,573	2,127	2,360	2,247	2,333	2,224	2,278	2,245	2,247
set/14	2,027	2,077	2,314	2,220	2,286	2,178	2,231	2,203	2,220
out/14	2,015	1,913	2,040	2,296	2,031	1,907	2,007	2,083	2,040
nov/14	1,968	2,692	2,091	2,260	2,160	2,048	2,299	2,348	2,260
dez/14	1,290	1,400	1,562	1,870	1,550	1,444	1,519	1,611	1,562



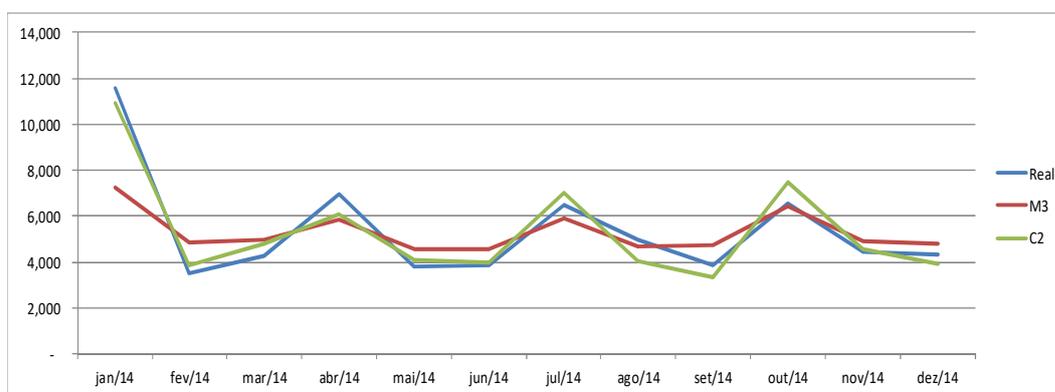
## 4. Imposto Renda Sobre Pessoa Jurídica (IRPJ)

Modelo	Erro acumulado percentual	Raiz do erro quadrático médio (RMSE)	Mean Percentual Error (MPE)	Mean Absolute Percentual Error (MAPE)	Mean Absolute Deviation (MAD)	Pesos			
						M1	M2	M3	
Modelos Originais									
SARIMA - M1	0.11%	R\$ 1,518	-0.77%	12.10%	R\$ 1,129	1			
Holt-Winters - M2	6.82%	R\$ 2,308	1.99%	13.00%	R\$ 1,523		1		
Fatorial Dinamico - M3	3.02%	R\$ 2,882	-5.00%	19.02%	R\$ 2,018			1	
Combinções									
Combinção Ótima - C1	-1.03%	R\$ 1,448	-3.36%	12.58%	R\$ 1,127	0.948	-0.34	0.392	
Regressão Linear - C2	10.09%	R\$ 1,780	9.17%	14.05%	R\$ 1,395	0.964	-0.208	0.134	
Performance - C3	2.28%	R\$ 1,572	-0.76%	12.29%	R\$ 1,242	0.585	0.253	0.162	
Média Simples - C4	3.31%	R\$ 1,925	-1.26%	13.95%	R\$ 1,473	0.333	0.333	0.333	
Mediana - C5	6.37%	R\$ 2,346	1.30%	13.69%	R\$ 1,569				
	Real	M1	M2	M3	C1	C2	C3	C4	C5
jan/14	22,154	22,787	15,602	13,833	21,718	20,572	19,518	17,408	15,602
fev/14	6,621	8,805	8,254	9,152	9,128	7,998	8,722	8,737	8,805
mar/14	8,681	10,551	9,915	9,741	10,450	9,414	10,259	10,069	9,915
abr/14	13,331	12,562	11,317	11,163	12,436	11,251	12,020	11,681	11,317
mai/14	7,138	6,656	7,370	8,731	7,226	6,054	7,173	7,586	7,370
jun/14	7,029	7,089	7,319	8,508	7,567	6,452	7,377	7,639	7,319
jul/14	11,905	12,007	11,360	10,922	11,802	10,675	11,667	11,430	11,360
ago/14	8,649	6,465	7,503	8,799	7,028	5,852	7,106	7,589	7,503
set/14	6,624	7,072	7,684	8,915	7,586	6,415	7,525	7,890	7,684
out/14	12,422	13,833	12,875	12,351	13,578	12,312	13,351	13,020	12,875
nov/14	8,135	8,068	8,146	9,271	8,513	7,326	8,283	8,495	8,146
dez/14	11,353	8,012	8,242	8,913	8,287	7,204	8,216	8,389	8,242

## 5. Contribuição Social sobre o Lucro Líquido (CSLL)

Modelo	Erro acumulado percentual	Raiz do erro quadrático médio (RMSE)	Mean Percentual Error (MPE)	Mean Absolute Percentual Error (MAPE)	Mean Absolute Deviation (MAD)	Pesos		
						M1	M2	M3
<b>Modelos Originais</b>								
SARIMAc- M1	10.15%	R\$ 1,234	5.78%	12.53%	R\$ 808	1		
Holt-Winters - M2	9.50%	R\$ 1,167	5.71%	10.87%	R\$ 708		1	
Fatorial Dinamico - M3	1.90%	R\$ 1,449	-5.37%	17.09%	R\$ 976			1
<b>Combinações</b>								
Combinação Ótima – C1	24.58%	R\$ 1,450	26.66%	26.66%	R\$ 1,324	1.183	1.702	-1.884
Regressão Linear – C2	0.96%	R\$ 590	0.32%	9.77%	R\$ 520	1.121	1.419	-1.326
Performance – C3	7.79%	R\$ 1,219	2.91%	10.99%	R\$ 725	0.352	0.393	0.255
Média Simples – C4	7.18%	R\$ 1,238	2.04%	11.50%	R\$ 746	0.333	0.333	0.333
Mediana – C5	5.96%	R\$ 1,157	1.06%	11.53%	R\$ 727			

	Real	M1	M2	M3	C1	C2	C3	C4	C5
jan/14	11,588	8,069	8,072	7,217	9,680	10,928	7,853	7,786	8069.46
fev/14	3,530	3,836	4,220	4,835	2,606	3,876	4,242	4,297	4220.04
mar/14	4,274	4,124	4,745	4,946	3,632	4,797	4,578	4,605	4745.24
abr/14	6,960	5,085	5,732	5,858	4,730	6,065	5,536	5,558	5732.31
mai/14	3,806	4,449	3,628	4,573	2,818	4,071	4,158	4,216	4449.2
jun/14	3,855	4,172	3,757	4,544	2,765	3,982	4,104	4,158	4171.98
jul/14	6,460	5,611	6,000	5,884	5,758	7,000	5,833	5,832	5883.98
ago/14	4,985	4,276	3,848	4,678	2,790	4,049	4,211	4,268	4276.49
set/14	3,844	3,687	3,860	4,734	2,008	3,332	4,022	4,093	3859.64
out/14	6,550	6,229	6,361	6,438	6,059	7,470	6,334	6,342	6360.54
nov/14	4,461	4,760	4,019	4,891	3,252	4,553	4,502	4,557	4760.41
dez/14	4,316	3,771	4,251	4,802	2,644	3,890	4,223	4,275	4250.85



## 6. Imposto de Renda Retido na Fonte (IRRF)

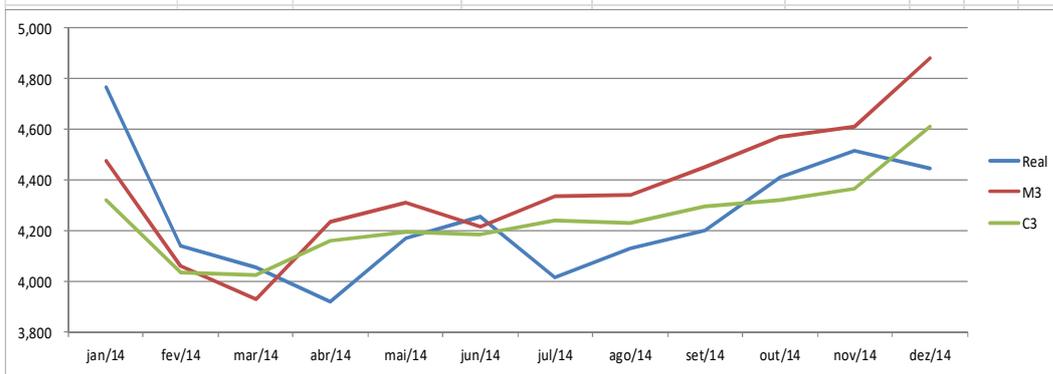
Modelo	Erro acumulado percentual	Raiz do erro quadrático médio (RMSE)	Mean Percentual Error (MPE)	Mean Absolute Percentual Error (MAPE)	Mean Absolute Deviation (MAD)	Pesos			
						M1	M2	M3	
<b>Modelos Originais</b>									
SARIMA - M1	0.60%	R\$ 539	0.39%	3.61%	R\$ 455	1			
Holt-Winters - M2	5.06%	R\$ 1,186	5.47%	7.03%	R\$ 946		1		
Fatorial Dinamico - M3	5.68%	R\$ 1,275	4.64%	6.38%	R\$ 929			1	
<b>Combinações</b>									
Combinação Ótima - C1	0.98%	R\$ 543	0.82%	3.75%	R\$ 479	0.915	0.082	0.003	
Regressão Linear - C2	22.58%	R\$ 2,966	23.06%	23.06%	R\$ 2,920	1.005	0.309	-0.549	
Performance - C3	1.92%	R\$ 596	1.69%	3.90%	R\$ 505	0.722	0.149	0.129	
Média Simples - C4	3.78%	R\$ 802	3.50%	4.61%	R\$ 614	0.333	0.333	0.333	
Mediana - C5	3.84%	R\$ 947	3.49%	4.98%	R\$ 682				
	Real	M1	M2	M3	C1	C2	C3	C4	C5
jan/14	16,092	15,510	13,729	13,270	15,357	12,540	14,956	14,169	13,729
fev/14	10,434	10,312	9,355	10,607	10,234	7,428	10,207	10,091	10,312
mar/14	11,799	11,270	10,835	11,000	11,233	8,631	11,170	11,035	11,000
abr/14	12,850	12,332	11,003	11,612	12,221	9,415	12,041	11,649	11,612
mai/14	11,100	12,209	10,564	11,375	12,072	9,286	11,857	11,383	11,375
jun/14	14,267	13,333	14,220	13,105	13,406	10,595	13,436	13,553	13,333
jul/14	11,027	11,423	10,951	11,339	11,384	8,635	11,342	11,238	11,339
ago/14	11,211	10,951	10,425	11,065	10,908	8,148	10,887	10,813	10,951
set/14	11,339	11,576	10,761	11,312	11,508	8,745	11,420	11,216	11,312
out/14	11,945	12,237	12,022	12,358	12,220	9,224	12,221	12,206	12,237
nov/14	13,970	13,718	12,647	12,565	13,626	10,792	13,410	12,977	12,647
dez/14	19,112	19,346	20,784	16,732	19,456	16,672	19,223	18,954	19,346

## 7. Imposto sobre Operações Financeiras (IOF)

Modelo	Erro acumulado	Raiz do erro quadrático	Mean Percentual	Mean Absolute Percentual	Mean Absolute	Pesos				
	percentual	médio (RMSE)	Error (MPE)	Error (MAPE)	Deviation (MAD)	M1	M2	M3		
Modelos Originais										
SARIMA - M1	-13.92%	R\$ 366	-14.31%	14.31%	R\$ 345	1				
Holt-Winters - M2	-11.11%	R\$ 306	-11.54%	11.54%	R\$ 275		1			
Fatorial Dinamico - M3	-19.83%	R\$ 507	-20.22%	20.22%	R\$ 492			1		
Combinções										
Combinção Ótima - C1	-20.83%	R\$ 528	-21.17%	21.17%	R\$ 517	1.868	-1.381	0.513		
Regressão Linear - C2	-208.03%	R\$ 5,159	-209.32%	209.32%	R\$ 5,158	-2.345	4.637	0.501		
Performance - C3	-13.60%	R\$ 359	-14.01%	14.01%	R\$ 337	0.338	0.485	0.176		
Média Simples - C4	-14.95%	R\$ 390	-15.36%	15.36%	R\$ 371	0.333	0.333	0.333		
Mediana - C5	-13.92%	R\$ 366	-14.31%	14.31%	R\$ 345					
	Real	M1	M2	M3	C1	C2	C3	C4	C5	
jan/14	2,172	2,739	2,678	2,903	2,907	7,448	2,738	2,773	2,739	
fev/14	2,472	2,693	2,657	2,869	2,833	7,441	2,706	2,740	2,693	
mar/14	2,201	2,684	2,657	2,708	2,734	7,379	2,675	2,683	2,684	
abr/14	2,427	2,858	2,779	2,990	3,035	7,680	2,843	2,876	2,858	
mai/14	2,364	2,804	2,740	2,990	2,987	7,629	2,806	2,845	2,804	
jun/14	2,583	2,865	2,789	2,913	2,995	7,671	2,836	2,856	2,865	
jul/14	2,399	2,824	2,757	2,931	2,971	7,629	2,810	2,837	2,824	
ago/14	2,567	2,819	2,763	2,984	2,981	7,695	2,821	2,855	2,819	
set/14	2,508	2,821	2,747	2,980	3,004	7,615	2,813	2,849	2,821	
out/14	2,502	2,841	2,775	3,133	3,083	7,771	2,860	2,916	2,841	
nov/14	2,757	2,870	2,785	3,117	3,114	7,745	2,872	2,924	2,870	
dez/14	2,799	3,076	2,930	3,132	3,306	7,940	3,015	3,046	3,076	

## 8. Contribuição ao Programa de Integração Social (CPIS)

Modelo	Erro acumulado	Raiz do erro quadrático	Mean Percentual	Mean Absolute Percentual	Mean Absolute	Pesos					
	percentual	médio (RMSE)	Error (MPE)	Error (MAPE)	Deviation (MAD)	M1	M2	M3			
Modelos Originais											
SARIMA - M1	1.40%	R\$ 237	1.12%	4.11%	R\$ 179	1					
Holt-Winters - M2	1.33%	R\$ 209	1.19%	3.99%	R\$ 173		1				
Fatorial Dinamico - M3	-2.72%	R\$ 234	-2.82%	4.82%	R\$ 205				1		
Combinções											
Combinção Ótima - C1	-1.57%	R\$ 189	-1.74%	3.45%	R\$ 147	0.447	-0.169	0.723			
Regressão Linear - C2											
Performance - C3	0.09%	R\$ 182	-0.08%	3.36%	R\$ 145	0.302	0.387	0.311			
Média Simples - C4	0.00%	R\$ 181	-0.17%	3.32%	R\$ 143	0.333	0.333	0.333			
Mediana - C5	0.48%	R\$ 207	0.31%	3.79%	R\$ 164						
	Real	M1	M2	M3	C1	C2	C3	C4	C5		
jan/14	4,765	4,194	4,292	4,476	4,381		4,320	4,321	4,292		
fev/14	4,140	4,194	3,897	4,061	4,148		4,038	4,051	4,061		
mar/14	4,058	4,194	3,973	3,931	4,041		4,027	4,033	3,973		
abr/14	3,922	4,194	4,072	4,233	4,243		4,159	4,167	4,194		
mai/14	4,171	4,194	4,107	4,313	4,295		4,197	4,205	4,194		
jun/14	4,257	4,194	4,150	4,218	4,218		4,184	4,187	4,194		
jul/14	4,018	4,194	4,200	4,334	4,294		4,240	4,242	4,200		
ago/14	4,132	4,194	4,174	4,341	4,304		4,232	4,236	4,194		
set/14	4,199	4,194	4,246	4,453	4,372		4,295	4,298	4,246		
out/14	4,412	4,194	4,214	4,572	4,464		4,319	4,327	4,214		
nov/14	4,517	4,194	4,305	4,611	4,477		4,367	4,370	4,305		
dez/14	4,446	4,194	4,725	4,882	4,601		4,613	4,600	4,725		



## 9. Contribuição para Financiamento da Seguridade Social (COFINS)

Modelo	Erro acumulado percentual	Raiz do erro quadrático médio (RMSE)	Mean Percentual Error (MPE)	Mean Absolute Percentual Error (MAPE)	Mean Absolute Deviation (MAD)	Pesos			
						M1	M2	M3	
Modelos Originais									
SARIMA - M1	-4.04%	R\$ 845	-4.14%	4.42%	R\$ 693	1			
Holt-Winters - M2	-6.15%	R\$ 1,174	-6.23%	6.54%	R\$ 1,038		1		
Fatorial Dinamico - M3	-1.81%	R\$ 624	-1.86%	3.49%	R\$ 554			1	
Combinções									
Combinção Ótima - C1	-1.06%	R\$ 490	-1.13%	2.49%	R\$ 387	0.604	-0.483	0.879	
Regressão Linear - C2	9.09%	R\$ 1,522	9.03%	9.03%	R\$ 1,452	0.508	-0.468	0.862	
Performance - C3	-3.15%	R\$ 711	-3.21%	3.66%	R\$ 578	0.298	0.154	0.547	
Média Simples - C4	-4.00%	R\$ 822	-4.08%	4.35%	R\$ 686	0.333	0.333	0.333	
Mediana - C5	-4.01%	R\$ 814	-4.10%	4.41%	R\$ 694				
	Real	M1	M2	M3	C1	C2	C3	C4	C5
jan/14	17,326	17,675	16,999	16,539	17,004	15,277	17,326	17,071	16,999
fev/14	15,089	15,171	15,111	14,903	14,965	13,478	15,089	15,062	15,111
mar/14	15,005	15,591	15,722	14,544	14,608	13,096	15,005	15,286	15,591
abr/14	14,922	16,497	16,543	15,889	15,940	14,331	14,922	16,310	16,497
mai/14	15,737	17,039	16,605	16,068	16,395	14,731	15,737	16,571	16,605
jun/14	15,964	16,347	16,756	15,810	15,677	14,087	15,964	16,304	16,347
jul/14	15,218	16,531	17,080	16,346	16,103	14,491	15,218	16,652	16,531
ago/14	15,720	16,619	17,298	16,455	16,147	14,528	15,720	16,790	16,619
set/14	16,104	16,585	17,428	16,665	16,249	14,631	16,104	16,893	16,665
out/14	16,834	16,547	17,586	17,276	16,686	15,064	16,834	17,136	17,276
nov/14	17,205	17,310	18,127	17,608	17,177	15,484	17,205	17,682	17,608
dez/14	16,592	17,547	18,260	17,088	16,799	15,094	16,592	17,632	17,547

