

**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**  
**INSTITUTO DE ECONOMIA**  
**PROGRAMA DE PÓS GRADUAÇÃO EM ECONOMIA**

**João Ricardo Cruz Moço**

**Estimação de um Modelo de Previsão do ICMS do**  
**Estado do Rio de Janeiro**

**RIO DE JANEIRO**

**2017**

**João Ricardo Cruz Moço**

**Estimação de um Modelo de Previsão do ICMS do  
Estado do Rio de Janeiro**

Dissertação de Mestrado apresentada ao Programa de Pós Graduação em Economia do Instituto de Economia da Universidade Federal do Rio de Janeiro, como requisito para a obtenção do título de Mestre em Economia.

**Orientadora: Prof<sup>a</sup>. Dra. Viviane Luporini**

**RIO DE JANEIRO**

**2017**

## FICHA CATALOGRÁFICA

M687 Moço, João Ricardo Cruz.  
Estimação de um modelo de previsão do ICMS do Estado do Rio de Janeiro /  
João Ricardo Cruz Moço. – 2017.  
88 p. ; 31 cm.

Orientador: Viviane Luporini.  
Dissertação (mestrado) – Universidade Federal do Rio de Janeiro, Instituto de  
Economia, Programa de Pós-Graduação em Economia da Indústria e da  
Tecnologia, 2017.

Bibliografia: f. 84 – 88.

1. Imposto sobre circulação de mercadorias e serviços. 2. Modelo de previsão. 3.  
Modelos econométricos. I. Luporini, Viviane, orient. II. Universidade Federal do Rio  
de Janeiro. Instituto de Economia. III. Título.

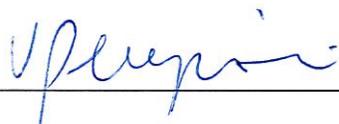
CDD 343.0558

**João Ricardo Cruz Moço**

**ESTIMAÇÃO DE UM MODELO DE PREVISÃO DO ICMS DO  
ESTADO DO RIO DE JANEIRO**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Economia da Indústria e da Tecnologia (PPGE), do Instituto de Economia da Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Ciências, em Economia.

Aprovada em: 22/06/2017



**VIVIANE LUPORINI (orientadora)**



**EDUARDO PONTUAL RIBEIRO**



**MARGARIDA MARIA GOMES PEREIRA SARMIENTO GUTIERREZ**

**ANTÔNIO LUIS LICHA (suplente interno)**

**JULIA DE MEDEIROS BRAGA (suplente externo)**

## AGRADECIMENTOS

Aos meus pais, Maria de Lourdes e Ricardo, pelo apoio dado ao longo de toda a vida, especialmente no período da graduação e do mestrado.

A minha namorada, Sara, pelo apoio e companheirismo, além da compreensão pelos períodos de ausência durante o período da dissertação.

Aos meus amigos de trabalho na AgeRio, Graciela, Edson, André e Carlos Guilherme pela confiança e flexibilidade dadas em diversos momentos quando tinha compromissos no mestrado.

Agradeço ao pessoal da SEFAZ-RJ – Henrique Reis, Raphael Philipe e Liliane Figueiredo – pela disponibilidade em tirar minhas dúvidas. Sem dúvida, sem essa ajuda o trabalho ficaria muito mais difícil.

A todos os colegas de mestrado, especialmente ao Rodrigo e ao André (de novo) que várias vezes me ajudaram de forma muito importante em trabalhos, dúvidas...

Aos amigos Luan, Romilson e Marcos pelas inúmeras contribuições dadas durante a graduação na UERJ que foram fundamentais para minha formação. Ao Luan também tenho que agradecer pela época de CATE – nesse momento também tenho que agradecer ao Jucá – e também durante o mestrado.

Por fim agradeço a minha orientadora, Viviane Luporini, pela paciência e pela disponibilidade ao longo do período da dissertação e em outros momentos ao longo do mestrado.

## RESUMO

**MOÇO, J. R. C. *Estimação de um Modelo de Previsão do ICMS do Estado do Rio de Janeiro*. Rio de Janeiro, 2017. Dissertação (Mestrado em Economia) – Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.**

O exercício da previsão das receitas governamentais é de extrema importância para uma adequada execução orçamentária. Uma boa acurácia na estimação permite que seja estipulado um nível de gasto que atenda as demandas da população, respeitando a restrição oriunda dos valores previstos. Essa importância é explicitada na Lei de Responsabilidade Fiscal, que obriga os entes da federação a realizarem esse trabalho e divulgá-lo na Lei de Diretrizes Orçamentária (LDO) e na Lei Orçamentária Anual (LOA). No Estado do Rio de Janeiro, isso é feito pela Secretaria de Estado de Fazenda e Planejamento (SEFAZ-RJ), que desde 2013 o faz utilizando métodos econométricos de previsão. Entre os itens previstos, o que mais se destaca-se individualmente é o ICMS, correspondendo a 40,44% das receitas do Estado em 2015. Conforme estimações apresentadas pela SEFAZ-RJ, os erros de previsão para tal tributo nos anos de 2014, 2015 e 2016 foram de, respectivamente, 1,92%, 14,75% e 12,64%. Os erros significativos de previsão no últimos dois anos motivaram este trabalho que tem como objetivo estimar modelos econométricos univariados, multivariados e com métodos de combinação e comparar seus resultados com o modelo oficial utilizado pela SEFAZ-RJ. As estimações foram feitas usando dados de janeiro de 2003 até junho de 2015 e as previsões realizadas para julho de 2015 até dezembro de 2016, tendo as estatísticas de erro de previsão computadas apenas com as previsões para os meses de 2016. Os resultados apontaram para relevante aumento da acurácia preditiva dos modelos aqui estimados quando comparados com o oficial. Em especial, destaca-se que o melhor modelo de combinação de previsão obtido gerou os resultados mais acurados para quatro das cinco estatísticas de erro consideradas, enquanto que, para a estatística restante, a do Erro % Anual, o modelo multivariado foi, de longe, o mais acurado.

**Palavras-Chave: Previsão. ICMS. Rio de Janeiro. Métodos Econométricos.**

## ABSTRACT

**MOÇO, J. R. C. *Estimação de um Modelo de Previsão do ICMS do Estado do Rio de Janeiro*. Rio de Janeiro, 2017. Dissertação (Mestrado em Economia) – Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.**

The exercise of government revenue forecasting is of extreme importance for adequate budget execution. A good accuracy in this estimation allows a level of spending that meets the demands of the population, respecting the restriction coming from the predicted values. This importance is made explicit in the Fiscal Responsibility Law, which obliges the federation's entities to carry out this work and to disclose it in the Law of Budgetary Guidelines (LDO) and in the Annual Budget Law (LOA). In the State of Rio de Janeiro, this is done by the Secretary of State for Finance and Planning (SEFAZ-RJ), which has done this since 2013 using econometric forecasting methods. Among the items forecast, the most significant is the ICMS, corresponding to 40.44% of the state revenues in 2015. According to estimates presented by SEFAZ-RJ, the prediction errors for such tax in the years of 2014, 2015 and 2016 were, respectively, 1.92%, 14.75% and 12.64%. Significant forecasting errors over the last two years have motivated this work, which aims to estimate univariate, multivariate, and combination econometric models and to compare their results with the official SEFAZ-RJ model. Estimates were made using data from January 2003 to June 2015 and forecasts for July 2015 to December 2016, with prediction error statisticians computed only with forecasts for the months of 2016. The results pointed to significant Increase in the predictive accuracy of the models here estimated when compared with the official one. In particular, it should be noted that the best prediction combination model obtained yielded the most accurate results for four of the five error statistics considered, while for the remaining statistic, that of the % Annual Error, the multivariate model was by far the most accurate.

**Keywords: Forecasting. ICMS. Rio de Janeiro. Econometric Methods.**

## Lista de Tabelas

Tabela 1 – Arrecadação Governo Geral (2015) por Base de Incidência .....	32
Tabela 2 – Carga Tributária (2015) - Base de Incidência Bens e Serviços .....	33
Tabela 3 – Componentes da Receita Corrente (2015) – ERJ .....	35
Tabela 4 – Componentes da Receita de Capital (2015) – ERJ.....	35
Tabela 5 – Componentes da Receita Tributária (2015) – ERJ.....	36
Tabela 6 – Seção Comércio CNAE – ICMS Estado do Rio de Janeiro - 2015.....	37
Tabela 7 – Seção Indústria de Transformação CNAE – ICMS ERJ - 2015 .....	37
Tabela 8 – Seção Eletricidade e Gás CNAE – ICMS ERJ - 2015 .....	38
Tabela 9 – Seção Inf. e Comunicação CNAE – ICMS RJ - 2015 .....	38
Tabela 10 – Propriedades FAC e FACP – Alguns Casos .....	46
Tabela 11 – Variáveis Explicativas – Modelos Multivariados .....	59
Tabela 12 – Testes de Causalidade de Granger.....	60
Tabela 13 – Testes ADF, PP e KPSS – Séries em Nível – Estatísticas de Teste.....	61
Tabela 14 – Testes ADF, PP e KPSS – Séries em 1ª Dif. – Estatísticas de Teste ...	61
Tabela 15 – Testes ADF, PP e KPSS – ICMS – Estatísticas de Teste .....	63
Tabela 16 – Resultados de Previsão: Modelo Original SEFAZ-RJ e Versões Alternativas .....	65
Tabela 17 – Modelo Original SEFAZ-RJ .....	66
Tabela 18 – Testes ADF para a série do ICMS-RJ com dif. Sazonal (D=12).....	68
Tabela 19 – Modelos SARIMA – Análise Estrutural – R <sup>2</sup> e Significância dos Parâmetros .....	70
Tabela 20 – Modelos SARIMA – Análise Estrutural – Critérios de Inf. ....	70
Tabela 21 – Modelos SARIMA – Análise dos Resíduos.....	71

Tabela 22 – Resultados de Previsão: Modelos SARIMA.....	71
Tabela 23 – Resultados de Previsão: SARIMA e Holt-Winters .....	72
Tabela 24 – Teste de Cointegração de Johansen.....	74
Tabela 25 – Modelos Multivariados – Análise dos Resíduos .....	76
Tabela 26 – Resultados de Previsão: Modelos Multivariados .....	76
Tabela 27 – Métodos de Combinação: Pesos para os Modelos de Previsão .....	78
Tabela 28 – Resultados de Previsão: Métodos de Combinação .....	78
Tabela 29 – Resultados de Previsão: Comparação Final .....	79

## **Lista de Figuras**

Figura 1 – Carga Tributária em Relação ao PIB – OCDE e Brasil - 2015 .....	32
Figura 2 – Tributos com Base de Incidência Bens e Serviços (% PIB) - OCDE e Brasi.....	34
Figura 3 – Base de Incidência do ICMS-RJ (2015) – Seções do CNAE .....	37
Figura 4 – Série ICMS – Estado do Rio de Janeiro – 2003-2016.....	62
Figura 5 – Sazonalidade do ICMS-RJ .....	67
Figura 6 – Série do ICMS com diferenciação Sazonal (D=12) .....	68
Figura 7 – Previsões 2016 – Modelos Considerados .....	80

## Lista de Abreviaturas e Siglas

ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
CTB	Carga Tributária Bruta
ERJ	Estado do Rio de Janeiro
FECP	Fundo Estadual de Combate à Pobreza e às Desigualdades Sociais
ICMS	Imposto Sobre Circulação de Mercadorias e Prestação de Serviços
IPVA	Imposto sobre a Propriedade de Veículos Automotores
ITD	Imposto de Transmissão Causa Mortis e Doação
LRF	Lei de Responsabilidade Fiscal
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MSE	<i>Mean Square Error</i>
RFB	Secretaria da Receita Federal do Brasil
RMSE	<i>Root Mean Square Error</i>
SARIMA	<i>Seasonal Autoregressive Integrated Moving Average Model</i>
SEFAZ	Secretaria de Estado de Fazenda
VAR	<i>Vector Autoregressive</i>
VECM	<i>Vector Error Correction Model</i>

## Sumário

Introdução.....	14
1 Exercício da Previsão Orçamentária.....	17
1.1 Relevância do Exercício da Previsão Orçamentária.....	17
1.2 Prática da Previsão por Entes da Federação .....	19
1.3 Previsão de Receitas no Estado do Rio de Janeiro .....	20
2 Revisão da Literatura.....	23
2.1 Comparação de Modelos de Previsão.....	23
2.2 Práticas de Previsão .....	28
3 Contexto Empírico.....	31
3.1 Arrecadação Tributária no Brasil .....	31
3.2 Receita Estado do Rio de Janeiro .....	34
3.3 ICMS no Estado do Rio de Janeiro .....	36
4 Metodologia e Dados.....	39
4.1 Metodologia.....	39
4.1.1 Procedimentos da Previsão .....	39
4.1.1.1 Ideia Geral .....	39
4.1.1.2 Modelos Utilizados .....	39
4.1.1.3 Periodicidades da Estimação e Previsão .....	39
4.1.1.4 Metodologia de Previsão.....	40
4.1.1.5 Comparação dos Modelos .....	41
4.1.2 Modelos Utilizados.....	42
4.1.2.1 Modelos Univariados .....	42
4.1.2.1.1 SARIMA .....	42
4.1.2.1.2 Modelo de Holt-Winters .....	50
4.1.2.2 Modelos Multivariados .....	53
4.1.2.2.1 VAR/VEC.....	53
4.1.2.3 Métodos de Combinação.....	56
4.1.2.4 Estatísticas de Erro de Previsão .....	58
4.2 Dados.....	58
4.2.1 Séries Utilizadas .....	58
4.2.2 Testes de Estacionariedade.....	60
4.2.3 Série do ICMS .....	62
5 Resultados.....	64
5.1 Análise do Modelo da SEFAZ-RJ.....	64
5.2 Análise Univariada .....	67

5.2.1	Exame Preliminar da Série do ICMS .....	67
5.2.2	Escolha das Especificações do Modelo SARIMA .....	69
5.2.3	Comparação dos Modelos SARIMA .....	69
5.3	Análise Multivariada .....	73
5.3.1	Descrição dos Modelos Multivariados Considerados .....	73
5.3.2	Análise da Cointegração .....	74
5.3.3	Comparação dos Modelos Multivariados .....	75
5.4	Combinação de Modelos de Previsão.....	77
5.5	Comparação dos Métodos de Previsão.....	79
6	Conclusão .....	82
	REFERÊNCIAS .....	84

## Introdução

O trabalho de previsão de receitas públicas é de extrema importância para a adequada execução orçamentária dos governos. Uma acurada estimação das receitas permite que sejam atendidas as demandas da população ao mesmo tempo que se respeita a restrição orçamentária governamental dada pelos valores previstos.

A Lei de Responsabilidade Fiscal (LRF) corrobora com essa importância. No seu artigo 11 é dito que “Constituem requisitos essenciais da responsabilidade na gestão fiscal a instituição, previsão e efetiva arrecadação de todos os tributos da competência constitucional do ente da Federação”.

Ademais, a LRF define que as projeções de receitas devem ser divulgadas por todos os entes da federação através da Lei de Diretrizes Orçamentárias e da Lei Orçamentária Anual, conforme estabelecido no §1º do artigo 4º e no inciso I artigo 5º da LRF.

No caso específico do Estado do Rio de Janeiro (ERJ) esse trabalho é feito pela Secretaria de Estado de Fazenda e Planejamento (SEFAZ-RJ) que desde 2013 utiliza métodos econométricos para as estimações das receitas tributárias do Estado. Os métodos são tanto univariados, SARIMA, quanto multivariados, VAR.

Em relação ao total de receitas do Estado, as tributárias são de longe a de maior destaque individual, correspondendo em 2015 - conforme o Boletim de Transparência Fiscal do último bimestre de 2015, publicado pela SEFAZ-RJ - por 54,89% do total de receitas do Estado. Dentro da arrecadação tributária, o item mais relevante é o ICMS, que responde por 73,67% dos tributos do Estado e por 40,44% do total das receitas.

Dada sua importância para o orçamento fluminense, o estudo da capacidade preditiva dos modelos oficiais de previsão do ICMS se mostra de extrema importância. Analisando os resultados das estimações realizadas a partir de 2013, ano que a SEFAZ-RJ passou a utilizar técnicas econométricas de previsão, se observou uma significativa piora de acurácia de seu modelo nos últimos dois anos. Os erros de previsão anuais foram os seguintes: 1,92% em 2014, 14,75% em 2015 e 12,64% em 2016.

No sentido de tentar aprimorar a metodologia oficial atualmente utilizada, este trabalho tem como objetivo comparar a acurácia preditiva de diferentes métodos econométricos com modelo oficial da SEFAZ-RJ. Os métodos aqui utilizados serão os univariados, o multivariados e os de combinação de previsão.

No caso univariado serão utilizadas as metodologias SARIMA (*Seasonal Autoregressive Integrated Moving Average Model*) e Holt-Winters. Nos multivariados, serão considerados os modelos VAR (Vetores Autoregressivos) e VECM (Modelo com Vetor de Correção de Erros). Já as técnicas de combinação utilizadas serão: Média Simples, *Least Squares Weights*, *Mean Square Error (MSE) Weights* e *Mean Square Error (MSE) Rank*.

Com dados de janeiro de 2003 até junho de 2015, os modelos aqui testados serão estimados. Com posse deles serão realizadas as previsões para o período de julho de 2015 até dezembro de 2016. Contudo, as estatísticas de erros de previsão utilizadas neste trabalho (Erro % Anual, MSE, RMSE, MAE e MAPE) serão calculadas apenas com as previsões dos meses de 2016.

A escolha de junho como mês de estimação dos modelos se justifica pelo fato de que, na média dos últimos quatro anos, a SEFAZ-RJ realiza as suas previsões no começo do mês de agosto, utilizando dados disponíveis até junho do ano em questão, uma vez que existe uma defasagem de dois meses da divulgação de algumas variáveis utilizadas no modelo.

Vale ressaltar que primeiro serão feitas comparações com diferentes modelos univariados e então escolhido um melhor. O mesmo procedimento será feito com os multivariados. Com o melhor modelo univariado e o melhor multivariado, serão utilizadas as técnicas de combinação de previsão citadas e com os resultados obtidos será escolhido o melhor modelo de combinação.

Destaca-se que as comparações citadas levarão em conta não apenas a capacidade preditiva dos modelos, mas também a estrutura dos mesmos, testada via, entre outras coisas, análise dos resíduos para identificar problemas como autocorrelação serial, heterocedasticidade e ausência de normalidade.

No final desse processo de comparação, se terá em mãos três modelos: o melhor univariado, o melhor multivariado e o melhor de combinação. Eles serão comparados entre si com base nas cinco estatísticas de erro anteriormente citadas.

Este trabalho está dividido da seguinte maneira: no capítulo seguinte a esta introdução é feita uma motivação do assunto da estimação de receitas públicas, além de apresentar as técnicas utilizadas pelos diferentes entes da federação, destacando em especial o Estado do Rio de Janeiro. No capítulo 2 é apresentada uma revisão da bibliografia do assunto. Na primeira parte são destacados trabalhos similares à esta dissertação, no sentido de que comparam a acurácia preditiva de diferentes modelos. Na segunda parte o foco é passado para textos que tratam da influência de fatores institucionais sobre o exercício da previsão. Na parte 3, é feita uma contextualização geral das receitas do Governo Federal e do Estado do Rio de Janeiro, destacando em especial a importância do tributo ICMS. No quarto capítulo será melhor explicada a metodologia utilizada neste trabalho, descrevendo os modelos e métodos de comparação empregados. Neste mesmo capítulo, serão expostas as séries de dados utilizadas. Por fim, na parte 5 do texto serão mostrados os resultados obtidos das diversas comparações realizadas.

# 1 Exercício da Previsão Orçamentária

## 1.1 Relevância do Exercício da Previsão Orçamentária

Clements e Hendry (2004, p.2) definem o exercício da previsão como sendo: “Any statement about the future. Such statements may be well founded, or lack any sound basis; they may be accurate or inaccurate on any given occasion, or on average; precise or imprecise; and model-based or informal”.

No campo econômico, o trabalho da previsão é utilizado para basicamente todas as séries com dados disponíveis. Variáveis como PIB, inflação, receita governamental, câmbio e outras são comumente previstas pelos economistas, antecipando seus comportamentos futuros a fim de otimizar as decisões tomadas no momento atual.

Especificamente para a previsão das receitas públicas, tem-se que esse exercício é de extrema importância para a adequada execução orçamentária dos governos, dado que uma boa acurácia na estimação das receitas para os anos vindouros permite que seja determinado um nível de gasto futuro que, ao mesmo tempo, consiga atender as demandas da população e respeitar a restrição oriunda dos valores previstos para a arrecadação.

Leal *et al.* (2007) destacam que a elaboração dos orçamentos governamentais está no cerne da *accountability* que os governos têm que realizar pelo seu uso dos recursos públicos nas sociedades democráticas. Dentro da elaboração do orçamento, a tarefa da previsão de receitas ocupa posição relevante.

A importância da previsão de receitas governamentais está presente em documentos públicos e na própria legislação brasileira. Por exemplo, conforme o Manual de Procedimentos de Receitas Públicas do Tesouro Nacional:

É notável a relevância da Receita Pública no processo orçamentário, cuja previsão dimensiona a capacidade governamental em fixar a Despesa Pública e, no momento da sua arrecadação, torna-se instrumento condicionante da execução orçamentária da despesa. (Manual de Procedimentos de Receitas Públicas do Tesouro Nacional, 2007, p.13).

Também conforme o referido Manual:

A projeção das receitas é fundamental na determinação das despesas, pois é a base para a fixação destas na Lei Orçamentária Anual, na execução do orçamento e para a determinação das necessidades de financiamento do Governo. Além disso, é primordial sua análise na concessão de créditos suplementares por excesso de arrecadação. (Manual de Procedimentos de Receitas Públicas do Tesouro Nacional, 2007, p.30).

Essa importância também é explicitada na Lei Complementar nº 101 de 04 de maio de 2000, a chamada Lei de Responsabilidade Fiscal (LRF). No seu artigo 11 é dito que “Constituem requisitos essenciais da responsabilidade na gestão fiscal a instituição, previsão e efetiva arrecadação de todos os tributos da competência constitucional do ente da Federação”.

A LRF também dispõe da forma como os entes da federação devem divulgar as previsões realizadas. No §1º do artigo 4º da LRF é expresso que:

Integrará o projeto de lei de diretrizes orçamentárias Anexo de Metas Fiscais, em que serão estabelecidas metas anuais, em valores correntes e constantes, relativas a receitas, despesas, resultados nominal e primário e montante da dívida pública, para o exercício a que se referirem e para os dois seguintes. (Lei Complementar Nº 101, 2000, §1º do artigo 4º).

Além disso, no inciso I artigo 5º da referida lei é dito que a Lei Orçamentária Anual “conterá, em anexo, demonstrativo da compatibilidade da programação dos orçamentos com os objetivos e metas constantes do documento de que trata o § 1º do artigo 4º da própria LRF.

Do que foi dito acima, pode-se resumir que a LRF exige que os entes da federação apresentem as previsões para as suas receitas no projeto de Lei de Diretrizes Orçamentárias (LDO) e no projeto de Lei Orçamentária Anual (LOA), sendo que o anexo da LOA atualiza os valores previstos no anexo de metas fiscais da LDO.

Para o caso do Estado do Rio de Janeiro (ERJ), que será o abordado neste trabalho, a aplicabilidade das exigências da LRF citadas acima é confirmada na constituição do ERJ, especialmente em seus artigos 209 e 210.

## 1.2 Prática da Previsão por Entes da Federação

No nível federal, o método de previsão utilizado é o chamado “Método dos Indicadores”. Essa abordagem realiza suas previsões via multiplicação das receitas por indicadores de preço e de quantidade, referentes aos períodos futuros. Esses índices, para os períodos previstos, podem ser obtidos via hipótese de que eles serão idênticos aos valores observados no momento da previsão ou via utilização de projeções dadas por órgãos oficiais de pesquisa ou pelas próprias secretarias de fazenda.

Conforme o Manual de Procedimentos de Receitas Públicas do Tesouro Nacional:

A metodologia de projeção de receitas orçamentárias adotada pelo Governo Federal está baseada na série histórica de arrecadação das mesmas ao longo dos anos ou meses anteriores (base de cálculo), corrigida por parâmetros de preço (efeito preço), de quantidade (efeito quantidade) e de alguma mudança de aplicação de alíquota em sua base de cálculo (efeito legislação). (Manual de Procedimentos de Receitas Públicas do Tesouro Nacional, 2007, p.30).

Seguindo a formulação proposta por Melo (2001), o Método dos Indicadores pode ser representado da seguinte maneira:

$$X_t = X_{t-1}(1 + \Delta P)(1 + \Delta Q)(1 + \Delta L)(1 + \Delta U) \quad (1.1)$$

Na equação acima, P indica índice preço, Q índice de quantidade, L índice que representa alteração na legislação tributária e U é um índice residual que trata de todos os outros fatores que também determinam a arrecadação n data t.

Conforme será mais bem explicado no capítulo 2, o método dos indicadores já motivou alguns estudos empíricos que investigaram a possibilidade de redução dos erros de previsão dados por esse método via utilização de especificações alternativas, especialmente as econométricas. Nesse sentido, pode-se citar como exemplos os trabalhos de Melo (2001), Siqueira (2002), Campos (2009) e Benelli (2013).

Uma breve revisão realizada pelo autor, com base nas LDO de 2016, dos métodos de previsão de receitas usados em alguns dos estados brasileiros mostrou relevante difusão do método dos indicadores e de versões similares. Estados como Bahia, Espírito Santo, Ceará e Santa Catarina são utilizadores de tal método. O uso de técnicas econométricas é algo menso comum. Além do Rio de Janeiro, outro estado que utiliza tais técnicas é o de Minas Gerais, que utiliza modelos econométricos univariados (ARIMA, Holt-Winters e ARIMAX).

### **1.3 Previsão de Receitas no Estado do Rio de Janeiro**

No Estado do Rio de Janeiro, o trabalho da previsão das receitas públicas, tanto para a LDO quanto para a LOA, é realizado pela Secretaria de Estado de Fazenda e Planejamento (SEFAZ-RJ)<sup>1</sup>. Via de regra, em fevereiro ou março, no caso da LDO, e, em maio ou junho, para a LOA, são publicados decretos do poder executivo do ERJ, estabelecendo as diretrizes para a elaboração das citadas leis orçamentárias.

Até o ano de 2016, nesses decretos era estabelecido que o trabalho da estimação da receita estadual devia ser feito pela Secretaria de Estado de Fazenda (SEFAZ-RJ) que seria incluído pela Secretaria de Estado de Planejamento e Gestão (SEPLAG-RJ) no projeto de LDO ou LOA, a depender do caso.

A explicação que segue abaixo considera o fluxo observado até o ano de 2016 quando SEFAZ-RJ e SEPLAG-RJ eram instituições diferentes. A ideia é que, mesmo com a junção das secretarias, o fluxo de elaboração das leis orçamentárias continua com a mesma lógica, só que dentro de uma única instituição.

Assim, após a publicação do decreto do poder executivo do estado, a SEFAZ-RJ estima seus modelos de previsão de receitas, que no caso das tributárias, se baseiam em metodologia econométrica.

No caso da LDO, de modo a respeitar o prazo de publicação dessa lei - que na média dos últimos quatro anos ocorreu em começo de julho - a SEFAZ-RJ

---

<sup>1</sup> O Decreto nº 45.896 de 27/01/2017 incorporou a Secretaria de Estado de Planejamento e Gestão - SEPLAG - à Secretaria de Estado de Fazenda - SEFAZ.

enviava sua nota técnica<sup>2</sup> para a SEPLAG-RJ incluir no projeto de Lei de Diretrizes Orçamentárias do ano no final do mês de março. Assim, há o intervalo de tempo nos meses de abril, maio e junho para término de elaboração da lei e discussão na Assembleia Legislativa.

Ideia semelhante vale para a elaboração da LOA. A SEFAZ-RJ elabora nota técnica com as previsões de receitas para a LOA, na média dos últimos quatro anos, no começo de agosto. Os valores previstos nessa nota técnica eram usados pela SEPLAG-RJ na elaboração do projeto de Lei Orçamentária do ano em questão, que, na mesma média citada, é aprovado no começo de janeiro do ano seguinte.

Assim, no mês de março, para a LDO, e de agosto, para a LOA, a SEFAZ-RJ estima os parâmetros dos seus modelos com os dados a partir de janeiro de 2003 até os disponíveis no mês da elaboração. Com o modelo estimado, é feita a previsão mensal, consolidada em base anual, para os três anos seguintes.

Conforme Nota Técnica SUPOF/SEFAZ-RJ 09/2016 de 12/07/2016<sup>3</sup>, as previsões para as receitas tributárias do Estado do Rio de Janeiro, a saber, ICMS, IPVA, ITD e FECF, são feitas com o uso de técnicas econométricas.

Para o caso do ICMS, motivação central deste trabalho dada sua importância em relação ao total das receitas do estado<sup>4</sup>, o modelo utilizado pela SEFAZ-RJ nos últimos quatro anos<sup>5</sup> para realização de previsão tem as seguintes características:

- Metodologia: VAR com séries em nível.
- Dados: Além da Série do ICMS do ERJ, são usadas as seguintes variáveis: Pesquisa Mensal de Comércio ERJ – PMC-RJ do IBGE; PIB Mensal Calculado pelo BACEN; Índice de Atividade Econômica Brasil – IBC-BR do BACEN; Taxa de Câmbio Efetiva Real (IPCA) e; série de Vendas, pelas Distribuidoras, dos Derivados de Petróleo em barris (ANP).

---

<sup>2</sup> Documento que apresenta as previsões de receitas junto com suas respectivas metodologias de cálculo.

<sup>3</sup> [http://www.transparencia.rj.gov.br/transparencia/content/conn/UCMServer/path/Contribution%20Folders/site\\_fazenda/transpFiscal/notas\\_tecnicas\\_supof/2016/NOTA%20T%c3%89CNICA%2009%20de%202016%20-%20Metodologia%20de%20previs%c3%a3o%20PLOA%202017.pdf?lve](http://www.transparencia.rj.gov.br/transparencia/content/conn/UCMServer/path/Contribution%20Folders/site_fazenda/transpFiscal/notas_tecnicas_supof/2016/NOTA%20T%c3%89CNICA%2009%20de%202016%20-%20Metodologia%20de%20previs%c3%a3o%20PLOA%202017.pdf?lve)

<sup>4</sup> Conforme explicação apresentada no capítulo 3, o ICMS correspondeu a 40,44% do total das receitas do ERJ em 2015.

<sup>5</sup> Período que a SEFAZ-RJ passou a usar técnicas econométricas na previsão de receitas.

Vale ressaltar que o modelo é estimado com duas variáveis endógenas: ICMS e PMC-RJ e três exógenas: PIB Mensal, Câmbio Efetivo Real e Vendas de Derivados de Petróleo.

No capítulo 5, o dos resultados, será apresentado com maiores detalhes o modelo utilizado pela SEFAZ-RJ, assim como serão destacados os resultados de previsão desse modelo em sua formulação original e em versões alternativas, no sentido de dar alguma intuição de suas fragilidades.

Desde o ano que a SEFAZ-RJ passou a realizar estimações para o ICMS usando técnicas econométricas, os seja em 2013, os erros de previsão anuais obtidos foram os seguintes: 1,92% em 2014, 14,75% em 2015 e 12,64% em 2016.

Esses percentuais de erros foram calculados com base nas previsões dadas nas notas técnicas da LOA e não nas da versão final dessa lei. Essa opção foi baseada no fato de que os valores dados nas notas refletem apenas os fatores técnicos de sua previsão e, assim, vão ao encontro do objetivo deste trabalho.

Percebe-se que tais erros têm proporções não desprezíveis para os últimos dois anos, possivelmente explicados pela forte piora no quadro econômico do estado no período. Dessa forma, abre-se uma relevante oportunidade para o estudo do modelo oficial usado na previsão do ICMS do Estado do Rio de Janeiro e de possíveis aprimoramentos. Isso será feito via comparação de especificações econométricas alternativas àquela proposta pela SEFAZ-RJ, a saber, modelos univariados, modelos multivariados e técnicas de combinação.

## **2 Revisão da Literatura**

Esta revisão de literatura está dividida em duas partes. Na primeira estão expostos trabalhos com objetivos semelhantes ao desta dissertação, isto é, comparar a acurácia de diferentes técnicas de previsão de arrecadação tributária. Já na segunda parte, as comparações metodológicas de previsão serão, parcialmente, deixadas de lado, para se privilegiar aspectos institucionais que influenciam o trabalho da previsão.

### **2.1 Comparação de Modelos de Previsão**

A literatura referente à previsão da arrecadação tributária no Brasil tem alguns de seus primeiros trabalhos no início dos anos 2000. Esses estudos comparavam o Método dos Indicadores, usado pela Secretária da Receita Federal do Brasil (RFB), com métodos econométricos univariados. Em geral, os resultados encontrados indicavam redução, em relação ao método oficial, dos erros de previsão quando se utilizava a metodologia univariada.

Peceguini (2001) compara métodos de previsão da arrecadação do ICMS anual e mensal no Estado de São Paulo para o ano de 2000, usando dados de 1995 a 1999 para estimar os modelos. Dos métodos univariados utilizados (Decomposição Clássica, Linha de Tendência e ARIMA), o da Decomposição Clássica apresentou, de forma geral, melhores previsões em relação ao modelo base de comparação, o chamado, pelos autores, Modelo Aritmético (que segue a mesma lógica do Método dos Indicadores).

Melo (2001) usou dois modelos univariados de previsão (Alisamento Exponencial de Holt-Winters Sazonal Aditivo e ARIMA) para predizer a arrecadação do Imposto de Renda (IR) no ano de 2000. Os resultados dos modelos univariados reduziram o erro de previsão médio em relação ao método oficial da Receita Federal.

Ampliando a análise, Siqueira (2002) fez a comparação entre os modelos univariados e o Método dos Indicadores levando em conta um conjunto de dez

tributos, sendo nove federais e o ICMS. Os dados para as estimações dos modelos foram de 1989 a 2000 e o ano de 2001 foi a base de comparação. Como resultado, encontrou-se que em oito das séries tratadas, a metodologia ARIMA se mostrou superior ao Método dos Indicadores. Nas duas restantes, o autor justificou a superioridade do método oficial dizendo que houve presença de quebras estruturais nas séries e alterações na legislação referente a esses dois tributos.

O trabalho de Guaranga e Mello (2002) estimou a arrecadação do ICMS no Estado do Rio Grande do Sul nos anos de 1999, 2000 e 2001, levando como base os dados de 1989 a 1998. Os métodos univariados usados apresentaram bons resultados em relação aos valores observados.

Um trabalho mais recente é o de Camargos (2008), que, através da metodologia Box-Jenkins, estimou modelos SARIMA para as séries dos tributos estaduais de São Paulo - ICMS, IPVA, ITCMD/ITBI e Taxas. Nesse trabalho não houve comparação com nenhuma metodologia oficial de previsão, mas apenas as estimações dos modelos.

Pessoa, Coronel e Lima (2013) analisaram a capacidade dos modelos ARIMA e ARFIMA em prever a arrecadação do ICMS do Estado de Minas Gerais. Não houve total predominância de um método sobre o outro, porém ambos os modelos tiveram bons resultados de previsão.

Azevedo, Silva e Gatsios (2017), ao observarem uma sequência de resultados não muito precisos da previsão do ICMS de alguns estados brasileiros, trabalharam em uma especificação alternativa – modelagem ARIMA – àquela utilizada pelos estados. Com uma amostra de seis unidades da federação (SP, MG, RJ, RS, PR e BA) e usando os anos de 2012 e 2013 como base de comparação, encontraram como resultado que o método ARIMA possui capacidade de gerar maior acurácia na previsão do ICMS em relação aos modelos oficiais dos estados.

Passando para os textos que se valeram de métodos multivariados, cita-se, inicialmente, Campos (2009) que, além dos métodos univariados, utilizou técnicas como VAR/VEC, equações simultâneas e modelos estruturais no exercício da previsão de três tributos federais arrecadados no âmbito da jurisdição do Estado de

São Paulo, a saber, Imposto de Importação (II), Imposto sobre a Renda das Pessoas Jurídicas (IRPJ) e Contribuição para o Financiamento da Seguridade Social (Cofins). A ideia básica era comparar os resultados de previsão dos métodos citados com o Método dos Indicadores. Os resultados apontaram para significativa redução do RMSE das previsões, especialmente no caso dos modelos multivariados.

No mesmo sentido, Benelli (2013) ampliou a análise, abordando todos os tributos federais, dividindo-os por base de incidência. Basicamente, os modelos multivariados usados foram o VAR e o VEC, além de métodos univariados (ARIMA) e de técnicas de combinação. Os resultados mostraram, assim como no trabalho de Campos (2009), superioridade em termos de previsão desses modelos que levam em conta várias variáveis na comparação com o método oficial da Receita Federal e com os modelos univariados.

Para as bases de incidência Renda e Folha de Salários, os melhores resultados vieram dos modelos multivariados, enquanto que para Bens e Serviços, Transações Financeiras e Outros Tributos a combinação via média simples das previsões uni e multivariados tiveram a melhor acurácia preditiva.

Ainda no âmbito dos modelos multivariados, porém agora tratando da previsão de tributos de nível estadual, a saber, o ICMS, tem-se o trabalho de Castanho (2011) que encontrou como resultado que as técnicas econométricas citadas apresentaram resultados mais acurados do que a metodologia utilizada pela Secretaria de Fazenda do Espírito Santo (SEFAZ-ES), que se vale de métodos mais simples como o da Linha de Tendência.

Corvalão (2002) já havia tratado do uso de métodos multivariados, modelo de regressão dinâmica, para a previsão do ICMS de Santa Catarina. Usando dados como faturamento da indústria, consumo de energia elétrica e consultas ao serviço de proteção ao crediário (SPC), o autor conseguiu reduzir os erros de previsão quando usou tais modelos na comparação com o método oficial da Secretaria de Fazenda de Santa Catarina (SEFAZ-SC) que segundo o autor se baseava em modelos univariados.

Um trabalho aplicado para o Estado do Rio de Janeiro foi o de (Lima, Ferreira e Honaiser, 2013), que analisou modelos uni e multivariados para a previsão da

receita corrente desse estado. Os modelos univariados utilizados foram do tipo SARIMA (*Seasonal Autoregressive Integrated Moving Average Model*) e estruturais, Já os multivariados foram: modelo BVAR (*Bayesian Vector Autoregressive Model*) e BVEC (*Bayesian Vector Error Correction Model*).

As séries de receitas analisadas foram: FECP, FPE, ICMS, IPI, IPVA, IRRF, ITCMD, SE, SUS e CIDE. Foi levado em conta na análise multivariada um conjunto de 16 séries referentes a diversos setores da atividade econômica, sendo a maioria delas concernentes ao ERJ e, as demais, ao Brasil e ao sudeste do país.

A comparação entre eles se deu da seguinte forma: foram feitas previsões para os 12 meses seguintes ao mês que foi feita a estimação, os erros obtidos foram somados e, assim, pôde-se obter um erro percentual anual. Esse procedimento foi repetido para todos os períodos a partir de certo ponto da amostra disponível até o final dela.

Na análise multivariada, as séries usadas nos modelos foram escolhidas com base nas significâncias estatísticas das variáveis explicativas. Com as variáveis escolhidas foram rodados os modelos multivariados (BVAR, BVEC e VEC) para todas as séries de receita e comparado seus erros anuais com os dos modelos univariados estruturais antes escolhidos.

Para a série do ICMS, o modelo multivariado usado se baseou nas seguintes séries: População (RJ), Produção Industrial (RJ), Comércio Varejista (RJ), Consumo de Energia Comercial (Sudeste) e Produção de Petróleo (RJ). Os resultados encontrados apontam para a superioridade do modelo univariado estimado em relação ao multivariado. Vale destacar que esse comportamento também foi observado na maioria das demais séries.

Refinando a análise, Mendonça e Medrano (2016) utilizaram técnicas de combinação de previsão para prever um total de nove tributos federais. As estimativas combinadas vieram dos seguintes modelos: Modelo Fatorial Dinâmico (MFD), SARIMA e o Método de Holt-Winters. Já as técnicas de combinação empregadas foram: combinação ótima, *performance*, regressão simples, média simples e mediana. Os resultados indicaram que, na maioria dos casos, as

previsões oriundas das combinações obtiveram resultados superiores às dos modelos individuais.

Kubo (2014) utilizou o método de otimização de carteira de investimento proposto por Markowitz para combinar as previsões dadas pelos modelos SARIMA, PLSR (Regressão com Mínimos Quadrados Parciais) e o Método dos Indicadores. O estudo foi feito com base na série do ICMS de São Paulo.

Outra metodologia de previsão foi utilizada por Mendonça e Medrano (2015). Tais autores aprimoram os trabalhos de Mendonça, Santos e Martins (2009) e de Mendonça, Medrano e Sachsida (2013). Isso foi feito com a estimação do Modelo Fatorial Dinâmico (MFD) bayesiano para a previsão de uma amostra de tributos representativos da carga tributária brasileira. Os resultados desse foram comparados com os do Modelo Linear Dinâmico, constatando-se a superioridade do primeiro.

Expandindo a análise para trabalhos realizados no exterior, pode-se destacar, de início, dois que usam índices estaduais de atividade econômica para prever arrecadação tributária estadual. O trabalho de Bram *et al.* (2004) foi um dos pioneiros no assunto. Esse usou os modelos VAR e VEC para analisar a capacidade do indicador de atividade econômica do Estado de Nova York prever a arrecadação tributária desse Estado. Os resultados encontrados indicam a melhora em termos de previsão quando se utiliza tais índices.

Balcilar *et al.* (2015) usam além dos métodos multivariados lineares já citados, alguns outros não lineares. Mas o foco principal do trabalho é o mesmo do outro citado acima. Assim, usando um índice de atividade da economia do estado de Nevada como base para os modelos multivariados, os autores encontram que o mesmo reduz consideravelmente os erros de previsão.

Favero e Marcellino (2005) realizaram previsões para uma série de variáveis fiscais, inclusive receita governamental, de países da zona do Euro. Os modelos utilizados foram ARMA, VAR, Equações Simultâneas e Combinação. De forma geral, os modelos ARMA e Combinação apresentaram os resultados mais acurados.

Willoughby e Guo (2008) compararam a precisão de diferentes técnicas de previsão usadas pelos estados dos Estados Unidos<sup>6</sup>. Os dados usados foram obtidos do *Government Performance Project* (GPP), um relatório de gestão governamental dos Estados Unidos.

Os autores encontraram como resultado principal que técnicas mais simples de previsão como, por exemplo, combinação de tendência linear com previsão consensual geraram, na maioria dos casos, resultados mais satisfatórios do que os de técnicas mais sofisticadas.

Williams e Kavanagh (2016) realizaram a previsão de 55 séries de receitas de diferentes estados dos Estados Unidos, usando métodos univariados de previsão. Entre os resultados encontrados, destaca-se que os métodos de tendência linear e de suavização exponencial apresentam os melhores resultados quando levado em conta dados nas formas anual e trimestral.

Além disso, foi visto que não houve melhora da previsão quando os dados foram passados da forma nominal para real e também não foi relevante outra especificação utilizada, a com os dados no formato anual.

Em suma, a literatura exposta acima, principalmente a brasileira, apontou para o ganho de acurácia preditiva das séries de tributos, no momento que se passou a empregar métodos econométricos de previsão, inclusive quando considerados os multivariados.

Pode-se dizer que as técnicas econométricas conseguiram reduzir os erros de previsão apresentados pelas metodologias oficiais da Secretaria da Receita Federal e das Secretarias de Fazenda mencionadas.

## **2.2 Práticas de Previsão**

Brender e Navon (2010), além de estimar um modelo multivariado para a arrecadação tributária de Israel, identificam as principais origens das incertezas do modelo elaborado. A principal conclusão foi que os erros de previsão foram

---

<sup>6</sup> Tendência Simples, Modelos de Série de Tempo, Regressão Linear, Julgamento, entre outras técnicas.

explicados especialmente pela má previsão das variáveis explicativas incluídas e não por uma possível má especificação do modelo construído.

Buettner e Kauder (2010) apontam os determinantes da acurácia preditiva das receitas de alguns países da OCDE. Incerteza sobre os fundamentos macroeconômicos, entendida como os erros de previsão das variáveis explicativas (PIB principalmente) do modelo, foi o principal ponto destacado. Composição da base tributária, período da realização da previsão e independência de manipulações do governo também se mostraram relevantes.

Na composição da base tributária, destaca-se que os erros foram maiores para a previsão de tributos sobre a renda e lucro das empresas. Também foi destacada que intervalos maiores entre o momento da previsão e a data prevista acarretam em maiores erros. Por fim, a independência do trabalho da previsão, via, por exemplo, participação de agentes externos ao governo, também aprimorou os resultados.

Costa (2011) buscou determinar os fatores institucionais que tivessem impacto sobre o exercício de previsão tributária dos estados do Brasil. Para isso, ele realizou uma regressão logística múltipla, tendo como variável dependente a chance do erro de previsão ser maior que 5,00% e como variável explicativa o autor utilizou algumas informações que representassem o contexto institucional das previsões no país.

Entre os resultados encontrados, destaca-se que a presença de forças opositoras no legislativo gerou redução das chances de erro acima de 5,00%. A ideia é que poderia haver maior facilidade para distorção dos resultados de previsão caso houvesse apenas um partido dominante com poderes sobre a previsão das receitas. Da mesma forma, participação de agentes externos ao governo também reduz as chances de erro de previsão.

Por outro lado, a existência de metas de arrecadação aumenta as chances de se observar erros acima de 5,00%. Segundo o autor, alguns estados possuem metas de arrecadação que condicionam parte dos salários dos servidores aos seus cumprimentos, podendo gerar viés sobre a previsão realizada.

Leal *et al.* (2007) fazem uma revisão da literatura do tema do impacto institucional/político sobre a acurácias das previsões de receitas de diversos países. As principais conclusões obtidas desses trabalhos foram as seguintes: há evidência de influência política sobre as previsões em alguns países europeus, enquanto que para os Estados Unidos os resultados não são tão conclusivos; a acurácia da previsão se reduz à medida que o intervalo de tempo da previsão para o período previsto aumenta; previsões de agências independentes tende a melhor as previsões.

## 3 Contexto Empírico

### 3.1 Arrecadação Tributária no Brasil

Nessa contextualização será feita, de princípio, uma breve descrição de algumas das principais informações referentes à carga tributária do Brasil no ano de 2015. Isso será feito dando maior ênfase aos tributos com base de incidência em bens e serviços, principalmente o ICMS que é a motivação central deste trabalho.

Exceto quando explicitado o contrário, os dados apresentados a seguir terão como fonte o documento “Carga Tributária no Brasil 2015: Análise por Tributos e Bases de Incidência”, divulgado pela RFB. Esse documento, divulgado anualmente, apresenta informações acerca da Carga Tributária Bruta (CTB) do Brasil, entendida como a soma do total da arrecadação da União, Estados e Municípios.

Dado que a publicação do documento da Receita Federal do Brasil ocorre no segundo semestre, a sua versão mais recente disponível é aquela divulgada no segundo semestre de 2016, referente ao ano de 2015. Logo, esse é o motivo pelo qual as informações aqui apresentadas são do ano de 2015.

Iniciando a análise, destaca-se que o total arrecadado em valores correntes pelo governo geral (União, Estados e Municípios) em 2015 foi R\$ 1.928.182.850,00, representando 32,66% do PIB brasileiro nesse ano<sup>7</sup>.

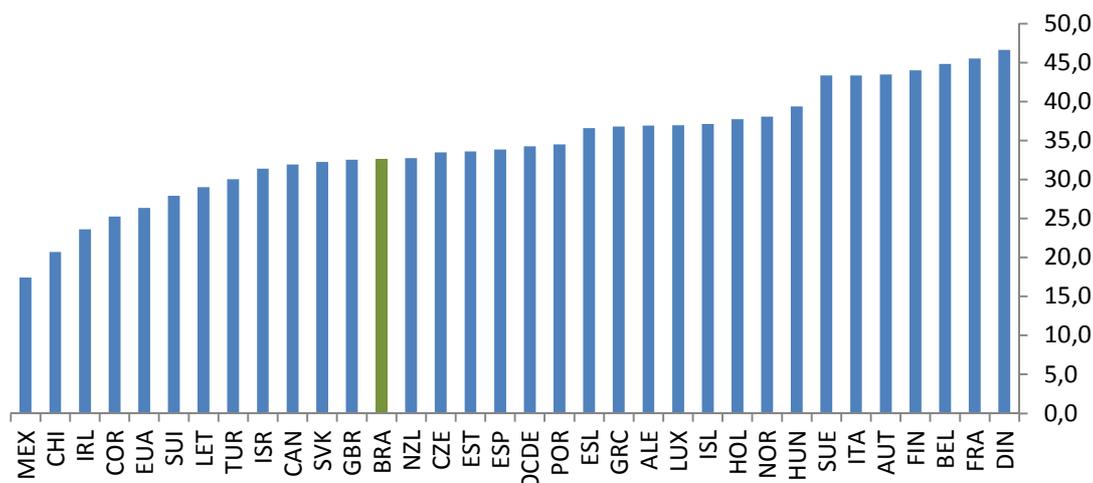
Na comparação com outros países, observa-se que, na análise da razão da carga tributária em relação ao PIB, o Brasil se apresenta abaixo da média dos países da OCDE (Organização para Cooperação e Desenvolvimento Econômico<sup>8</sup>), que é de 34,27%.

---

<sup>7</sup> Conforme dados do IBGE, o PIB brasileiro em 2015, em valores correntes, foi de aproximadamente R\$ 5.940,33 bilhões de reais.

<sup>8</sup> A OCDE é um fórum internacional que reúne boa parte das principais economias do mundo e tem como objetivo debater políticas de estímulo econômico e social. Atualmente, o Brasil não é membro da OCDE, mas é considerado com um importante parceiro da instituição, contribuindo em uma série de trabalhos.

Figura 1: Carga Tributária em Relação ao PIB – OCDE e Brasil - 2015:



Fonte: Elaboração Própria com dados da OCDE.

Na análise por entes da federação tem-se que a União foi responsável por 68,26% (R\$ 1.316.190,50 milhões) do total da arrecadação em 2015, enquanto que estados e municípios tiveram participação de, respectivamente, 25,37% (R\$ 489.103,22 milhões) e 6,37% (R\$ 122.889,13 milhões).

Analisando o total arrecadado por base de incidência, tem-se que em 2015 a base “Bens e Serviços” foi a mais importante entre as consideradas pela Receita Federal. Em relação ao total arrecadado e ao PIB, ambos de 2015, essa base de incidência teve participações de, respectivamente, 49,68% e 16,22%.

Tabela 1 - Arrecadação Governo Geral (2015) por Base de Incidência

Tipo de Base	Arrecadação (R\$ Milhões)	% PIB	% Arrecadação Total
Renda	R\$ 352.305,41	5,97%	18,27%
Folha de Salários	R\$ 498.034,80	8,44%	25,83%
Propriedade	R\$ 85.696,87	1,45%	4,44%
Bens e Serviços	R\$ 957.923,77	16,22%	49,68%
Transações Financeiras	R\$ 34.686,30	0,59%	1,80%
Outros	R\$ -464,30	-0,01%	-0,02%
<b>Total</b>	<b>R\$ 1.928.182,85</b>	<b>32,66%</b>	<b>100,00%</b>

Fonte: Elaboração Própria com dados da RFB.

Na publicação do seu trabalho referente à carga tributária de 2015, a Receita Federal apresenta os tributos que compõem a base de incidência “Bens e Serviços” em alguns grupos. Por exemplo, há uma diferenciação para gerais e seletivos, no

âmbito de um mesmo tributo. Enquanto gerais trata de produtos que recebem uma alíquota geral do tributo em questão, seletivos trata de produtos que recebem alíquotas específicas<sup>9</sup>.

Contudo, para fins de simplificação, abaixo será mostrada tabela com os tributos que compõem a base “Bens e Serviços” sem contar com tais diferenciações. Destaca-se que o tributo ICMS apresenta maior participação em relação ao total da base, tendo magnitude superior ao dobro da apresentada pelo segundo tributo mais relevante.

Tabela 2: Carga Tributária (2015) - Base de Incidência Bens e Serviços

Tributos	Arrecadação (R\$ Milhões)	% PIB	% Arrecadação Total
ICMS	R\$ 396.513,49	6,71%	20,56%
Cofins	R\$ 190.293,05	3,22%	9,87%
IPI	R\$ 47.607,74	0,81%	2,47%
ISS	R\$ 58.038,52	0,98%	3,01%
PIS	R\$ 39.825,13	0,67%	2,07%
CIDE	R\$ 3.271,18	0,06%	0,17%
Comércio Exterior	R\$ 38.969,36	0,66%	2,02%
Taxas	R\$ 48.436,19	0,82%	2,51%
Cont. Previdência	R\$ 22.156,65	0,38%	1,15%
Outros	R\$ 122.812,46	1,91%	5,85%
Total	R\$ 957.923,77	16,22%	49,68%

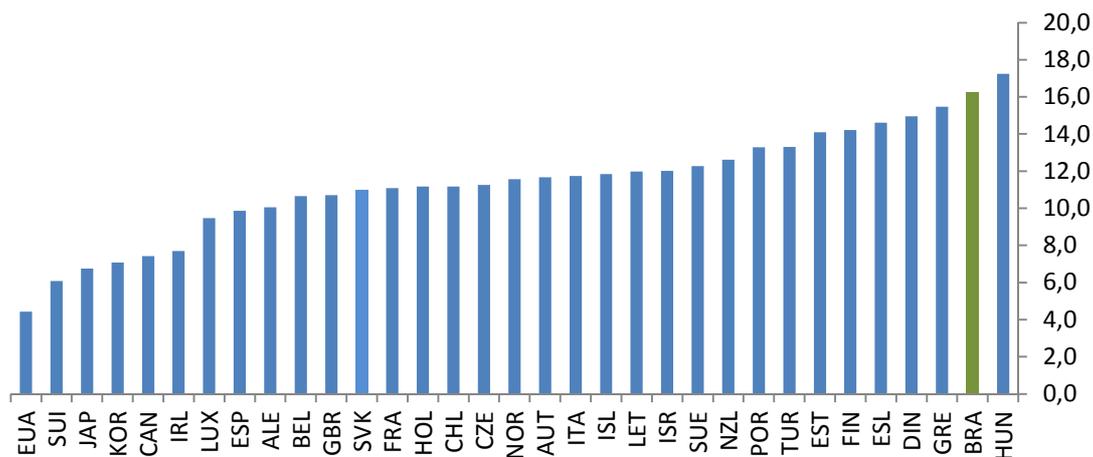
Fonte: Elaboração Própria com dados da RFB.

A relevância do ICMS pode ser observada não somente quando analisada no âmbito da base de “Bens e Serviço”. Individualmente, o ICMS é o tributo mais relevante entre todos aqueles que compõem a carga tributaria brasileira.

Quando comparado com os países da OCDE, a relevância da base de “Bens e Serviços” se torna ainda mais evidente. Nessa análise, o Brasil só fica atrás da Hungria (17,23%) em termos da comparação dessa base com o PIB do país.

<sup>9</sup> Por exemplo, os produtos automóveis, bebidas, combustíveis, energia elétrica, tabaco e telecomunicações possuem alíquotas específicas dentro do tributo ICMS.

Figura 2: Tributos com Base de Incidência Bens e Serviços (% PIB) - OCDE e Brasil:



Fonte: Elaboração Própria com dados da OCDE - 2015.

### 3.2 Receita Estado do Rio de Janeiro

Nesta parte será feita análise das receitas do governo do ERJ, dando ênfase para a participação do ICMS no total arrecadado. A fonte das informações citadas, exceto quando explicitado o contrário, será o Boletim de Transparência Fiscal, publicado pela Secretaria de Estado de Fazenda do Rio de Janeiro (SEFAZ-RJ).

A versão utilizada se refere ao último bimestre de 2015, que possui informações do acumulado desse ano. Em 2016, houve uma mudança na periodicidade da publicação do Boletim, deixando de ser bimestral para se tornar quadrimestral. Até o momento a versão referente ao último quadrimestre de 2016 ainda não havia sido publicada. Assim, optou-se pelo uso do Boletim do último bimestre de 2015.

No ano de 2015, o Estado do Rio de Janeiro teve uma receita de R\$ 79.086,90 milhões, sendo que 91,44% (R\$ 72.316,11 milhões) se referiam a receitas correntes e 8,56% (R\$ 6.770,79 milhões) a receitas de capital.

Conforme explicação apresentada no Manual de Procedimentos de Receitas Públicas do Tesouro Nacional:

Receitas operacionais são os ingressos de recursos financeiros oriundos das atividades operacionais, para aplicação em despesas correspondentes, também em atividades

operacionais, que não decorre de uma mutação patrimonial. (Manual de Procedimentos de Receitas Públicas do Tesouro Nacional, 2007, p.27).

Também conforme o Manual de Procedimentos de Receitas Públicas do Tesouro Nacional:

Receitas de Capital são derivadas da obtenção de recursos mediante a constituição de dívidas, amortização de empréstimos e financiamentos ou alienação de componentes do ativo permanente, constituindo-se em meios para atingir a finalidade fundamental do órgão ou entidade, ou mesmo, atividades não operacionais visando ao estímulo às atividades operacionais do ente. (Manual de Procedimentos de Receitas Públicas do Tesouro Nacional, 2007, p.27).

Na análise das receitas correntes, percebe-se que há relevante concentração da receita na base tributária, já que essa corresponde por 54,89% do total em 2015.

Tabela 3: Componentes da Receita Corrente (2015) – ERJ

Receita	Arrecadação (R\$ Milhões)	% Total Receita
Tributária	R\$ 43.414,67	54,89%
Contribuições	R\$ 1.914,27	2,42%
Patrimonial	R\$ 6.477,26	8,19%
Transferências	R\$ 5.931,51	7,50%
Intraorçamentária	R\$ 3.484,79	4,41 %
Agropecuária	R\$ 0,05	0,00%
Serviços	R\$ 358,97	0,45%
Indústria	R\$ 128,15	0,16 %
Outros	R\$ 10.606,43	13,41%
<b>Total</b>	<b>R\$ 72.316,10</b>	<b>91,44%</b>

Fonte: Elaboração Própria com dados da SEFAZ-RJ.

Tabela 4: Componentes da Receita de Capital (2015) – ERJ

Receita	Arrecadação (R\$ Milhões)	% Total Receita
Operações de Crédito	R\$ 4.994,87	6,32%
Alienações de Bens	R\$ 1.084,01	1,37%
Amortização de Empréstimos	R\$ 310,85	0,39%
Transferência de Capital	R\$ 380,60	0,48%
Intraorçamentária	R\$ 0,45	0,00 %
Outros	R\$ 0,00	0,00%
<b>Total</b>	<b>R\$ 6.770,78</b>	<b>8,56%</b>

Fonte: Elaboração Própria com dados da SEFAZ-RJ.

Dentro das receitas tributárias, também pode se observar uma característica de concentração, uma vez que o ICMS corresponde por 73,67% (R\$ 31.982,75 milhões) do total das receitas tributárias do Estado. Em relação ao total das receitas, o ICMS tem participação de 40,44%.

Tabela 5: Componentes da Receita Tributária (2015) – ERJ

Tributos	Arrecadação (R\$ Milhões)	% Tributária	% Total Receita
ICMS	R\$ 31.982,75	73,67%	40,44%
Adicional ICMS	R\$ 3.018,74	6,95%	3,82%
IRRF	R\$ 2.988,13	6,88%	2,91%
IPVA	R\$ 2.299,82	5,30%	3,78%
ITD	R\$ 849,91	1,96%	1,07%
Taxas	R\$ 2.275,32	5,24%	2,88%
<b>Total</b>	<b>R\$ 43.414,67</b>	<b>100,00%</b>	<b>54,89%</b>

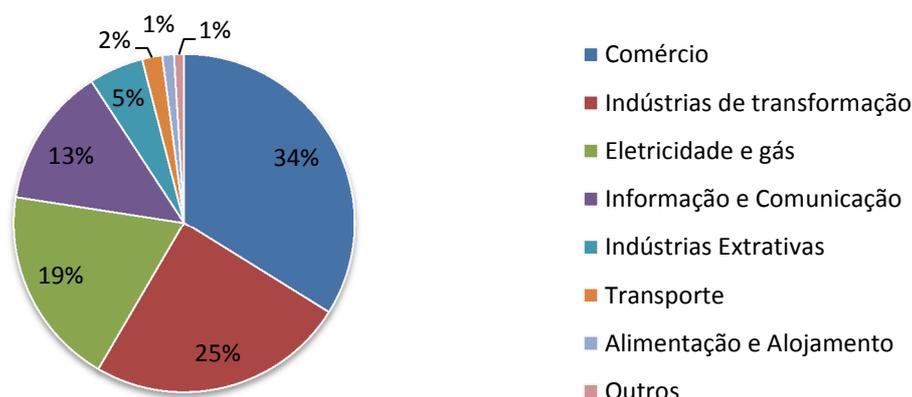
Fonte: Elaboração Própria com dados da SEFAZ-RJ.

### 3.3 ICMS no Estado do Rio de Janeiro

O ICMS é a principal fonte de arrecadação do Estado do Rio de Janeiro, correspondendo a 40,44% do total arrecadado em 2015 e a 73,67% da arrecadação tributária do estado nesse mesmo ano. O valor arrecadado no estado do Rio de Janeiro corresponde a cerca de 8,00% do total arrecadado desse tributo no país em 2015.

Na análise por base de incidência segundo CNAE 2.0 das atividades econômicas fluminenses, observa-se uma concentração da arrecadação em quatro setores (seções do CNAE): comércio, indústria de transformação, eletricidade e gás e informação e comunicação.

Figura 3: Base de Incidência do ICMS-RJ (2015) – Seções do CNAE:



Fonte: Elaboração Própria com dados da SEFAZ-RJ.

A característica da concentração não é tão claramente observada na análise dos subsetores que compõem cada uma das quatro principais seções citadas acima. Nas seções comércio e indústria de transformação, os subsetores mais relevantes representam, respectivamente, apenas 13,86% e 15,51% do total das seções.

Tabela 6: Seção Comércio CNAE – ICMS Estado do Rio de Janeiro - 2015

Subclasse CNAE	Arrecadação (R\$ Milhões)	% Total Comércio	% Total Receita
Álcool, Biodiesel e Gasolina	R\$ 1.487,36	13,86%	4,69%
Cigarros	R\$ 743,44	6,93%	2,35%
Vestuário e Acessórios	R\$ 625,38	5,83%	1,97%
Cosméticos e Perfumaria	R\$ 542,52	5,05%	1,71%
Medicamentos	R\$ 516,62	4,81%	1,63%
Outros	R\$ 6.819,74	63,53%	21,51%
<b>Total</b>	<b>R\$ 10.735,06</b>	<b>100,00%</b>	<b>33,87%</b>

Fonte: Elaboração Própria com dados da SEFAZ-RJ.

Tabela 7: Seção Indústria de Transformação CNAE – ICMS ERJ - 2015

Subclasse CNAE	Arrecadação (R\$ Milhões)	% Total Indústria	% Total Receita
Refino de Petróleo	R\$ 1.206,44	15,51%	3,81%
Cervejas	R\$ 823,49	10,59%	2,60%
Automóveis	R\$ 415,50	5,34%	1,31%
Medicamentos	R\$ 353,42	4,54%	1,11%
Outros	R\$ 4.978,28	64,01%	15,70%
<b>Total</b>	<b>R\$ 7.777,12</b>	<b>100,00%</b>	<b>24,53%</b>

Fonte: Elaboração própria com dados da SEFAZ-RJ.

Tabela 8: Seção Eletricidade e Gás CNAE – ICMS Estado do Rio de Janeiro - 2015

Subclasse CNAE	Arrecadação (R\$ Milhões)	% Total Eletricidade	% Total Receita
Distribuição de Energia Elétrica	R\$ 4.495,58	74,41%	14,18%
Geração de Energia Elétrica	R\$ 733,60	12,14%	2,31%
Produção de Gás	R\$ 350,34	5,80%	1,11%
Comércio de Energia Elétrica	R\$ 295,01	4,88%	0,93%
Outros	R\$ 167,06	2,77%	0,53%
<b>Total</b>	<b>R\$ 6.041,60</b>	<b>100,00%</b>	<b>19,06%</b>

Fonte: Elaboração Própria com dados da SEFAZ-RJ.

Tabela 9: Seção Inf. e Comunicação CNAE – ICMS Estado do Rio de Janeiro - 2015

Subclasse CNAE	Arrecadação (R\$ Milhões)	% Total Inf. e Co	% Total Receita
Telefonia Móvel Celular	R\$ 1.716,63	40,51%	5,42%
Telefonia Fixa	R\$ 1.652,19	38,99%	5,21%
Serviço Móvel Especializado	R\$ 455,64	10,75%	1,44%
Outros	R\$ 412,95	9,75%	1,31%
<b>Total</b>	<b>R\$ 4.237,41</b>	<b>100,00%</b>	<b>13,37%</b>

Fonte: Elaboração Própria com dados da SEFAZ-RJ.

## **4 Metodologia e Dados**

### **4.1 Metodologia**

#### **4.1.1 Procedimentos da Previsão**

##### **4.1.1.1 Ideia Geral**

O objetivo deste trabalho é comparar a acurácia preditiva, para a previsão da série do ICMS do ERJ, de três especificações econométricas distintas, a saber, modelos univariados, modelos multivariados e combinação de modelos de previsão.

Tal comparação será realizada especialmente com base nas estatísticas de erro calculadas para cada modelo, mas também pela análise estrutural dos mesmos que se dará com a avaliação dos resíduos das estimações.

##### **4.1.1.2 Modelos Utilizados**

Mais à frente serão detalhadamente descritos cada um dos modelos utilizados neste trabalho. De forma preliminar, destaca-se que os univariados serão o SARIMA (*Sazonal Autoregressive Moving Average*) e o de Holt-Winters. Os multivariados serão VAR ou VEC, a depender dos resultados das análises de cointegração das séries. Por fim, as técnicas de combinação utilizadas serão: Média Simples, *Least Squares Weights*, *Mean Square Error (MSE) Weights* e *Mean Square Error (MSE) Rank*. As combinações serão feitas com os melhores modelos univariados e multivariados (um de cada).

##### **4.1.1.3 Periodicidades da Estimação e Previsão**

Os modelos serão estimados com dados de janeiro de 2003 até junho de 2015 e as previsões serão feitas para os meses de julho de 2015 até dezembro de

2016. Contudo, as estatísticas de erros serão calculadas tomando como base apenas as previsões mensais para os 12 meses de 2016.

O motivo para a escolha de junho de 2015 para a realização da previsão é a tentativa de simular o trabalho de fato realizado pela SEFAZ-RJ. Conforme exposto no capítulo 1, a previsão para a LDO ocorre, em média, no final de março (com dados até janeiro devido à defasagem da divulgação de algumas séries usadas no modelo) e a da LOA, em média, no começo de agosto (com dados até junho pelo mesmo motivo citado para a LDO).

Nota-se que o período usado neste trabalho reflete o caso da LOA. A razão para essa escolha se deve ao fato de que o valor considerado na elaboração do orçamento estadual é o apresentado na LOA, conforme definido no Anexo de Compatibilização.

Também é válido ressaltar que as estatísticas dos erros de previsão serão calculadas apenas com as estimativas para os doze meses de 2016, apesar de que, conforme já citado, a previsão se inicie já em julho de 2015. A razão é que os meses de 2015 não têm nenhuma importância para o orçamento de 2016, mas apenas são feitos porque a previsão para 2016 é realizada em meados de 2015 e não no último mês do ano.

#### **4.1.1.4 Metodologia de Previsão**

Conforme Hyndman e Athanasopoulos (2013, seção 2/3), o exercício da previsão deve ser iniciado com a divisão da amostra em duas partes: *“training set”* ou *“in-sample data”* e *“test set”* ou *“out-of-sample data”*. A parte *“training-set”* da amostra é aquela na qual o modelo é estimado com os dados observados das variáveis utilizadas, enquanto que o *“test set”* representa a parte da amostra na qual serão realizadas as previsões. No caso deste trabalho, o *“training set”* vai de janeiro de 2003 até junho de 2015 e o *“test set”* vai de julho de 2015 até dezembro de 2016.

As previsões aqui realizadas estarão baseadas na metodologia conhecida como Previsão Dinâmica (*Dynamic Forecast*), na qual a previsão para um período a

frente não utiliza os valores observados das defasagens das séries usadas, mas sim os valores previstos pelo próprio modelo. Alternativa a essa especificação é a Previsão Estática (*Static Forecast*), aquela que toma como base os valores observados da série. Obviamente, o trabalho da SEFAZ-RJ é caracterizado como Previsão Dinâmica.

As previsões aqui consideradas foram feitas no software Eviews 9, que permite algumas caracterizações adicionais, além da escolha pelo método da Previsão Dinâmica. Por exemplo, pode-se escolher entre Determinístico ou Estocástico para o tipo de simulação. A diferença entre os dois é que na estocástica é incluído um termo de erro aleatório no modelo, usado junto com as defasagens das variáveis na previsão. Este trabalho optou pela simulação Estocástica.

A escolha da simulação Estocástica permite uma caracterização adicional, dada pela seleção da opção "*Include Coefficient Uncertainty*", que faz com que os parâmetros estimados sejam randomicamente alterados no início de cada previsão. Essa opção também foi incluída neste trabalho.

#### **4.1.1.5 Comparação dos Modelos**

A comparação dos modelos será feita com base tanto nas estatísticas de erro de previsão, quanto na solidez do modelo, medida em termos da análise dos seus resíduos. A ideia é que determinado modelo pode apresentar resultado muito bom na previsão, mas não possuir a mesma qualidade na análise dos seus resíduos, de modo que a confiabilidade no mesmo para o exercício de futuras previsões fica abalada.

A análise dos resíduos se concentrará em três pontos principais: autocorrelação serial, homocedasticidade e normalidade. Conforme Hyndman e Athanasopoulos (2013, seção 2/6), o caso mais grave seria a observação de autocorrelação, uma vez que esse seria um indício de que há informações deixadas nos resíduos que poderiam ser úteis na previsão.

Já a não observação da homocedasticidade acarretaria em variância do estimador via MQO dos parâmetros não sendo mais eficiente e a não normalidade faria com que alguns testes estatísticos pudessem ficar comprometidos em pequenas amostras.

Já as estatísticas de erro de previsão utilizadas serão: Média dos Erros ao Quadrado (MSE), Raiz da Média dos Erros ao Quadrado (RMSE), Média dos Erros Absolutos (MAE), Média dos Erros Absolutos Percentuais (MAPE) e Erro Percentual Anual (% Anual).

Tanto as estatísticas de erro de previsão quanto os testes dos resíduos serão explicados com maiores detalhes a seguir junto com a descrição dos modelos univariados e multivariados.

## **4.1.2 Modelos Utilizados**

### **4.1.2.1 Modelos Univariados**

#### **4.1.2.1.1 SARIMA**

O modelo univariado mais usado na previsão de séries econômicas é o ARMA (*Autoregressive Moving Average*). Esse modelo leva em conta nas suas previsões os valores observados no passado da série em estudo (parte autoregressiva) e as observações também no passado dos erros de previsão.

A parte autoregressiva é dada por uma combinação linear dos valores observados da série em estudo. Isto é, a regressão que é feita nesse caso é da variável em consideração contra ela mesma no passado. A formulação básica de um AR(p), onde p significa o número de termos autoregressivos pode ser representada por:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (4.1)$$

Já a parte média móvel do ARMA leva em conta os erros de previsão observados no passado. A formulação básica de um MA(q), onde q indica o número de termos de média móvel é representada abaixo:

$$Y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} \quad (4.2)$$

A combinação dos dois modelos citados acima, AR(p) e MA(q), é o ARMA(p,q):

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t \quad (4.3)$$

Destaca-se que a utilização da representação ARMA é válida apenas em casos nos quais a série temporal em questão seja estacionária. Basicamente, uma série será considerada como estacionária quando a mesma não tiver suas propriedades estatísticas dependentes do tempo. Conforme Enders (1995, p.69) um processo estocástico é dito covariância-estacionário ou estacionário-fraco quando são válidas as seguintes restrições para todas as datas t e t-s:

$$E(y_t) = E(y_{t-s}) = \mu \quad (4.4)$$

$$E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2] = \sigma_y^2 \quad [Var(y_t) = Var(y_{t-s}) = \sigma_y^2] \quad (4.5)$$

$$E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s [Cov(y_t, y_{t-s})] \\ = Cov(y_{t-j}, y_{t-j-s}) \quad (4.6)$$

Resumindo o apresentado acima, uma série de tempo será considerada covariância-estacionária quando sua média e variância forem iguais para qualquer data t e todas as autocovariâncias dependerem apenas da distancia temporal entre as observações e não da data t em questão.

Para séries que não forem estacionárias, a simples utilização do modelo ARMA pode não ser a mais adequada. Sob tal especificação, as séries poderiam apresentar problemas como, por exemplo, o fato de que os choques poderiam ser não temporários, de modo que a série não teria reversão para sua média de longo-prazo (Enders, 1995, p.212)

A forma mais tradicional de tratamento de uma série não estacionaria é realizar a sua diferenciação, que simplesmente é feita via a subtração do valor da

série em certa data pelo valor na data imediatamente anterior. Esse tipo de transformação é mais eficiente para tratar a não estacionariedade da média, uma vez que corrige problemas ligados à tendência e a sazonalidade.

Outra transformação útil para o objetivo citado é a passagem do logaritmo na mesma. Isso é mais relevante quando se deseja estabilizar a variância da série. Os testes mais tradicionais para a estacionariedade serão apresentados a seguir.

O teste mais usual para a estacionariedade de uma série de tempo é o de Dickey-Fuller. A partir de uma equação,  $y_t = \alpha_1 y_{t-1} + \varepsilon_t$ , tem-se que quando  $\alpha_1 = 1$ , a série em questão será um passeio aleatório e, assim, não estacionária. Antes de partir para o teste, são feitas transformações na equação antes citada. Deve-se subtrair cada lado da equação por  $y_{t-1}$ , de modo a obter  $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$ , onde  $\gamma = \alpha_1 - 1$ .

A hipótese nula testada no modelo é que  $\gamma = 0$ , equivalente a dizer que  $\alpha_1 = 1$ . Se a hipótese nula não for rejeitada, não se nega a hipótese da série ter uma raiz unitária e, assim, não ser estacionária. A estimativa do parâmetro usado no teste se dá via Mínimos Quadrados Ordinários (MQO). Com a estatística de teste obtida pode se comparada com o valor tabelado na tabela de Dickey-Fuller.

Vale destacar que esse teste pode ser realizado via diferente especificações, dependendo dos componentes determinísticos incluídos no modelo. Nesse sentido, há três casos básicos para o teste: sem tendência e constante, apenas com constante e com constante e tendência:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t \quad (4.7)$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \varepsilon_t \quad (4.8)$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_1 t + \varepsilon_t \quad (4.9)$$

Um problema relacionado à especificação apresentada acima é que, por simplificação, se supõe que o erro do teste de Dickey-Fuller é um ruído branco, contudo, isso não é necessariamente verdade. Quando os erros não forem ruído branco, se faz necessário realizar transformação no teste original via o acréscimo de

termos autoregressivos no modelo. Esse é o chamado teste de Dickey-Fuller Aumentado (ADF). As especificações apresentadas acima agora têm:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (4.10)$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (4.11)$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_1 t + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (4.12)$$

O teste ADF demanda a escolha da ordem dos termos autoregressivos a serem incluídos no modelo. Outra Alternativa para o caso dos resíduos não serem ruído branco é o teste de Phillips Perron (PP), que tem a vantagem de não demandar a especificação da ordem dos termos autoregressivos a serem incluídos. Suas equações são idênticas ao do teste de Dickey-Fuller.

Mais um teste de estacionariedade é o de Kwiatkowski, Phillips, Schmidt e Shin (KPSS). Diferente dos dois testes antes citados, o KPSS tem como hipótese nula a estacionariedade do processo em questão.

A metodologia Box-Jenkins (Box e Jenkins, 1976) se baseia na realização de previsões usando modelos ARIMA(p,d,q). Sua formatação está dividida em distintos passos: identificação, estimação e testes de adequação.

O primeiro passo do processo é o da identificação que pode ser feita via duas definições básicas: i) a série é estacionária e, conseqüente, definição do valor do componente d e; ii) quais termos AR e MA devem ser incluídos no modelo.

A primeira questão já foi discutida acima com os testes de estacionariedade. Depois de feita tal análise e, se for o caso, realizada a diferenciação na série, parte-se para a análise do segundo ponto. Isso é feito via análise do correlograma da série, que representa suas funções de autocorrelação (FAC) e autocorrelação parcial (FACP).

Conforme modelagem apresentada por Enders (1995, p.78), a obtenção da FAC se dá via a divisão das autocovariâncias  $\gamma_s$  pela autocovariância de ordem zero, que é a variância,  $\gamma_0$ . Por exemplo, para o caso de um ARMA(1,0),  $y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ , tem-se que a sua variância,  $\gamma_0$ , é dada por  $\gamma_0 = \sigma^2 / [1 - (\alpha_1)^2]$ . Já suas autocovariâncias,  $\gamma_s$ , são dadas por  $\gamma_s = \sigma^2 (\alpha_1)^s / [1 - (\alpha_1)^2]$ .

Então dividindo cada  $\gamma_s$  por  $\gamma_0$  se obtém as  $\rho_s$ , isto é, as autocorrelações do processo AR(1) em questão. O exemplo dado, como dito, trata do caso AR(1), mas a mesma ideia vale para todos os outros processos possíveis do modelo ARIMA.

A FAC apresenta a relação entre observações da série em estudo que possuem a mesma distância temporal sem que se descontem os efeitos indiretos causados pelas observações intermediárias. Nesse sentido, a FACP surge como alternativa, uma vez que trata desses efeitos indiretos. O que a FACP apresenta de novidade são autocorrelações parciais,  $\phi_{ss}$ .

Uma parte do correlograma é dada pela plotagem de  $\rho_s$  contra  $s$ , isto é, da apresentação de cada autocorrelação para cada distancia temporal  $s$ . O correlograma é complementado com a plotagem dos  $s$  contras as  $\phi_{ss}$ .

A utilidade do correlograma para a identificação da série estudada no processo Box-Jenkins em questão se dá pelo fato de que os processos em verificação (AR(1), MA(1), ARMA(1,1)...) apresentam padrões definidos para as suas FAC e FACP.

Tabela 10: Propriedades FAC e FACP – Alguns Casos:

Processo	FAC	FACP
AR(1): $\alpha_1 > 0$	Decaimento Exponencial: $\rho_s = \alpha_1^s$	$\phi_{11} = \rho_1$ ; $\phi_{ss} = 0$ para $s \geq 2$
AR(1): $\alpha_1 < 0$	Decai Oscilando: $\rho_s = \alpha_1^s$	$\phi_{11} = \rho_1$ ; $\phi_{ss} = 0$ para $s \geq 2$
AR(p)	Decai para Zero. Coeficientes podem Oscilar	Picos até lag p. Demais $\phi_{ss} = 0$
MA(1): $\beta > 0$	Pico Positivo no lag 1 e $\rho_s = 0$ para os Demais	Decai Oscilando: $\phi_{11} > 0$
MA(1): $\beta < 0$	Pico Negativo no lag 1 e $\rho_s = 0$ para os Demais	Decai: $\phi_{11} < 0$
ARMA(p,q)	Decai Direto ou Oscilando a partir do lag q	Decai Direto ou Oscilando a partir do lag p

Fonte: Enders (1995, p.85).

Obviamente, apenas a inspeção visual do correlograma não é algo totalmente confiável para a escolha da melhor especificação da série em estudo. Um auxílio se dá via utilização de critérios de informação. Uma regra de bolso nesse sentido é a seguinte: através da análise da FAC e da FACP escolher valores  $p$  e  $q$  máximos, então se estima todos os modelos possíveis com  $0 \leq p \leq p_{\max}$  e  $0 \leq q \leq q_{\max}$ . Depois, deve utilizar os critérios de informação para selecionar os modelos mais competitivos entre os considerados (Martin, Hurn e Harris, 2012, p.507).

A ideia básica dos critérios de informação é classificar os modelos com base em dois pontos fundamentais: i) uma função da soma dos quadrados dos resíduos e ii) um termo que penalize a perda de graus de liberdade quando da inclusão de novas variáveis no modelo. A ideia é que a inclusão de variáveis adicionais no modelo pode gerar redução da soma dos quadrados dos resíduos, porém, ao custo de perda de graus de liberdade.

Os modelos de critério de informação mais comuns são o Critério de Informação de Akaike (AIC), o de Critério de Informação Bayesiano de Schwartz (SBC) e o Critério de Informação de Hannan-Quinn (HQ). Suas representações seguem abaixo (Lutkepohl, 2004, p.33 e 34), destacando que  $n$  representa o número de parâmetros estimados ( $p + q +$  um termo de constante se for o caso) e  $T$  é o número total de observações usadas no cálculo:

$$AIC(n) = \log \tilde{\sigma}_u^2(n) + \frac{2}{T}n \quad (4.13)$$

$$SBC(n) = \log \tilde{\sigma}_u^2(n) + \frac{\log T}{T}n \quad (4.14)$$

$$HQ(n) = \log \tilde{\sigma}_u^2(n) + \frac{2 \log \log T}{T}n \quad (4.15)$$

Destaca-se que  $\tilde{\sigma}_u^2(n) = T^{-1} \sum_{t=1}^T \hat{u}_t(n)^2$  representa o estimador da variância dos erros baseado nos resíduos  $\hat{u}_t(n)$  obtidos na estimação via Mínimos Quadrados Ordinários (MQO) do modelo tratado.

O segundo passo da metodologia Box-Jenkins é o da estimação. O método mais usual para a estimação de modelos ARIMA é o da Maximização da Função de Verossimilhança. Conforme Hamilton (1994), para esse método, primeiro deve-se

especificar a função de densidade de probabilidade a ser usada, assumindo que os termos de erro tenham distribuição normal. Então, o processo de estimação começa com a derivação função de verossimilhança e então são obtidos os valores dos parâmetros do vetor  $\Theta$  de modo a maximizar o valor da função citada.

O terceiro passo da metodologia Box-Jenkins se baseia no diagnóstico dos resíduos. Espera-se que tais resíduos sejam ruído branco. Caso isso não se verifique, há o risco de surgimento de problemas ligados a viés nos parâmetros estimados e a variância de tais estimações (questão da eficiência<sup>10</sup>) que podem interferir na qualidade das previsões consideradas. Nesse sentido, destacam-se as seguintes propriedades estatísticas:

- Ausência de autocorrelação serial entre os resíduos,  $E\{\varepsilon_t \varepsilon_{t-s} | X\} = 0$  para qualquer  $t, s$ ;
- Homocedasticidade,  $Var\{\varepsilon_t | X\} = \sigma^2 = constante$ ;
- Normalidade:  $\varepsilon_t | X \sim N(0, \sigma^2)$ .

Para o caso da autocorrelação, uma forma de teste se dá via o teste Box-Pierce (Portemanteau Q-Test). A ideia desse teste é realizar a estimação do modelo em análise e com os resíduos obtidos calcular suas autocorrelações para cada distancia temporal  $k$ , isto é,  $r_k$ . Com tais autocorrelações é possível obter uma estatística de teste dada por  $Q = T \sum_{k=1}^h r_k^2$ , onde  $h$  indica a quantidade de termos de autocorrelação considerados e  $T$  representa o número total de observações. A hipótese nula é que a série é não autocorrelacionada.

Outro teste de autocorrelação é o de Breusch-Godfrey (LM Test). Esse teste realiza a estimação do modelo em estudo, obtendo-se os resíduos da regressão. Com esses resíduos uma nova regressão seria feita colocando o resíduo na data  $t$  em função dos resíduos nas datas passadas. Com isso, se obteria um coeficiente de determinação  $R^2$ , que seria usado na estatística de teste do modelo, da dada por  $TR^2$ , onde  $T$  representa o número de observações usadas no modelo. A hipótese nula é que a série é não autocorrelacionada.

---

<sup>10</sup> Isto é, o estimador pode não apresentar a variância mínima.

Para a hipótese da normalidade o teste usado é o de Jarque-Bera. Esse teste leva em conta dois componentes básicos: a assimetria e a curtose. A estatística de teste é dada por  $JB = \frac{T}{6} \left( \hat{S}^2 + \frac{(K-3)^2}{4} \right)$ , onde T indica o total de observações consideradas, S a assimetria da distribuição dos resíduos e K a curtose de tal distribuição. Para que a hipótese nula de que há normalidade seja aceita, S deve tender a 0 e K deve tender a 3, valores que caracterizam uma distribuição normal.

Por fim, na análise da homocedasticidade da série em estudo, será feita usando teste de White. Esse teste é bastante similar ao teste de Breusch-Godfrey anteriormente apresentado para a autocorrelação serial. A diferença básica é que agora os erros seriam considerados elevados ao quadrado, de modo a atentar para a variância. Sua estatística de teste seria dada por  $TR^2$ , sendo T o tamanho da amostra e  $R^2$  o coeficiente de determinação da regressão dos resíduos ao quadrado contra suas defasagens. A hipótese nula é de que a série é homocedástica.

Outro componente importante que pode ser introduzido na análise Box-Jenkins é o da sazonalidade. Em muitas séries econômicas os valores observados em certas datas sofrem grande influência de fatores sazonais, de modo que se torna relevante a introdução da versão sazonal do modelo ARIMA, o SARIMA (*Sazonal Autoregressive Moving Average*).

O SARIMA pode ser representado por  $ARIMA(p,d,q)(P,D,Q)_T$ , onde os termos p,d,q representam a parte não sazonal do modelo e P,D,Q representam a parte sazonal. Já T indica a periodicidade dos dados.

A análise da sazonalidade pode entrar já na primeira etapa da metodologia Box-Jenkins, a da identificação. Isto seria feito analisando a FAC e a FACP da série original sem nenhuma diferenciação, contudo ao invés de olhar-se para os primeiros termos AR e MA, dar-se-ia atenção para tais termos nas suas versões sazonais, isto é, a partir da observação 12 (dado que os dados são mensais). Caso observada a existência de sazonalidade, deveria ser feita uma diferenciação, tal como dito, antes, só que dessa vez sazonal, subtraindo as observações da 12ª defasagem.

A partir de então, voltaria para a análise antes apresentada, isto é, para as primeiras defasagens da FAC e da FACP para ver se a série é estacionária ou não e, se for o caso, realizar nova diferenciação simples e prosseguir com a análise.

#### 4.1.2.1.2 Modelo de Holt-Winters

Antes de abordar especificamente o modelo de Holt-Winters, será feita uma breve revisão do método de Alisamento Exponencial de forma geral, destacando suas principais variações. Inicialmente será exposto o “Método de Alisamento Exponencial Simples” (SES na sigla em inglês<sup>11</sup>). Depois serão introduzidos dois novos que levam em conta a tendência da série em estudo (“Método de Tendência Linear de Holt” e “Método de Tendência Exponencial”). Por fim, será destacado o “Método de Holt-Winters”, que tem como principal diferença em relação aos demais o tratamento dado às questões sazonais da série em estudo.

A característica principal desses tipos de modelo é a realização das previsões com base em uma média ponderada das observações passadas, sendo que o peso para cada observação se reduz à medida que a distância temporal cresce.

A versão mais simples do método de Alisamento Exponencial é a chamada “Alisamento Exponencial Simples”. Esse caso é mais apropriado para séries que não apresentem tendência e sazonalidade, já que esses componentes não são levados em consideração na análise.

Conforme Hyndman e Athanasopoulos (2013, seção 7/1), esse modelo pode ser entendido como uma combinação de dois métodos mais simples de previsão: i) o método *naïve*, considerado como a previsão para um período à frente sendo igual à do período anterior e; ii) o método da média simples das previsões passadas, no qual o valor previsto para um período à frente é dado por uma média simples dos valores já observados. Sua formulação é dada por:

---

<sup>11</sup> “Simple Exponential Smoothing”.

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots \quad (4.16)$$

onde  $0 \leq \alpha \leq 1$  é o *parametro de suavização*

Percebe-se que  $\alpha$ , o parâmetro de suavização, determina a taxa pela qual o peso das observações mais antigas irá se reduzir ao longo do tempo.

Uma forma alternativa de apresentar o “Método de Suavização Exponencial Simples” se dá através da decomposição de seus componentes. No caso apresentado, há apenas um componente, o de nível  $l_t$ , além da equação de previsão:

$$\text{Equação de Previsão:} \quad \hat{y}_{t+1|t} = l_t \quad (4.17)$$

$$\text{Equação de Suavização:} \quad l_t = \alpha y_t + (1 - \alpha)l_{t-1} \quad (4.18)$$

Para que o modelo descrito acima seja factível, deve-se estimar definir o termo  $l_0 = \hat{y}_{1|0}$ , isto é, a previsão na data zero para a variável na data 1. Como tal termo não pode ser definido com base nos dados disponíveis, ele deverá ser definido de outra maneira.

A definição desse termo pode se dar de forma arbitrária, fazendo  $l_0 = y_1$ . Alternativamente, ele poderia ser definido com a utilização de técnicas de otimização. Da mesma forma, o parâmetro de suavização,  $\alpha$ , também pode ser definido via otimização<sup>12</sup>. A forma mais tradicional de realizar as otimizações citadas e encontrar os dois termos citados é o da minimização do somatório do quadrado dos resíduos, dados por  $e_t = y_t - \hat{y}_{t|t-1}$ , para  $t = 1, 2, 3, \dots, T$ .

A primeira versão modificada do modelo simples de suavização exponencial é o “Método de Tendência Linear de Holt”. Basicamente, esse método acrescenta um componente de tendência linear ao método simples. Sua representação na forma de componentes pode ser descrita, além da equação de previsão, por mais duas equações, a de nível e a de tendência:

$$\text{Equação de Previsão:} \quad \hat{y}_{t+1|t} = l_t + hb_t \quad (4.19)$$

$$\text{Equação de Nível:} \quad l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (4.20)$$

<sup>12</sup> Em algumas situações esse parâmetro é definido de forma subjetiva, via, por exemplo, experiências anteriores.

$$\text{Equação de Tendência: } b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (4.21)$$

Os parâmetros que já apareciam no método simples mantém o mesmo significado nessa nova versão. Em relação aos novos termos, tem-se que  $b_t$  representa a estimativa da tendência da série na data  $t$ . Já o parâmetro  $\beta$  representa o parâmetro de suavização da tendência, sendo que  $0 \leq \beta \leq 1$ . Destaca-se que  $h$  indica o número de períodos à frente que a estimação está sendo realizada.

Outra maneira de se acrescentar um componente de tendência ao modelo simples é o chamado “Método de Tendência Exponencial”. Assim como no método de tendência de Holt, sua decomposição é composta por uma equação de nível, uma de tendência e outra de previsão:

$$\text{Equação de Previsão: } \hat{y}_{t+1|t} = l_t b_h^t \quad (4.22)$$

$$\text{Equação de Nível: } l_t = \alpha y_t + (1 - \alpha)(l_{t-1} b_{t-1}) \quad (4.23)$$

$$\text{Equação de Tendência: } b_t = \beta \left( \frac{l_t}{l_{t-1}} \right) + (1 - \beta)b_{t-1} \quad (4.24)$$

Finalmente, o “Método Sazonal de Holt-Winters” aprimorou os métodos já citados via incorporação de mecanismo que incluísse a sazonalidade dentro do modelo de previsão. Esse mecanismo pode ser introduzido no modelo de duas maneiras: aditiva e multiplicativa. A aditiva deve ser empregada quando as variações sazonais mantêm um padrão constante ao longo da série, enquanto que a versão multiplicativa deveria ser usada quando houvesse mudanças na sazonalidade proporcionais às mudanças de nível da série (Hyndman e Athanasopoulos, 2013, seção 7/5).

A versão aditiva do “Método Sazonal de Holt-Winters” pode ser representada na forma de decomposição em equações de nível, tendência e sazonalidade, além da equação de previsão:

$$\text{Equação de Previsão: } \hat{y}_{t+1|t} = l_t + h b_t + s_{t-m+h_m} \quad (4.25)$$

$$\text{Equação de Nível: } l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} b_{t-1}) \quad (4.26)$$

$$\text{Equação de Tendência: } b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (4.27)$$

$$\text{Equação de Tendência: } s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (4.28)$$

A interpretação para a equação de nível agora é que ela é dada por uma média ponderada da observação na data t ajustada sazonalmente com a previsão sem ajuste sazonal para essa mesma data. Já a equação de tendência tem exatamente a mesma representação dos métodos anteriormente citados. Por fim, a nova equação, a de sazonalidade indica uma média ponderada entre o índice de sazonalidade na data t com esse mesmo índice na data t-m, onde m é definido pela periodicidade dos dados.

Já a versão multiplicativa do Método de Holt-Winters, pode ser apresentada na decomposição de seus componentes da seguinte maneira:

$$\text{Equação de Previsão: } \hat{y}_{t+1|t} = (l_t + hb_t)s_{t-m+h_m} \quad (4.29)$$

$$\text{Equação de Nível: } l_t = \alpha \left( \frac{y_t}{s_{t-m}} \right) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (4.30)$$

$$\text{Equação de Tendência: } b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (4.31)$$

$$\text{Equação de Tendência: } s_t = \gamma \frac{y_t}{(l_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m} \quad (4.32)$$

#### 4.1.2.2 Modelos Multivariados

##### 4.1.2.2.1 VAR/VEC

O trabalho da previsão pode ser aprimorado com o uso de modelos multivariados, aqueles que levam em conta na previsão, além da própria série em estudo, outras que tenham influência sobre a mesma. O modelo mais tradicional nesse sentido é o VAR (Vetores Autorregressivos), que tem uma versão específica para o caso de variáveis cointegrantes, que é o VECM (Modelo com Vetor de Correção de Erros).

Conforme Lutkepohl e Kratzig (2004, p. 88), os modelos VAR podem ser representados para o caso de um conjunto de K séries de tempo  $y_t = (y_{1t}, \dots, y_{kt})'$  e para o caso de um modelo de ordem p (Var(p)):

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (4.33)$$

Na especificação acima,  $A_i$  são  $(K \times K)$  matrizes de coeficientes e  $u_t = (u_{1t}, \dots, u_{kt})'$  é um vetor que representa os termos de erro. Assume-se que os termos de erro têm distribuição normal e são independentes do tempo, isto é,  $u_t \sim N(0, \Sigma_u)$ .

Destaca-se que os modelos VAR podem ser usados para dois objetivos básicos: previsão e análise estrutural. Para o primeiro caso, deve-se usar o VAR na forma reduzida e no segundo, na forma estrutural. No VAR reduzido capturam-se as propriedades dinâmicas entre as séries consideradas. Isso não seria suficiente para uma análise estrutural do modelo, uma vez que diferentes teorias econômicas podem ser compatíveis com um mesmo modelo reduzido.

No caso do VAR estrutural o objetivo central é analisar choques independentes entre as variáveis em estudo, de modo a se poder evitar análises erradas da causalidade (análise de impulso-resposta e decomposição da variância, por exemplo). Nesse caso, deve-se estimar o VAR na forma estrutural via processo de identificação. Contudo, como o objetivo deste trabalho é a realização de previsões, não se fará análise mais aprofundada acerca do VAR estrutural.

Destaca-se que a representação do VAR acima (eq. nº 4.33) pode ser complementada com a inclusão de termos determinísticos como intercepto, tendência linear e dummies sazonais. Por exemplo, para o caso da inclusão de um termo de tendência linear ao VAR, ter-se-ia:

$$y_t = v_0 + v_{1t} + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (4.34)$$

Outra forma de se aprimorar o modelo VAR antes apresentado se dá via inclusão de variáveis exógenas ao modelo, que são aqueles que influenciam a determinação das variáveis endógenas do modelo VAR, mas não são definidas por ela, mas sim exogenamente.

Mais um ponto que deve ser levado em conta na determinação do VAR é a escolha da ordem  $p$  do mesmo, isto é, da quantidade de defasagens das variáveis em questão. Isso pode ser feito via critérios de informação.

Depois de determinada a forma a ser utilizada pelo modelo VAR, parte-se para o processo de estimação. No caso de um modelo com  $y_1, \dots, y_t$  e  $K$  equações do modelo VAR devem ser estimadas separadamente via Mínimos Quadrados Ordinários (MQO). O resultado da estimação são eficientes e consistentes, equivalentes ao método Mínimos Quadrados Generalizados (GLS) (Lutkepohl e Kratzig, 2004, p. 93).

Destaca-se que antes de realizar os procedimentos descritos acima, deve-se tratar da questão da estacionaridade das séries usadas no modelo. Primeiro, deve-se fazer os testes de raiz unitária para saber se o VAR em questão deve ser estimado em nível ou em diferenças.

Mesmo que os testes apontem para a existência de raiz unitária nas séries envolvidas, antes de partir para a estimação do VAR em diferença, deve-se realizar novo teste, o de cointegração. A ideia é que series  $I(1)$  podem apresentar alguma relação de longo prazo entre elas, isto é, podem cointegrar. Se isso for verdade, o VAR deve ser adaptado para o caso VECM.

Formalmente, um conjunto de series  $I(1)$  será cointegrado se existir pelo menos uma combinação linear entre elas que gere uma série  $I(0)$ , isto é, estacionária.

Por exemplo, para o caso de duas variáveis  $I(1)$ ,  $y_t$  e  $x_t$ , cuja relação de longo prazo (cointegração) seja dada por  $y_t = \beta x_t + u_t$ , o termo  $u_t$  representa os desvios de curto prazo da relação de cointegração de longo prazo (Martin, Hurn e Harris, 2012, p.698).

Tais desvios podem ser corrigidos por ambas as variáveis endógenas do modelo. Supondo, por exemplo, que  $x_t$  tenha se afastado da relação de equilíbrio para certo  $y_t$ , dado algum choque aleatório. A correção desse erro poderia se dar tanto por pelo próprio  $x_t$ ,  $\Delta x_t$ , quanto por  $y_t$ ,  $\Delta y_t$ . Assim, ter-se-ia:

$$\Delta x_t = \alpha_x(x_{t-1} - x_{t-1}^*) \quad (4.35)$$

$$\Delta y_t = \alpha_y(x_{t-1} - x_{t-1}^*) \quad (4.36)$$

Nas expressões acima, os termos  $\alpha_x$  e  $\alpha_y$  representam os vetores de correção de erro. Destaca-se que apesar de apresentados separadamente, o mais comum é que a correção do erro não ocorra com o movimento de apenas uma variável, mas sim de todas consideradas.

Acrescentando termos de erro aleatório às equações apresentada acima, tem-se um VECM de ordem zero. Outra representação para esse modelo é mostrada abaixo:

$$x_t = (1 + \alpha_x)x_{t-1} - \alpha_x\beta y_{t-1} + v_t \quad (4.37)$$

$$y_t = \alpha_y x_{t-1} + (1 - \alpha_y\beta)y_{t-1} + u_t \quad (4.38)$$

Destaca-se que nesse caso tem-se um VAR de ordem 1, sendo que seus parâmetros apresentam restrições: eles não são lineares. Assim, o modelo em questão deve ser estimado com algum método não linear. Ressalta-se que a escolha do número de defasagens a ser incluída no modelo é dada pelo número ótimo de defasagens do VAR dessas mesmas variáveis descontada uma unidade.

Uma forma de testar a existência de cointegração se dá com a utilização do teste apresentado por Johansen (1988). Basicamente, esse teste calcula o posto da matriz dos coeficientes do vetor de cointegração, dado por  $\Pi$  na expressão:  $\Delta X_t = C + \Pi X_{t-1} + \Lambda \Delta X_{t-1} + e_t$ .

Caso o posto de  $\Pi = 0$ , então não há cointegração; se  $\Pi = n$  (número de variáveis do modelo), então não há cointegração e as séries são  $I(0)$  e; caso  $\Pi = r < n$ , então há  $r$  independentes relações de cointegração.

#### 4.1.2.3 Métodos de Combinação

Desde trabalhos mais pioneiros como o de Bates e Granger (1969), uma literatura vem apontando para a possibilidade de melhoria do trabalho da previsão com a utilização da combinação de diferentes modelos de previsão.

Timmermann (2006) faz uma interessante revisão do assunto, apontando que as razões para a maior acurácia dos modelos de combinação não são totalmente entendidas, mas cita algumas possibilidades para isso.

A premissa básica do trabalho de combinação é criar uma nova previsão com base na combinação de diferentes modelos, sendo que para cada um deles é dado certo peso. Dessa forma, o trabalho se concentra em grande medida na forma pela qual esses pesos serão determinados.

Este trabalho utiliza quatro maneiras para determiná-los e, assim, realizar a previsão, a saber: Média Simples, *Least Squares Weights*, *Mean Square Error (MSE) Weights* e *Mean Square Error (MSE) Rank*. Para todos eles, a combinação será feita com o melhor modelo de previsão univariado e o melhor multivariado.

Na média simples, a combinação se dará com uma média aritmética dos dois modelos considerados. Para o *Least Squares Weights*, será feita regressão dos valores observados contra as previsões e, assim, determinados parâmetros, para cada modelo, que serão usados na combinação. Destaca-se que a regressão tem que ser feita dentro de algum subperíodo do total de meses da previsão. O subperíodo escolhido foi o de julho de 2015 até dezembro de 2015. Já as previsões com a combinação ocorrerão nos meses de 2016.

Para o *Mean Square Error (MSE) Weights*, no mesmo subperíodo citado acima, calcula-se o MSE dos dois modelos considerados e, depois, obtêm-se os pesos para cada um deles, dados pela razão de seus respectivos MSE sobre a soma do MSE de ambos os modelos. Com os pesos, são achadas as previsões da combinação para os meses de 2016.

Por fim, no *MSE Rank*, a ideia é basicamente a mesma do *MSE Weights* a não ser pelo fato de que são obtidos *ranks* (classificações) com base nos MSE para cada modelo de previsão. Os pesos de cada um dos modelos são obtidos pelas razões de cada um desses *ranks* pela soma de todos eles.

#### 4.1.2.4 Estatísticas de Erro de Previsão

A forma de *analisar* a capacidade preditiva dos modelos se dará via utilização de estatísticas calculadas sobre os erros de previsão dos modelos. Os erros de previsão são aqui definidos como  $e_i = y_i - \hat{y}_i$ , isto é, o erro na data  $i$  é dado pela diferença entre o valor observado e o estimado nessa data. As estatísticas de erros usadas neste trabalho serão:

- Média dos Erros ao Quadrado (MSE):  $\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$
- Raiz da Média dos Erros ao Quadrado (RMSE):  $\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$
- Média dos Erros Absolutos (MAE):  $\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$
- Média dos Erros Absolutos Percentuais (MAPE):  $\frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$
- Erro Percentual Anual (% Anual):  $\frac{\sum_{t=1}^{12} \hat{y}_t - \sum_{t=1}^{12} y_t}{\sum_{t=1}^{12} y_t} * 100$

Conforme (Hyndman e Athanasopoulos, 2013, seção 2/5), as medidas de acurácia MSE, RMSE e MAE tem o inconveniente de serem dependentes da escala de medida da série utilizada, de modo a tornar mais difícil a comparação entre séries que possuam diferentes escalas. Esse problema não é observado no caso do MAPE e do % Anual.

## 4.2 Dados

### 4.2.1 Séries Utilizadas

Obviamente, a principal série usada neste trabalho é a do ICMS do ERJ. As demais séries consideradas são as seguintes: Índice de Atividade Econômica do Brasil (IBC-BR), Índice de Atividade Econômica do Rio de Janeiro (IBC-RJ), Pesquisa Mensal de Comércio - RJ (PMC-RJ), Pesquisa Industrial Mensal de Produção Física – RJ (PIM-PF-RJ) e *Composite Leading Indicator* (CLI-BR).

Conforme explicado no capítulo 3, quatro seções do CNAE 2.0 respondem por cerca de 91,00% da base de incidência do ICMS do Rio de Janeiro. Dessa forma, optou-se por escolher séries econômicas que refletissem o comportamento dessas seções de forma agregada.

A característica agregada citada se justifica pelo fato de que, também conforme explicado no capítulo 3, dentro de cada uma dessas seções não havia concentração de importância dentro de uma ou outra subclasse, mas sim uma distribuição entre várias delas, cada uma com um pequeno peso em relação ao total<sup>13</sup>.

No sentido de aprimorar a análise, optou-se por incluir mais uma variável, a CLI-BR calculada pela OCDE. De diferente das demais utilizadas que são séries coincidentes, o CLI é uma série antecedente, isto é, busca apontar sinais de comportamento futuro da atividade econômica, identificando em especial mudanças do ciclo econômico.

Tabela 11: Variáveis Explicativas – Modelos Multivariados

Variável	Abrangência	Fonte	Descrição
IBC	BR	BCB	Nº Índice sem Ajuste Sazonal - Série 24.363
IBC	RJ	BCB	Nº Índice sem Ajuste Sazonal - Série 20.444
PMC	RJ	IBGE	Nº Índice sem Ajuste Sazonal
PIM-PF	RJ	IBGE	Nº Índice sem Ajuste Sazonal
CLI	BR	OCDE	Série em Nº Índice

Fonte: Elaboração Própria.

Uma forma de confirmar a relevância das séries citadas em explicar as variações do ICMS se dá via realização do Teste de Causalidade de Granger (Granger, 1969). A ideia é que, em um sistema VAR bivariado, por exemplo, se as defasagens da variável X ajudarem a explicar as variações na variável Y (em um conjunto informacional que já contém os lags de Y), então X Granger-causa Y.

Esse teste se dá com a estimação de dois modelos: um que Y depende apenas das suas próprias defasagens e outro no qual Y depende além do seu

<sup>13</sup> A subclasse com maior peso individual era a de Distribuição de Energia Elétrica com apenas 14,18% do total.

passado, dos lags de X. Então, faz-se o teste com a hipótese de que as defasagens de X são conjuntamente não significativas, indicando que X não-Granger causa Y.

Tabela 12: Testes de Causalidade de Granger

Hipótese	Nº Obs.	P-Valor
IBC-BR não-Granger causa ICMS ICMS não-Granger causa IBC-BR	156	0,000 0,000
IBC-RJ não-Granger causa ICMS ICMS não-Granger causa IBC-RJ	156	0,000 0,000
PMC-RJ não-Granger causa ICMS ICMS não-Granger causa PMC-RJ	156	0,000 0,000
PIM-PF-RJ não-Granger causa ICMS ICMS não-Granger causa PIM-PF-RJ	156	0,029 0,010
CLI-BR não-Granger causa ICMS ICMS não-Granger causa CLI-BR	156	0,139 0,291

Fonte: Elaboração Própria.

Nota: Para todos os casos, nº de lags = 12.

A hipótese de não causalidade de Granger é rejeitada, para as séries do IBC-BR, IBC-RJ e PMC-RJ, ao nível de significâncias de 1%. Para a PIM-PF-RJ, a hipótese nula é rejeitada ao nível de 5%. Por fim, o CLI destoou dos demais resultados, uma vez que a hipótese nula foi aceita ao nível de significância de 10%. Apesar disso, optou-se por incluí-lo no modelo, dado que essa série possui característica distinta das demais que pode aprimorar os resultados de previsão.

#### 4.2.2 Testes de Estacionariedade

Conforme antes explicado, para a escolha da melhor representação dos modelos, é necessário analisar a existência ou não de estacionariedade nas séries utilizadas. Isso será feito com os testes acima apresentados: ADF, PP e KPSS

Tabela 13: Testes ADF, PP e KPSS – Séries em Nível – Estatísticas de Teste

Série	ADF		PP		KPSS	
	Sem Det	Tend e Cte	Sem Det	Tend e Cte	Cte	Tend e Cte
IBC-RJ	0,70	1,44	1,62	-3,10	1,49***	0,34***
IBC-BR	0,50	-0,69	0,99	-3,28*	1,41***	0,31***
PMC-RJ	0,11	1,56	-0,35	-10,29***	1,56***	0,23***
PIM-PF-RJ	-0,04	-1,15	0,01	-5,13***	0,40*	0,36***
CLI-BR	0,33	-3,00	0,26	-2,77	0,28	0,15**

Fonte: Elaboração Própria.

Nota: Para ADF e PP, H0: Série tem uma raiz unitária. Para KPSS, H0: Série não tem raiz unitária.

Nota: Número de Lags para o ADF foi definido com base no critério de informação SBC.

Nota: \*  $p < 0,10$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$ .

Nota: Sem Det: Sem termos determinísticos; Tend e Cte: Com tendência e constante.

Tabela 14: Testes ADF, PP e KPSS – Séries em 1ª Dif. – Estatísticas de Teste

Série	ADF		PP		KPSS	
	Sem Det	Tend e Cte	Sem Det	Tend e Cte	Cte	Tend e Cte
IBC-RJ	-1,19	-2,70	-28,33***	-33,17***	0,31	0,14*
IBC-BR	-1,97**	-3,01	-20,40***	-26,43***	0,32	0,08
PMC-RJ	-0,66	-1,02	-43,43***	-62,76***	0,21	0,17**
PIM-PF-RJ	-4,28***	-4,80***	-35,94***	-51,03***	0,31	0,12*
CLI-BR	-4,40***	-4,38***	-3,50***	-3,50**	0,05	0,05

Fonte: Elaboração Própria.

Nota: Para ADF e PP, H0: Série tem uma raiz unitária. Para KPSS, H0: Série não tem raiz unitária.

Nota: Número de Lags para o ADF foi definido com base no critério de informação SBC.

Nota: \*  $p < 0,10$ , \*\*  $p < 0,05$ , \*\*\*  $p < 0,01$ .

Nota: Sem Det: Sem termos determinísticos; Tend e Cte: Com tendência e constante.

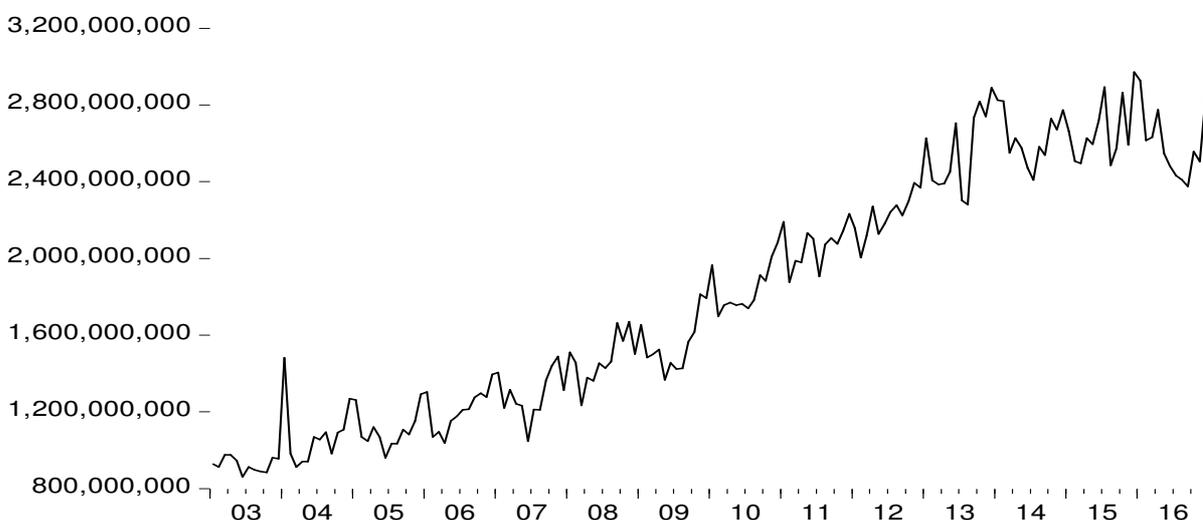
As tabelas 13 e 14 indicam que, em nível, todas as séries testadas apresentam raiz unitária. Porém, quando elas estão em sua versão de primeira diferença, observa-se o oposto, isto é, elas se tornam estacionárias.

### 4.2.3 Série do ICMS

Conforme exposto na figura 4, a série do ICMS do ERJ apresentou uma trajetória bem determinada de crescimento entre os anos de 2003 e 2013, quando saiu, em termos nominais, de uma arrecadação anual de R\$ 11.094.920.312,21 para R\$ 30.727.000.000,00. Contudo, a partir de 2014, o que se observou foi uma estabilidade da arrecadação até 2016 (R\$ 31.573.740.000,00, R\$ 31.982.770.000,00 e R\$ 31.099.590.000,00 em, respectivamente, 2014, 2015 e 2016).

A principal razão para essa piora foi a queda da atividade econômica fluminense no período, refletindo o quadro observado em nível nacional. Para os anos de 2014, 2015 e 2016, as variações do PIB brasileiro foram de, respectivamente, 0,50%, -3,80% e -3,60%. Já para o IBC-RJ, para esses mesmos períodos, as variações foram de 0,94%, -1,88% e -3,39%.

Figura 4: Série ICMS – Estado do Rio de Janeiro – 2003-2016:



Fonte: Elaboração Própria.

Além do componente de tendência, fica claro no gráfico a presença de considerável sazonalidade na série. Especialmente no começo do ano (provavelmente em janeiro) ocorrem picos de arrecadação não observados nos demais meses. Essa questão sazonal será mais bem estudada no capítulo dos resultados, especificamente na parte referente ao modelo SARIMA.

Tabela 15: Testes ADF, PP e KPSS – ICMS – Estatísticas de Teste

Série	ADF		PP		KPSS	
	Sem Det	Tend e Cte	Sem Det	Tend e Cte	Cte	Tend e Cte
ICMS	1,34	-5,95***	2,20	-6,08***	1,59***	0,20**
ICMS 1ª Dif	-12,83***	-12,96***	-23,17***	-27,09***	0,08	0,06

Fonte: Elaboração Própria.

Nota: Para ADF e PP, H0: Série tem uma raiz unitária. Para KPSS, H0: Série não tem raiz unitária.

Nota: Número de Lags para o ADF foi definido com base no critério de informação SBC.

Nota: \* p<0,10, \*\* p<0,05, \*\*\* p<0,01.

Nota: Sem Det: Sem termos determinísticos; Tend e Cte: Com tendência e constante.

Os resultados dos testes de raiz unitária apontam para a não estacionariedade da série do ICMS em nível. Na verdade, essa hipótese é aceita apenas em dois casos, ADF e PP com componente de tendência. Isso vai ao encontro do mostrado na figura 4. Tomando a primeira diferença da série, sob todos os testes realizados, a série se torna estacionária.

## 5 Resultados

### 5.1 Análise do Modelo da SEFAZ-RJ

Antes de tratar das modelagens do ICMS do ERJ propostas neste trabalho, será realizada uma breve análise do modelo usado atualmente pela SEFAZ-RJ. Isso será feito via: i) descrição das suas principais características e; ii) comparação dos resultados de previsão da sua formulação original com outras alternativas no intuito de obter alguma intuição acerca das suas possíveis fragilidades.

Nesse sentido, o modelo utilizado pela SEFAZ-RJ nos últimos quatro anos<sup>14</sup> (2013 a 2016) para realização da previsão do ICMS do Estado do Rio de Janeiro apresenta as seguintes características:

- Metodologia: VAR com as séries em nível e usando duas defasagens das variáveis endógenas e os valores contemporâneos das exógenas.
- Variáveis usadas na Estimação do Modelo: As duas variáveis endógenas do modelo são: ICMS do ERJ e Pesquisa Mensal de Comércio ERJ – PMC-RJ do IBGE<sup>15</sup>. As variáveis exógenas são: PIB Mensal – Valores Correntes<sup>16</sup>; Índice da Taxa de Câmbio Efetiva Real (IPCA)<sup>17</sup> e; Vendas, pelas Distribuidoras, dos Derivados de Petróleo (no ERJ e medidos em m<sup>3</sup>).
- Variáveis usadas na Previsão: As duas variáveis endógenas são previstas pelo próprio modelo considerado. Em relação às exógenas, as previsões para o PIB e para o câmbio têm como base o Boletim Focus, do Banco Central do Brasil, mais recente disponível no momento da previsão. No caso do PIB, primeiro são calculados os equivalentes mensais das taxas de crescimento anuais do PIB do ano de elaboração da previsão e do ano seguinte. Depois, tais estimativas são usadas na projeção da série do PIB via suposição que as taxas de crescimento

---

<sup>14</sup> Período que a SEFAZ passou a usar técnicas econométricas na previsão das receitas do ERJ.

<sup>15</sup> Série em número índice sem ajuste sazonal.

<sup>16</sup> Código da série 4380 no Sistema Gerador de Séries Temporais (SGS) do Banco Central do Brasil. Essa série é calculada mensalmente pelo Banco Central com base em informações econômicas do mês de referência e é usada para o cálculo da relação entre agregados econômicos mensais (como dívida pública ou saldo de crédito, por exemplo) e o PIB. Assim, ela ajuda a contornar o problema, para este caso, da periodicidade trimestral do PIB oficial do país calculado pelo IBGE. Ela se refere a todo o Brasil e não apenas ao ERJ.

<sup>17</sup> Código da série 11752 no Sistema Gerador de Séries Temporais (SGS) do Banco Central do Brasil.

mensais da série seguirão as estimativas obtidas. Para o câmbio, deve-se primeiro calcular a variação percentual entre as taxas previstas no Boletim Focus para o final dos anos em questão e a taxa nominal do mês de estimação. Obtidas as taxas, deve-se calcular as equivalentes mensais e usá-las nas previsões da taxa de câmbio efetiva real. Por fim, para a série de vendas dos derivados de petróleo, a estimativa é realizada com um modelo SARIMA construído no software Demetra+ com a especificação RSA5.

- Transformações nas Séries usadas na Estimação: Antes da estimação dos modelos, as séries do ICMS e do PIB Mensal são deflacionadas com o uso do IPCA e todas as cinco séries são modificadas para suas versões em logaritmo.
- Periodicidade dos Dados: Mensal. Os dados para estimação vão de Jan/2003 até o mês da estimação (final de março para LDO e começo de agosto para LOA).
- Períodos de Previsão: Três anos seguintes ao da elaboração das previsões.

Descrito o modelo usado pela SEFAZ-RJ para previsão do ICMS, passa-se agora para o exame da sua capacidade de previsão e a das suas versões alternativas aqui consideradas. Tais versões modificarão a original via exclusão de alguma ou todas as variáveis exógenas incluídas no modelo primário.

Os períodos de estimação e de previsão são os mesmos considerados no restante deste trabalho, isto é, estimação realizada com os dados disponíveis até junho de 2015 e previsão feita de julho de 2015 até dezembro de 2016, sendo que os cálculos das estatísticas de erro terão como base apenas os meses de 2016.

Tabela 16: Resultados de Previsão: Modelo Original SEFAZ-RJ e Versões Alternativas

Modelo	% Anual	MSE	RMSE	MAE	MAPE
Original SEFAZ-RJ	12,74%	R\$ 383.510,46	R\$ 619,28	R\$ 529,93	20,62%
SEFAZ-RJ sem Exóg.	7,63%	R\$ 72.518,51	R\$ 269,29	R\$ 234,94	9,38%
SEFAZ-RJ sem PIB	4,57%	R\$ 467.870,03	R\$ 684,01	R\$ 492,11	18,99%
SEFAZ-RJ sem Câmbio	19,60%	R\$ 300.710,69	R\$ 548,37	R\$ 507,84	20,16%
SEFAZ-RJ sem Petróleo	10,32%	R\$ 298.759,91	R\$ 546,59	R\$ 451,60	17,55%

Fonte: Elaboração Própria.

Nota: Para as previsões das séries do PIB e do câmbio, utilizou-se o Boletim Focus BCB de 26/06/2015.

Nota: Na previsão da série do petróleo, utilizou-se o modelo SARIMA (2,1,1)(0,1,0)<sub>12</sub> no software Eviews 9.

A tabela 16 indica, de forma geral, que a exclusão das variáveis exógenas aumenta a capacidade preditiva do modelo da SEFAZ-RJ. No caso extremo, quando todas as três variáveis exógenas são retiradas, os erros de previsão caem para cerca da metade do modelo original.

Duas explicações principais para o padrão descrito acima podem ser apontadas: i) a incapacidade preditiva dos métodos usados na previsão das variáveis exógenas. No caso do PIB, enquanto que a mediana do Boletim Focus considerado projetava queda de 1,49% para 2015 e expansão de 0,50% para 2016, os valores efetivos foram -3,80% em 2015 e -3,60% em 2016. Para o câmbio se esperava para 2015 e 2016, respectivamente, dólar a R\$ 3,20 e R\$ 3,37 com base nas medianas do Boletim Focus. Contudo, os valores realizados foram R\$ 3,87 e R\$ 3,35. Para o petróleo, as estimativas foram 8.711.614,66 m<sup>3</sup> e 8.038.390,50 m<sup>3</sup> para 2015 e 2016, respectivamente, e os valores observados foram 8.663.521,44 m<sup>3</sup> e 8.056.805,39 m<sup>3</sup> e; ii) a não significância estatística de algumas das variáveis exógenas consideradas no modelo, conforme modelo exposto na tabela 17.

Tabela 17: Modelo Original SEFAZ-RJ

Modelo	LOG_ICMS	LOG_PMC
LOG_ICMS_DEFLACIONADO (-1)	0,3464*** (4,1521)	0,2365** (1,9789)
LOG_ICMS_DEFLACIONADO (-2)	0,0444 (0,6279)	-0,0995 (-0,9824)
LOG_PMC (-1)	0,3186*** (6,8143)	0,0010 (0,0144)
LOG_PMC (-2)	-0,1196** (-2,2404)	-0,0647 (-0,8461)
LOG_PIB	0,2630** (1,9901)	0,2269 (1,1980)
LOG_PETRÓLEO	0,0419 (0,3179)	0,9312*** (4,9303)
LOG_CÂMBIO	0,0551 (1,0994)	-0,1091 (-1,5193)
CONSTANTE	3,4583*** (5,2032)	-5,6598*** (-5,9424)

Fonte: Elaboração Própria.

Nota: \* p<0,10, \*\* p<0,05, \*\*\* p<0,01.

Nota: Estatísticas-t em ().

Outras alternativas poderiam ser trabalhadas via, por exemplo, diferenciação das séries no caso de constatação da não estacionariedade das séries, não utilização da forma logarítmica e a não retirada dos efeitos da inflação sobre as séries do ICMS e do PIB.

## 5.2 Análise Univariada

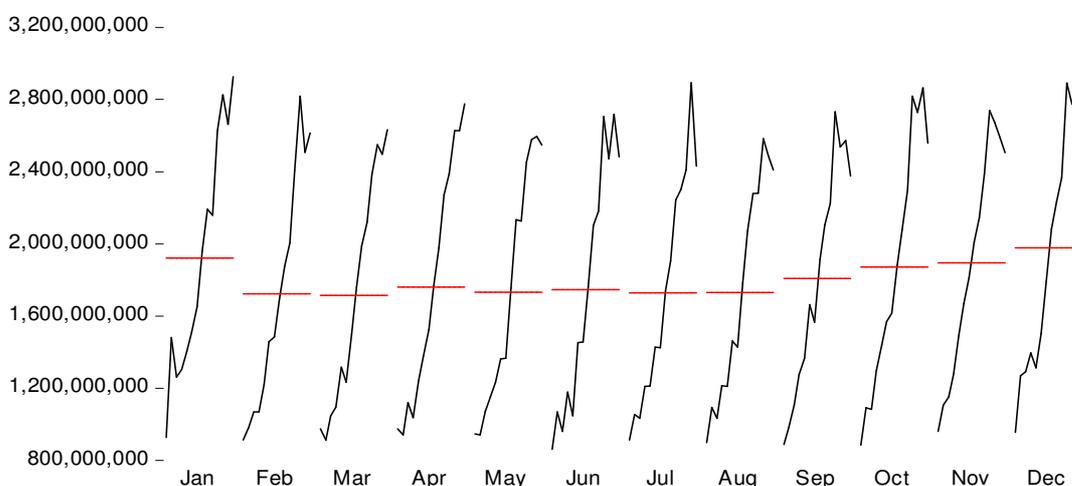
### 5.2.1 Exame Preliminar da Série do ICMS

A análise da série do ICMS apresentada no capítulo 4 indicou que a mesma possui clara tendência crescente até o ano de 2013 e certa estabilidade desde então. Esse sinal de não estacionariedade foi confirmado com os testes de raiz unitária expostos nesse mesmo capítulo.

Da mesma forma, a análise indicou a possibilidade de existência de componente sazonal na série, especialmente pelos picos observados no primeiro e nos quatro últimos meses dos anos.

A não estacionariedade da série do ICMS pode ser mais bem observada quando se leva em conta a figura 5 exposta abaixo. Esse gráfico, obtido no software Eviews 9, apresenta, para cada mês, a média de arrecadação com base nos anos da amostra, assim como a dispersão em torno dessa média.

Figura 5: Gráfico Sazonalidade do ICMS-RJ

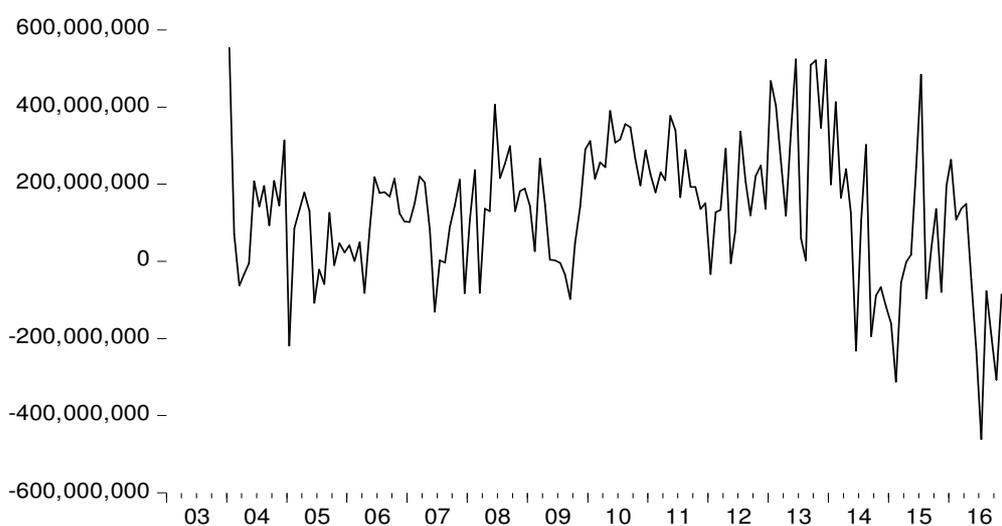


Fonte: Elaboração Própria usando o Software Eviews 9.

Pela observação do gráfico, percebe-se que a dispersão é parecida para todos os meses. Já a média apresenta pico significativo em janeiro, se mantendo constante entre fevereiro e agosto e a partir de setembro ocorre um processo de crescimento da mesma que se estende até dezembro.

Dito isso, optou-se pela diferenciação sazonal na série do ICMS (D=12). A figura 6 indica que a série resultante da diferenciação aparenta não ser estacionária, dado que há sinais de algumas tendências nos dados.

Figura 6: Série do ICMS com diferenciação Sazonal (D=12)



Fonte: Elaboração Própria.

Essa percepção foi corroborada com os resultados dos testes de estacionariedade realizados na série do ICMS do ERJ após a diferenciação sazonal. Com isso, optou-se por realizar uma diferenciação de primeira diferença (d=1) na série citada.

Tabela 18: Testes ADF para a série do ICMS-RJ com diferenciação Sazonal (D=12)

Série	Estatística-t		
	Sem Const. E Tend.	Const.	Const. e Tend.
SEFAZ com Dif. Sazonal	-1,45	-2,00	-1,91

Fonte: Elaboração Própria.

Nota: H0: Série é não estacionária.

Nota: \* p<0,10, \*\* p<0,05, \*\*\* p<0,01.

## 5.2.2 Escolha das Especificações do Modelo SARIMA

Feitas as transformações citadas acima, passa-se agora para a escolha das especificações dos modelos univariados SARIMA usados nas comparações. Essa tarefa se dará com a análise do correlograma da série do ICMS resultante das transformações citadas.

O correlograma dessa série apresenta picos nas defasagens 1 e 2 tanto na FAC quanto na FACP, indicando a necessidade de considerar-se termos ar  $\leq 2$  e ma  $\leq 2$ . Já na parte sazonal, observam-se picos nas defasagens 12 da FAC e 12 e 13 da FACP.

Contudo, os resultados das regressões feitas sob diferentes especificações, indicaram que os termos MA(2), AR(2) e sma(1) não apresentavam significância estatística. Assim, eles foram retirados da análise, de modo que restaram as seguintes possibilidades: SARIMA (p,d,q)(P,D,Q)<sub>12</sub> com  $p \leq 1$ ,  $d = 1$ ,  $q \leq 1$ ,  $P \leq 1$ ,  $D = 1$ ,  $Q = 0$ .

## 5.2.3 Comparação dos Modelos SARIMA

A comparação dos modelos SARIMA considerados será feita da seguinte forma: primeiro será analisada a estrutura dos modelos, levando em conta a significância estatística dos parâmetros, o coeficiente de determinação dos mesmos e seus valores para os critérios de informação utilizados. Depois, será testada a solidez dos modelos via análise dos resíduos. Nesse sentido, serão apresentados os resultados de testes de autocorrelação, homocedasticidade e normalidade. Por fim, serão destacadas as estatísticas de erro de previsão.

Iniciando a análise pela estrutura do modelo, a tabela 19 mostra que, em geral, os parâmetros estimados apresentam significância estatística, exceto pelos termos ar (1) dos modelos 7 e 8. Além disso, três modelos, 5, 6 e 8 se destacam na comparação dos coeficientes de determinação. Esses três modelos também apresentam os menores valores para os critérios de informação, conforme exposto na tabela 20.

Tabela 19: Modelos SARIMA – Análise Estrutural – R<sup>2</sup> e Significância dos Parâmetros

#	Modelos	R <sup>2</sup>	Significância dos Parâmetros		
			AR(1)	MA(1)	SAR(1)
1	SARIMA(0,1,0)(0,1,0) <sub>12</sub>	0,14	-	-	-
2	SARIMA(0,1,0)(1,1,0) <sub>12</sub>	0,43	-	-	-0,663*** (-8,583)
3	SARIMA(0,1,1)(0,1,0) <sub>12</sub>	0,33	-	-0,621*** (-8,787)	-
4	SARIMA(1,1,0)(0,1,0) <sub>12</sub>	0,23	-0,345*** (-4,190)	-	-
5	SARIMA(0,1,1)(1,1,0) <sub>12</sub>	0,56	-	-0,600*** (-8,453)	-0,671*** (-8,790)
6	SARIMA(1,1,0)(1,1,0) <sub>12</sub>	0,51	-0,397*** (-4,944)	-	-0,690*** (-9,087)
7	SARIMA(1,1,1)(0,1,0) <sub>12</sub>	0,33	0,134 (0,966)	-0,706*** (-6,997)	-
8	SARIMA(1,1,1)(1,1,0) <sub>12</sub>	0,56	0,036 (0,248)	-0,624*** (-5,445)	-0,669*** (-8,724)

Fonte: Elaboração Própria.

Nota: \* p<0,10, \*\* p<0,05, \*\*\* p<0,01.

Nota: Estatísticas-t em ().

Nota: Em todos os modelos foram usadas duas dummies (fev/2004) e (jan/2005) para controle de outliers.

Tabela 20: Modelos SARIMA – Análise Estrutural – Critérios de Inf.

#	Modelos	Critérios de Informação		
		AIC	BIC	HQ
1	SARIMA(0,1,0)(0,1,0) <sub>12</sub>	40,59	40,66	40,62
2	SARIMA(0,1,0)(1,1,0) <sub>12</sub>	40,25	40,34	40,29
3	SARIMA(0,1,1)(0,1,0) <sub>12</sub>	40,37	40,45	40,40
4	SARIMA(1,1,0)(0,1,0) <sub>12</sub>	40,49	40,58	40,53
5	SARIMA(0,1,1)(1,1,0) <sub>12</sub>	40,01	40,11	40,05
6	SARIMA(1,1,0)(1,1,0) <sub>12</sub>	40,11	40,22	40,16
7	SARIMA(1,1,1)(0,1,0) <sub>12</sub>	40,37	40,48	40,42
8	SARIMA(1,1,1)(1,1,0) <sub>12</sub>	40,02	40,15	40,07

Fonte: Elaboração Própria.

Passa-se agora para a averiguação da solidez dos modelos via teste de resíduos. A tabela 21 indica que todos quase todos os modelos passam nos testes de homocedasticidade e de normalidade, padrão oposto ao observado nos testes de autocorrelação, que são satisfeitos apenas pelos modelos 5 e 8.

Tabela 21: Modelos SARIMA – Análise dos Resíduos

#	Modelos	Autocorrelação		Normalidade	Homocedasticidade
		Q-Test	LM-Test	Jarque-Bera	White
1	SARIMA(0,1,0)(0,1,0) <sub>12</sub>	Todas	0,000	0,017	0,723
2	SARIMA(0,1,0)(1,1,0) <sub>12</sub>	Todas	0,000	0,401	0,906
3	SARIMA(0,1,1)(0,1,0) <sub>12</sub>	Muitas	0,000	0,307	0,992
4	SARIMA(1,1,0)(0,1,0) <sub>12</sub>	Muitas	0,000	0,077	0,675
5	SARIMA(0,1,1)(1,1,0) <sub>12</sub>	24	0,123	0,919	0,996
6	SARIMA(1,1,0)(1,1,0) <sub>12</sub>	Todas	0,000	0,492	0,944
7	SARIMA(1,1,1)(0,1,0) <sub>12</sub>	A partir da 12	0,000	0,142	0,815
8	SARIMA(1,1,1)(1,1,0) <sub>12</sub>	24	0,112	0,915	0,976

Fonte: Elaboração Própria.

Nota: Exceto para o Q-Test, os demais valores são p-valores.

Nota: No Q-Teste, é dito para quais defasagens não se rejeita a hipótese de autocorrelação.

Por fim, na comparação da capacidade preditiva dos modelos selecionados, conforme tabela 22 abaixo as estatísticas de erro para cada um deles. Levando em conta a medida de erro mais simples, a do Erro % Anual, o modelo com melhor desempenho foi o 3, seguido dos modelos 7, 6 e 5.

Os modelos 5 e 6 tiveram os melhores resultados na análise das outras quatro estatísticas de erros consideradas, sendo que eles apresentaram resultados praticamente idênticos para tais análises dos erros.

Tabela 22: Resultados de Previsão: Modelos SARIMA

#	Modelo	% Anual	MSE	RMSE	MAE	MAPE
1	SARIMA(0,1,0)(0,1,0) <sub>12</sub>	17,67%	R\$ 283.815,94	R\$ 532,74	R\$ 457,90	18,19%
2	SARIMA(0,1,0)(1,1,0) <sub>12</sub>	7,14%	R\$ 55.515,68	R\$ 235,62	R\$ 197,68	7,86%
3	SARIMA(0,1,1)(0,1,0) <sub>12</sub>	5,59%	R\$ 54.373,54	R\$ 233,18	R\$ 198,77	7,87%
4	SARIMA(1,1,0)(0,1,0) <sub>12</sub>	13,89%	R\$ 187.646,07	R\$ 433,18	R\$ 368,50	14,66%
5	SARIMA(0,1,1)(1,1,0) <sub>12</sub>	6,64%	R\$ 48.771,29	R\$ 220,84	R\$ 183,81	7,31%
6	SARIMA(1,1,0)(1,1,0) <sub>12</sub>	6,60%	R\$ 48.299,65	R\$ 219,77	R\$ 183,48	7,30%
7	SARIMA(1,1,1)(0,1,0) <sub>12</sub>	5,90%	R\$ 57.730,77	R\$ 240,27	R\$ 204,62	8,10%
8	SARIMA(1,1,1)(1,1,0) <sub>12</sub>	6,86%	R\$ 51.701,24	R\$ 227,38	R\$ 190,82	7,59%

Fonte: Elaboração Própria.

Levando em conta todos os critérios acima apresentados, chegou-se a conclusão que o modelo 5, SARIMA(0,1,1)(1,1,0)<sub>12</sub>, é o que melhor consegue conciliar as propriedades estatísticas desejáveis aos resultados de previsão.

Esse modelo apresentou significância estatística em todos os seus parâmetros, teve o maior coeficiente de determinação e os menores valores para os critérios de informação entre os modelos considerados. Da mesma forma, passou em todos os testes realizados para os resíduos.

Ademais, com exceção do critério erro % anual (que obteve resultado não distante do melhor modelo, o de nº 3), obteve os menores valores das estatísticas de erro, ao lado do modelo 6, para as quatro demais estatísticas consideradas.

Uma maneira de confirmar a capacidade preditiva do modelo SARIMA escolhido antes de compará-lo com os resultados da análise multivariada, se dá via confronto com outro modelo univariado, a saber, o de Holt-Winters. Conforme apresentado no capítulo referente à metodologia, essa especificação consegue levar com conta as características de tendência e de sazonalidade, que, conforme já exposto, estão presentes na série original do ICMS do ERJ.

Tabela 23: Resultados de Previsão: SARIMA e Holt-Winters

Modelo	% Anual	MSE	RMSE	MAE	MAPE
SARIMA(0,1,1)(1,1,0) <sub>12</sub>	6,64%	R\$ 48.771,29	R\$ 220,84	R\$ 183,81	7,31%
Holt-Winters Mult.	10,94%	R\$ 100.932,85	R\$ 317,70	R\$ 283,49	9,80%
Holt-Winters Adit.	9,98%	R\$ 90.698,16	R\$ 301,16	R\$ 258,53	9,06%

Fonte: Elaboração Própria.

Como se pode observar a o modelo SARIMA(0,1,1)(1,1,0)<sub>12</sub> obteve previsões mais acuradas do que as dos dois casos do modelo Holt-Winters considerados, de modo que o SARIMA passou no teste de comparação com uma especificação univariada alternativa.

## **5.3 Análise Multivariada**

### **5.3.1 Descrição dos Modelos Multivariados Considerados**

O capítulo 3 mostrou que a base de incidência do ICMS do ERJ estava concentrada em quatro setores (seções do CNAE): comércio (34,00%), indústria de transformação (25,00%), eletricidade e gás (19,00%) e informação e comunicação (13,00%). Também, não houve relevante concentração em apenas uma subclasse de atividade, dado que a mais relevante individualmente foi a de distribuição de energia elétrica (pertencente ao setor de eletricidade e gás), respondendo por 14,18% do total do ICMS.

Dessa forma, para a análise multivariada, foi optado por priorizar séries que explicassem de forma agregada a economia fluminense e a brasileira, assim como os dois setores mais importantes citados, em detrimento de outras que tinham caráter mais setorial. As séries escolhidas foram: IBC-RJ (BCB), IBC-BR (BCB), PMC-RJ (IBGE) e PIM-PF-RJ (IBGE).

Destaca-se que essas séries são caracterizadas como coincidentes, isto é, retratam o comportamento contemporâneo da atividade econômica. No sentido de enriquecer a análise, optou-se por um introduzir outra variável no estudo multivariado, a CLI (OCDE), que é antecedente, isto é, busca antecipar movimentos futuros da economia.

Serão considerados doze modelos multivariados: quatro bivariados, cada um com o ICMS mais uma das quatro variáveis coincidentes destacadas; quatro com três variáveis, cada um com ICMS, algum dos dois IBC e PMC-RJ ou PIM-PF e; mais quatro com três variáveis, cada um com o CLI mais uma das quatro séries coincidentes, além do ICMS. Em todos os casos serão introduzidas dummies para correção dos outliers dos resíduos dos modelos estimados.

### 5.3.2 Análise da Cointegração

Com base nos testes de raiz unitária expostos no capítulo 4, todas as séries aqui consideradas são  $I(1)$ , de modo que se faz necessária a realização de testes de cointegração antes de partir diretamente para a regressão com variáveis em diferença.

O teste de cointegração escolhido foi o de Johansen, realizado no software Eviews 9. Esse teste demanda a especificação por parte do pesquisador de alguns dos seus componentes. Um deles é o número de lags, que foi determinado pela subtração de uma unidade do número ótimo dado pela análise dos critérios de informação<sup>18</sup> do modelo na forma de VAR em nível.

Outra escolha a ser feita é a dos componentes determinísticos a serem incluídos na equação de cointegração e no modelo VECM usado. Conforme explicação dada pelo manual do usuário do software Eviews, a quarta opção de componentes (entre cinco disponíveis), aquela que introduz uma constante e um termo de tendência na equação de cointegração, é a mais adequada para os modelos considerados, uma vez que a série do ICMS, usada em todos os casos abordados, apresenta tendência determinística.

Tabela 24: Teste de Cointegração de Johansen

#	Modelos	Nº de Lags	Nº de Vetores de Cointegração		VAR ou VEC
			Traço	Máximo Autovalor	
1	ICMS, IBC-RJ	12	1	1	VEC
2	ICMS, IBC-BR	12	1	1	VEC
3	ICMS, PMC-RJ	13	0	0	VAR
4	ICMS, PIM-PF-RJ	12	1	0	VAR
5	ICMS, IBC-RJ, PMC	11	3	3	VAR
6	ICMS, IBC-RJ, PIM-PF-RJ	12	3	1	VEC
7	ICMS, IBC-BR, PMC	11	1	1	VEC
8	ICMS, IBC-BR, PIM-PF-RJ	1	2	1	VEC

<sup>18</sup> No caso de resultados conflitantes para os critérios de informação selecionados, usou-se o critério de Hannan-Quinn como desempate.

9	ICMS, IBC-RJ, CLI	12	1	1	VEC
10	ICMS, IBC-BR, CLI	11	2	2	VEC
11	ICMS, PMC-RJ, CLI	11	1	2	VEC
12	ICMS, PIMPF-RJ, CLI	4	0	0	VAR

Fonte: Elaboração Própria.

Nota: Em todos os modelos usou-se a especificação de teste com termos determinísticos constante e tendência no vetor de cointegração.

Destaca-se que nos casos que apresentaram vetores de cointegração diferentes de zero, mas com valores distintos (modelos 6, 8 e 11), optou-se por escolher a forma de VEC com o menor número de vetores de cointegração entre os dois conflitantes. Já no caso em que o número de vetores de cointegração era igual ao de variáveis do modelo (modelo 5), usou-se a modelagem em VAR, dado que se fosse assumido que o resultado estivesse correto, ter-se-ia que aceitar que as séries são estacionárias em nível, indo de encontro aos testes apresentados no capítulo 4.

### 5.3.3 Comparação dos Modelos Multivariados

Antes de passar para a comparação da capacidade preditiva dos modelos multivariados, respeitando as conclusões dos testes de cointegração dados na tabela 24, serão expostos os resultados dos testes dos resíduos das regressões a fim de avaliar possíveis fragilidades dos modelos, que comprometeriam a solidez da estimação dos mesmos conforme explicado no capítulo 4.

Destaca-se que foram introduzidas dummies, específicas para cada modelo, com o intuito de corrigir os impactos dos outliers das séries dos resíduos das estimações realizadas. As dummies foram selecionadas com base nas maiores dispersões observadas nas séries de resíduos. As dummies usadas para cada modelo estão expostas na tabela 25 que também traz as conclusões dos mesmos.

Tabela 25: Modelos Multivariados – Análise dos Resíduos

#	Modelos	Autoc.	Normal.	Homoced	Dummies
		LM-Test	Jarque-Bera	White	
1	ICMS, IBC-RJ	22	0,960	0,179	02/13, 09/13, 02/15
2	ICMS, IBC-BR	Vários	0,257	0,109	-
3	ICMS, PMC-RJ	2,4,12,19	0,997	0,125	06/07, 03/14
4	ICMS, PIM-PF-RJ	1,11	0,469	0,138	01/12
5	ICMS, IBC-RJ, PMC	12,16	0,218	0,090	06/07,03/09,01/12,04 /11,03/14
6	ICMS, IBC-RJ, PIM-PF-RJ	7,8,12,16	0,697	0,090	01/12, 09/13
7	ICMS, IBC-BR, PMC	1,2,12,13,2 1,22	0,304	0,182	01/04,04/15
8	ICMS, IBC-BR, PIM-PF-RJ	Vários	0,275	0,005	01/04
9	ICMS, IBC-RJ, CLI	22	0,956	0,179	02/13, 09/13, 02/15
10	ICMS, IBC-BR, CLI	12,13,18	0,680	0,124	01/04
11	ICMS, PMC-RJ, CLI	1,12,19,23	0,300	0,113	01/04
12	ICMS, PIMPF-RJ, CLI	Vários	0,000	0,850	01/04, 01/12, 03/08, 08/11

Fonte: Elaboração Própria.

Nota: Exceto para o LM-Test, os demais valores são p-valores.

Nota: No LM-Test, é dito para quais defasagens se rejeita a hipótese de autocorrelação.

A tabela 25 indica que apenas os modelos 8 e 12 apresentaram resultados insatisfatórios em mais de um dos testes realizados. Ambos rejeitaram a hipótese de ausência de autocorrelação para todas as defasagens consideradas (até a 24<sup>a</sup>) e, respectivamente, rejeitaram as hipóteses de homocedasticidade e normalidade.

Tabela 26: Resultados de Previsão: Modelos Multivariados

#	Modelo	% Anual	MSE	RMSE	MAE	MAPE
1	ICMS, IBC-RJ	8,71%	R\$ 67.261,81	R\$ 259,35	R\$ 225,65	9,03%
2	ICMS, IBC-BR	4,97%	R\$ 39.417,74	R\$ 198,54	R\$ 164,46	6,56%
3	ICMS, PMC-RJ	4,89%	R\$ 46.771,72	R\$ 216,27	R\$ 177,73	7,06%
4	ICMS, PIM-PF-RJ	10,03%	R\$ 101.099,40	R\$ 317,96	R\$ 278,06	11,11%
5	ICMS, IBC-RJ, PMC	6,73%	R\$ 60.401,39	R\$ 245,77	R\$ 200,27	8,00%
6	ICMS, IBC-RJ, PIM-PF-RJ	5,18%	R\$ 34.313,09	R\$ 185,24	R\$ 158,81	6,36%
7	ICMS, IBC-BR, PMC	0,74%	R\$ 38.489,44	R\$ 196,19	R\$ 169,18	6,57%

8	ICMS, IBC-BR, PIM-PF-RJ	4,78%	R\$ 48.987,80	R\$ 221,33	R\$ 194,45	7,69%
9	ICMS, IBC-RJ, CLI	8,17%	R\$ 63.162,95	R\$ 251,32	R\$ 215,92	8,65%
10	ICMS, IBC-BR, CLI	14,59%	R\$ 176.141,86	R\$ 419,69	R\$ 378,02	15,02%
11	ICMS, PMC-RJ, CLI	6,83%	R\$ 67.913,48	R\$ 260,60	R\$ 206,45	8,24%
12	ICMS, PIMPF-RJ, CLI	6,47%	R\$ 60.323,40	R\$ 245,61	R\$ 214,04	8,53%

Fonte: Elaboração Própria.

Os resultados das previsões dos modelos apontam para a escolha de um dos três seguintes modelos: 2, 6 e 7. Com exceção do critério Erro % Anual, esses três modelos foram os mais eficiente para as demais quatro estatísticas de erro.

O modelo aqui escolhido como o melhor multivariado é o 7, aquele que leva em conta, além do ICMS, o IBC-RJ e a PMC-RJ. Além de apresentar resultados satisfatórios na análise dos resíduos, esse modelo esteve com os outros dois citados entre os melhores nas especificações consideradas. Mas de maior importância, há o fato de que no Erro % Anual foi muito pequeno, apenas 0,74%.

## 5.4 Combinação de Modelos de Previsão

Conforme explicado no capítulo 4, os métodos de combinação considerados serão os seguintes: Média Simples, *Least Squares Weights*, *Mean Square Error (MSE) Weights* e *Mean Square Error (MSE) Rank*. As combinações serão feitas com dois modelos, o melhor univariado e o melhor multivariado, que são, respectivamente, SARIMA(0,1,1)(1,1,0)<sub>12</sub> e o VEC com ICMS, IBC-BR, PMC.

Para todos os métodos citados, à exceção das Médias Simples, necessita-se determinar um subperíodo dos meses da previsão para a determinação dos pesos para cada modelo. Como as previsões até aqui feitas vão de julho de 2015 até dezembro de 2016, contudo as estatísticas de erro calculadas apenas para os meses de 2016, torna-se natural a opção para a determinação dos pesos durante o período julho 2015 até dezembro de 2015 e as estatísticas de erro dos métodos de combinação calculadas com as previsões para os meses de 2016.

Tabela 27: Métodos de Combinação: Pesos para os Modelos de Previsão

#	Modelo	Modelo 1	Modelo 2
1	Média Simples	-	-
2	<i>Least Squares Weights</i>	-0.480	1.480
3	<i>MSE Weights</i>	0.526	0.474
4	<i>MSE Rank</i>	0.666	0.333

Fonte: Elaboração Própria.

Nota: Modelo 1: SARIMA(0,1,1)(1,1,0)<sub>12</sub> e Modelo 2: VEC com as variáveis ICMS, IBC-BR, PMC-RJ.

Nota: No caso do *Least Squares Weights*, os valores dados não são pesos, mas sim os parâmetros estimados.

Tabela 28: Resultados de Previsão: Métodos de Combinação

#	Modelo	% Anual	MSE	RMSE	MAE	MAPE
1	Média Simples	3,69%	R\$ 35.752,90	R\$ 189,08	R\$ 158,69	6,27%
2	<i>Least Squares Weights</i>	-2,09%	R\$ 55.929,26	R\$ 236,49	R\$ 203,04	7,76%
3	<i>MSE Weights</i>	3,54%	R\$ 35.509,32	R\$ 188,44	R\$ 157,53	6,22%
4	<i>MSE Rank</i>	4,68%	R\$ 38.341,82	R\$ 195,81	R\$ 166,23	6,58%

Fonte: Elaboração Própria.

À exceção do método 2 de combinação, *Least Squares Weights*, todos os três restantes apresentaram estatísticas de erro menores do que as do modelo SARIMA(0,1,1)(1,1,0)<sub>12</sub>, o melhor univariado, indicando que a combinação conseguiu aprimorar os resultados individuais do método univariado.

Exatamente o mesmo padrão acima, porém com menos intensidade, foi observado quando levando em conta o melhor método multivariado. A única diferença é que os métodos de combinação não conseguiram superar o VEC com as variáveis ICMS, IBC-BR, PMC-RJ quando comparados com base no Erro % anual.

Deixando de lado o método 2 de combinação, os demais três tiveram resultados muito semelhantes. Porém, como o método 3, *MSE Weights*, obteve resultados ligeiramente melhores do que os demais, ele foi escolhido como o melhor método de combinação.

## 5.5 Comparação dos Métodos de Previsão

Depois de apresentados isoladamente os três métodos de previsão aqui tratados (univariados, multivariados e combinação), passa-se para a comparação entre as melhores versões de cada um deles. Também será acrescentada a previsão dada pelo modelo utilizado pela SEFAZ-RJ, referência básica deste trabalho.

Tabela 29: Resultados de Previsão: Comparação Final

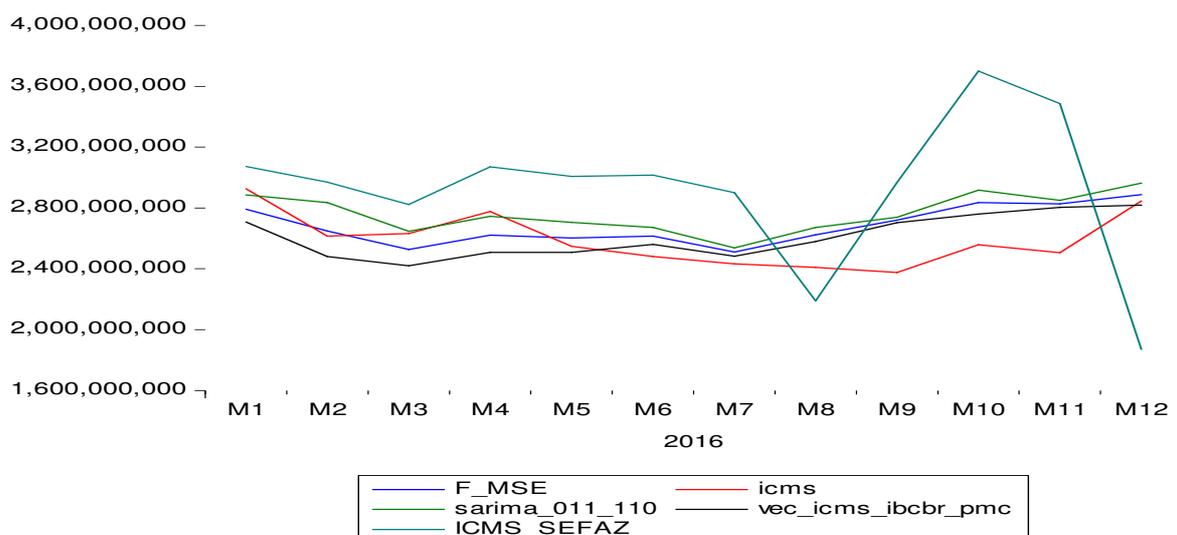
#	Modelo	% Anual	MSE	RMSE	MAE	MAPE
1	SARIMA(0,1,1)(1,1,0) <sub>12</sub>	6,64%	R\$ 48.771,29	R\$ 220,84	R\$ 183,81	7,31%
2	ICMS, IBC-BR, PMC	0,74%	R\$ 38.489,44	R\$ 196,19	R\$ 169,18	6,57%
3	MSE <i>Weights</i>	3,54%	R\$ 35.509,32	R\$ 188,44	R\$ 157,53	6,22%
4	SEFAZ-RJ	12,74%	R\$ 383.510,46	R\$ 619,28	R\$ 529,93	20,62%

Fonte: Elaboração Própria.

De princípio, observa-se, nos três modelos estimados, significativa melhoria da previsão do ICMS do Rio de Janeiro quando comparado com o método oficial da SEFAZ-RJ. Mesmo o modelo SARIMA(0,1,1)(1,1,0)<sub>12</sub>, o de pior previsão entre os três estimados, gera, na média das estatísticas de erro consideradas, redução de cerca de 65,00% de seus valores.

A tabela 29 também indica que em quatro das cinco estatísticas de erro consideradas, o modelo MSE *Weights* obteve os melhores resultados de previsão, porém seguidos de perto do modelo VEC (ICMS, IBC-BR e PMC-RJ). A única estatística de erro que não apresentou esse padrão foi o Erro % Anual, que teve no modelo VEC, de longe, o mais acurado.

Figura 7: Previsões 2016 – Modelos Considerados



Fonte: Elaboração Própria.

Conforme exposta na figura 7, todos os três modelos considerados conseguiram capturar de forma considerável os movimentos ao longo do ano da série do ICMS. Todos eles apresentaram pico no primeiro mês do ano e tiveram certa estabilidade entre fevereiro e julho, seguindo o padrão da série original do ICMS. Contudo os três modelos iniciaram uma tendência de crescimento já no mês de julho, enquanto que os dados originais só tiveram esse comportamento a partir de setembro.

Esse padrão vai ao encontro de uma característica já conhecida nos trabalhos de previsão: as previsões tendem a ter piores resultados conforme se aumenta a distancia temporal entre o momento da previsão e o do período que se esta sendo previsto.

Já o modelo da SEFAZ-RJ sempre apresentou a pior aderência ao comportamento da série original quando comparado com os modelos aqui elaborados. Até julho a tendência do modelo foi parecida à dos dados originais, contudo de forma significativamente superestimada. Entre agosto e dezembro observou-se irregularidade dos dados com quedas e altas não observadas na série efetiva.

Respondendo de forma objetiva a pergunta central deste trabalho, tem-se que a especificação multivariada VEC (ICMS, IBC-BR, PMC-RJ) foi a que apresentou a melhor acurácia preditiva quando comparada com as melhores versões univariada e

de combinação. Tal afirmativa se baseia fortemente no resultado do Erro % anual de previsão do modelo VEC (0,74%), dado que as quatro demais estatísticas de erro tiveram resultados ligeiramente inferiores ao método da combinação. Além do mais, vale lembrar que esse modelo VEC passou nos principais testes de análise de resíduos.

## 6 Conclusão

Este trabalho teve como objetivo comparar a acurácia preditiva do modelo oficial de previsão do ICMS do Estado do Rio de Janeiro, aquele utilizado pela SEFAZ-RJ com as outras especificações econométricas aqui estipuladas. A motivação principal para isso foi a piora dos resultados de previsão do modelo oficial nos anos de 2015 e 2016 (erros anuais de, respectivamente, 14,75% e 12,64%).

Entre os modelos univariados testados o escolhido como melhor foi o SARIMA(0,1,1)(1,1,0)<sub>12</sub>. Tal modelo conseguiu conciliar as propriedades estatísticas desejáveis aos resultados de previsão.

Esse modelo apresentou significância estatística em todos os seus parâmetros, teve o maior coeficiente de determinação e os menores valores para os critérios de informação entre os modelos considerados. Além disso, passou em todos os testes de análise dos resíduos.

Ademais, obteve os menores valores das estatísticas de erro, ao lado do modelo SARIMA(1,1,0)(1,1,0)<sub>12</sub>, para as quatro das cinco estatísticas consideradas. Na estatística restante, a do critério erro % anual, o modelo escolhido obteve resultado não distante do melhor modelo, o SARIMA(0,1,1)(0,1,0)<sub>12</sub>.

Na análise multivariada, o modelo escolhido foi o VEC que levou em conta, além do ICMS, o IBC-RJ e a PMC-RJ. Além de apresentar resultados satisfatórios na análise dos resíduos, esse modelo esteve com outros dois modelos entre os melhores em quatro das cinco estatísticas de erro consideradas. Na outra estatística, a do Erro % Anual, o resultado do modelo escolhido foi muito melhor que a dos outros dois, com erro de apenas 0,74%.

Já em relação aos modelos de combinação, à exceção do método *Least Squares Weights*, todos os três restantes apresentaram estatísticas de erro menores do que as do melhor univariado. Padrão parecido, porém menos intenso, foi observado quando analisado junto ao melhor método multivariado. A única diferença é que os métodos de combinação não superaram o melhor multivariado quando comparados com as estatísticas de Erro % anual.

Na análise final dos melhores modelos escolhidos para cada especificação (univariada, multivariada e combinação), tem-se que o modelo multivariado VEC (ICMS, IBC-BR, PMC-RJ) foi o que apresentou a melhor acurácia preditiva quando comparado com as melhores versões univariada e de combinação. Isso se justificou principalmente pelo seu Erro % anual de previsão de apenas 0,74%, dado que, para as quatro demais estatísticas de erro, os resultados do modelo multivariado foram ligeiramente piores do que as do melhor modelo dado pelos métodos de combinação. Além do mais, vale lembrar que esse modelo VEC passou nos principais testes de análise de resíduos.

## REFERÊNCIAS

AZEVEDO, R.R., SILVA, J.M., GATSIOS, R.C. Análise crítica dos modelos de previsão de série temporal com base no ICMS estadual. Revista de Gestão, Finanças e Contabilidade, v. 7, n. 1, p. 164-184, 2017.

BALCILAR, M., GUPTA, R., MAJUMDAR, A., MILLER, S. Forecasting Nevada gross gaming revenue and taxable sales using coincident and leading employment indexes. Discussion Paper Series, Department of Economics, Eastern Mediterranean University. 2015.

BATES, J. M., GRANGER, C. W. J. The combination of forecasts. Operational Research Quarterly, v. 20, p. 451-468, 1969.

BENELLI, F.C. Previsão da Receita Tributária Federal por Base de Incidência. 2013. 144 f. Dissertação (Mestrado Profissional em Economia do Setor Público), Universidade de Brasília, Brasília, 2013.

Boletim de Transparência Fiscal – 6º Bimestre de 2015. Secretaria de Estado De Fazenda do Rio de Janeiro - SEFAZ-RJ. 2015.

BOX, G. E. P., JENKINS, G. M. Time series analysis forecasting and control. San Francisco: Holden- Day, 1976. Edição revisada.

BRAM, J., HAUGHWOUT A. F., ORR J., RICH R., ROSEN R. The Linkage between Regional Economic Indexes and Tax Bases: Evidence from New York. Federal Reserve Bank of New York Staff Reports, n. 188, 2004.

BRASIL. Lei Complementar, n. 101, de 04 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Diário Oficial da União, Brasília, DF, 05 de maio de 2000b.

BRASIL. Ministério da Fazenda. Secretaria do Tesouro Nacional. Receitas Públicas: manual de procedimentos. Brasília: STN/Coordenação-Geral de Contabilidade, 2007. 233 p.

BRENDER, A., NAVON, G. Predicting government tax revenues and analyzing forecast uncertainty. *Israel Economic Review*, vol. 7, N. 2, 81–111, 2010.

BUETTNER, T., KAUDER, B. Revenue forecasting practices differences across countries and consequences for forecasting performance. *Ceslfo Working Paper* n. 2628, Category 1: Public Finance. 2010.

CAMARGOS, A.A.B. Modelos de Previsão da Arrecadação Tributária do Estado de São Paulo: ICMS, IPVA, ITCMD e Taxas. (Mestrado Profissional – IBMEC São Paulo). São Paulo. 2008.

CAMPOS, C.V.C. Previsão da Arrecadação de Receitas Federais: Aplicações de Modelos de Séries Temporais para o Estado de São Paulo. Dissertação (Mestrado – Universidade de São Paulo – Ribeirão Preto – USP-RP). Ribeirão Preto. 2009.

Carga Tributária no Brasil 2015: Análise por Tributos e Bases de Incidência. Centro de Estudos Tributários e Aduaneiros – Secretaria da Receita Federal do Brasil – Ministério da Fazenda. 2015.

CASTANHO, B.J.S. Modelos para previsão de receitas tributárias: o ICMS do Estado do Espírito Santo. 2011. 119 f. Dissertação (Mestrado em Economia) – Centro de Ciências Jurídicas e Econômicas, Universidade Federal do Espírito Santo, Espírito Santo, 2011.

CLEMENTS, M. P., HENDRY, D.F. An Overview of Economic Forecasting. In: *A Companion to Economic Forecasting*. Blackwell Publishing. 2004.

CORVALÃO, E. D. Previsão da Arrecadação do ICMS em Santa Catarina: Aplicação da Abordagem Geral para Específico em Modelos Dinâmicos. 2002. 78 f.

Dissertação (Programa de pós-graduação em engenharia de produção) – Centro Tecnológico, Universidade Federal de Santa Catarina, Santa Catarina, 2002.

COSTA, E.A.A. Fatores institucionais que influenciam a previsão das receitas orçamentárias: um estudo de caso dos governos estaduais brasileiros. 2011. 103 f. Dissertação (Mestrado em Administração) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação – FACE, Universidade de Brasília, Brasília, 2011.

ENDERS, W. Applied econometric time series. 1.ed. New Jersey: Wiley, 1995.

FAVERO, C. A., MARCELLINO, M. Modelling and Forecasting Fiscal Variables for the Euro Area. Oxford Bulletin of Economics and Statistics, 67: 755–783, 2005.

GRANGER, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Method. Econometrica, 37 (3): 424–438, 1969.

GUARAGNA, P.R.S., MELLO, M.R. Um Modelo de Previsão para a Arrecadação do ICMS. Brasília: ESAF, 2002, 46 p. Monografia premiada em 2º lugar no VII Prêmio Tesouro Nacional – 2002. Tributação, Orçamentos e Sistemas de Informação sobre Administração Financeira Pública.

HAMILTON, J.D. Time Series Analysis. Princeton University Press. 1994.

HYNDMAN, R.J., ATHANASOPOULOS, G. Forecasting: principles and practice. OTexts: Melbourne, Austrália. <http://otexts.org/fpp/>. 2013.

JOHANSEN, S. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, 12: 231–254, 1988.

KUBO, S.H. Previsão de arrecadação tributária baseada em um método de otimização de portfólio para a combinação de previsões. 2014. 275 f. Tese (Doutorado em Administração) – Departamento de Administração, Faculdade de

Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2014.

LEAL, T., PEREZ, J.J., TUJULA, M., VIDAL, J. Fiscal Forecasting: Lessons from the Literature and Challenges. Working Papers Series, n. 843. Banco Central Europeu. 2007.

LIMA, E. C. R., FERREIRA, L. R., HONAISSER, V. H. M. B. Projeção das Receitas Correntes do Estado do Rio de Janeiro: Aplicações de Modelos Univariados e Multivariados. Syn(thesis) , v. 6, n. 2, p. 207-233, 2013.

LUTKEPOHL, H., KRATZIG, M. Applied Time Series Econometrics. Cambridge University Press, 2014.

MARTIN, V., HURN, S., HARRIS, D. Econometric modelling with time series: specification, estimation and testing. Cambridge University Press, 2012.

MELO, B. S. V. de. Modelo de Previsão para a Arrecadação Tributária. (Mestrado – Universidade de Brasília - UnB). Brasília. 2001.

MENDONÇA, M. J., MEDRANO, L. A. Aplicação de um modelo fatorial dinâmico para previsão da arrecadação tributária no Brasil. Rio de Janeiro: Ipea, 2015. (Texto para Discussão, n. 2064).

MENDONÇA, M. J., MEDRANO, L. A. Um Modelo de Combinação de Previsões para a Arrecadação de Receita Tributária no Brasil. Rio de Janeiro: Ipea, 2016. (Texto para Discussão, n. 2186).

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

MENDONÇA, M. J.; MEDRANO, L. A. T; SACHSIDA, A. Um modelo econométrico para previsão de receita tributária no Brasil. *Economia Aplicada*, v. 17, n. 2, p. 295-329, 2013.

PECEGUINI, E.E. *Análise Comparativa de Métodos de Previsão Aplicados à Arrecadação do ICMS - Estado de São Paulo. (Aperfeiçoamento/Especialização em Economia do Setor Público - Fundação Instituto de Pesquisas Econômicas)*. São Paulo. 2001.

PESSOA, F. DE M. C.; CORONEL, D. A.; LIMA, J. E. DE. Previsão de arrecadação de ICMS para o estado de Minas Gerais: uma comparação entre modelos ARIMA e ARFIMA. *Revista Brasileira de Gestão e Desenvolvimento Regional*, v. 9, n. 2, p. 47–64, 2013.

SIQUEIRA, M. L. *Modelos de Séries Temporais para a Previsão da Arrecadação Tributária Federal. (Mestrado – Universidade Federal de Pernambuco)*. Pernambuco. 2002.

TIMMERMANN, A. Forecast combinations. In: ELLIOTT, G.; GRANGER, C.; TIMMERMANN, A. (Ed.). *Handbook of forecasting*. 1st ed. Amsterdam: Elsevier, 2006. v. 1. p. 135-196.

WILLIAMS, D.W., KAVANAGH, S.C. Local government revenue forecasting methods: competition and comparison. *Journal of public budgeting, accounting & financial management*, 28 (4), 488-526, 2016.

WILLOUGHBY, K.G., GUO, H. The State of the Art: Revenue Forecasting in US. State Governments. Em Jinping Sun e Thomas Lynch (Eds.) “Government Budget Forecasting: Theory and Practice”. *Public Administration and Public Policy*, n. 142, p. 27-42, 2008.