

Francisco Alves de Oliveira Júnior

Combinação de Previsões: O Caso do ICMS no Estado de Minas Gerais

Brasil

2014

Francisco Alves de Oliveira Júnior

Combinação de Previsões: O Caso do ICMS no Estado de Minas Gerais

Monografia apresentada ao Programa de Especialização em Estatística do Departamento de Estatística da Universidade Federal de Minas Gerais como requisito para obtenção do título de Especialista em Estatística

Universidade Federal de Minas Gerais – UFMG

Departamento de Estatística

Especialização

Orientador: Ela Mercedes Medrano de Toscano

Brasil

2014

Francisco Alves de Oliveira Júnior

Combinação de Previsões: O Caso do ICMS no Estado de Minas Gerais/
Francisco Alves de Oliveira Júnior. – Brasil, 2014-

71 p. : il. (algumas color.) ; 30 cm.

Orientador: Ela Mercedes Medrano de Toscano

Monografia (Especialização) – Universidade Federal de Minas Gerais – UFMG
Departamento de Estatística
Especialização, 2014.

1. Previsão. 2. Séries Temporais. 3. Arrecadação ICMS. 4. Combinação de
Previsões. I. Ela Mercedes Medrano de Toscano. II. Universidade Federal de
Minas Gerais. III. Departamento de Estatística. IV. Combinação de Previsões: O
Caso do ICMS no Estado de Minas Gerais

CDU 02:141:005.7

Francisco Alves de Oliveira Júnior

Combinação de Previsões: O Caso do ICMS no Estado de Minas Gerais

Monografia apresentada ao Programa de Especialização em Estatística do Departamento de Estatística da Universidade Federal de Minas Gerais como requisito para obtenção do título de Especialista em Estatística

Trabalho aprovado. Brasil, 10 de Abril de 2014:

Ela Mercedes Medrano de Toscano
Orientador

Aureliano Angel Bressan
Convidado 1

Luis Alberto Toscano
Convidado 2

Brasil
2014

Agradecimentos

A minha orientadora, Ela, que despertou meu interesse por séries temporais.

Ao professor Marcelo, pelo incentivo ao uso do R, que foi indispensável para a realização deste trabalho.

A Rose, pelo suporte durante todo o curso.

A minha namorada, Samira, pela presença e alegria durante os sábados infundáveis.

A minha família e amigos, que proporcionaram os momentos de descanso fundamentais para a finalização deste trabalho.

Resumo

Um dos principais determinantes para a implementação de políticas públicas pelos entes federados é a determinação da disponibilidade de recursos. No caso dos estados brasileiros destaca-se a arrecadação proveniente do Imposto sobre Circulação de Mercadorias e Serviços (ICMS). O objetivo deste trabalho é utilizar a combinação de previsão como forma de melhorar a acurácia das previsões de ICMS para o Estado de Minas Gerais obtidas por modelos univariados de séries temporais. Para isso foram calculadas previsões usando os modelos SARIMA e de Alisamento Exponencial que posteriormente foram combinadas mediante cinco procedimentos de combinação, a saber; combinação ótima, regressão linear, performance, média simples e mediana. Os resultados obtidos indicam que a combinação de previsão, seja por qual procedimento de combinação, consegue melhorar a performance da previsão de forma significativa.

Palavras-chaves: previsão. séries temporais. receita icms. combinação de previsões.

Abstract

A major determinant for the implementation of public policies by public agencies is determining the availability of resources. In the case of the Brazilian States, are noteworthy the collections of the VAT. The objective of this monograph is to use the combination of forecasts as a way of improving the accuracy of forecasts of VAT for the State of Minas Gerais obtained by univariate time series models. For this, forecasts of SARIMA models and exponential smoothing were made which were later combined by five different procedures, namely; optimum combination, linear regression, past performance, simple mean and median. The results obtained indicate that the combination of forecasts, for whatever combination procedure used, can improve the prediction performance significantly.

Key-words: forecasting. time series. vat revenue. combination of forecasts

Lista de ilustrações

Figura 1 – Exemplo de um Ruído Branco Gaussiano	25
Figura 2 – Função de Autocorrelação Parcial de um Ruído Branco Gaussiano	26
Figura 3 – Evolução Nominal do ICMS no período 2003-2012	43
Figura 4 – Arrecadação ICMS sem tendência	45
Figura 5 – Arrecadação ICMS sem tendência - Série Padronizada	45
Figura 6 – Identificação dos Modelos	46
Figura 7 – Ajuste dos Modelos	48
Figura 8 – Erros de Previsão Padronizados - Fora da Amostra - Modelos Originais	51
Figura 9 – Erros de Previsão Padronizados - Fora da Amostra - Combinações	51
Figura 10 – Erros de Previsão Padronizados - Fora da Amostra - Melhor Combinação	52
Figura 12 – Intervalo de Previsão de 95% - (Fora da Amostra)	60
Figura 14 – Intervalo de Previsão de 95% - (Fora da Amostra)	62
Figura 16 – Intervalo de Previsão de 95% - (Fora da Amostra)	64

Lista de tabelas

Tabela 1 – Estatísticas Descritivas Arrecadação Nominal ICMS - Em Milhões . . .	44
Tabela 2 – Estatísticas de Estimação dos Modelos (jan/2003 – dez/2012, n=120))	47
Tabela 3 – Estatísticas de Ajuste dos Modelos (jan/2003 – dez/2012, n=120) . .	47
Tabela 4 – Acurácia Fora da Amostra (jan/2013 – dez/2013, n=12) - Em Milhões	50
Tabela 5 – Razão entre MAPE e MAPE Método Ingênuo (jan/2013 – dez/2013, n=12)	50

Sumário

	Introdução	18
1	Séries Temporais	19
1.1	Estacionariedade	19
1.2	Operador de Defasagem	20
1.3	Exemplos de Processos Estocásticos	20
1.4	Função de Autocorrelação e Autocorrelação Parcial	24
1.4.1	Distribuição Amostral da Função de Autocorrelação	26
2	Métodos de Previsão	29
2.1	Suavização Exponencial	29
2.1.1	Suavização Exponencial Simples (SES)	29
2.1.2	Suavização Exponencial de Holt (SEH)	30
2.1.3	Suavização Exponencial de Holt-Winters (SEHW)	31
2.2	Análise de Intervenção	33
2.3	Acurácia das Previsões	34
2.4	Métodos Ingênuos de Previsão	35
2.5	Critérios de Seleção de Modelos	36
2.6	Combinação de Previsões	37
2.6.1	Procedimentos de Combinação	38
3	Metodologia	41
3.1	Base de Dados	41
3.2	Software	41
4	Resultados	43
4.1	Análise Exploratória	43
4.2	Ajuste dos Modelos	46
4.3	Comparação dos Modelos	49
	Conclusão	53
	Referências	55
	Apêndices	57
	APÊNDICE A – Gráficos de Diagnóstico	59
	APÊNDICE B – Script R	67

Introdução

A previsão de receitas públicas é o processo por meio do qual se estima o montante de recursos que serão arrecadados por um ente público em determinado período. A previsão é fundamental no processo orçamentário, apesar de não se confundir com este. Somente após a determinação da disponibilidade de recursos é possível iniciar a definição de quais objetivos serão perseguidos pelo governo por meio da implementação de políticas públicas.

Ao longo do ciclo orçamentário, que compreende a elaboração do Plano Plurianual, da Lei de Diretrizes Orçamentárias e da Lei Orçamentária Anual, o poder público se depara em diversos momentos com a necessidade de proceder as estimativas fiscais. A necessidade de um processo de previsão robusto foi tida como condição necessária para o equilíbrio das contas públicas pela Lei de Responsabilidade Fiscal, ao estabelecer que:

Art. 11. 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.

[...]

Art. 12. As previsões de receita observarão as normas técnicas e legais, considerarão os efeitos das alterações na legislação, da variação do índice de preços, do crescimento econômico ou de qualquer outro fator relevante e serão acompanhadas de demonstrativo de sua evolução nos últimos três anos, da projeção para os dois seguintes àquele a que se referirem, e da metodologia de cálculo e premissas utilizadas. (BRASIL, 2000).

No âmbito da federação brasileira, a principal fonte de receita para arcar com as despesas públicas são os impostos. No caso dos estados, em sua grande maioria, a principal arrecadação é a proveniente do Imposto sobre Operações relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual e Intermunicipal e de Comunicação (ICMS), previsto na Constituição Federal de 1988.

Atualmente o foco de acadêmicos e profissionais encontra-se nos aspectos quantitativos referentes a métodos e modelos. No entanto, alguns estudos começam a explorar vertentes organizacionais e processuais¹ que também impactam na acurácia e credibilidade das previsões. Uma vertente quantitativa ainda pouco explorada para a previsão de impostos no setor público é a combinação de previsões. Portanto, podemos enunciar o

¹ COSTA, E. A. de A. *Fatores Institucionais que Influenciam a Previsão das Receitas Orçamentárias: Um Estudo de Caso dos Governos Estaduais Brasileiros*. Dissertação (Mestrado) — Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2011. Disponível em: <repositorio.unb.br/bitstream/10482/9937/1/2011_EduardoAugustoAbreuCosta.pdf>.

problema de pesquisa como: A utilização da combinação de previsão é capaz de melhorar a acurácia da previsão do ICMS em relação a métodos univariados de séries temporais.

Especialmente com o advento da Lei de Responsabilidade Fiscal, esse problema se tornou mais importante tendo em vista que desvios de previsão de receitas de magnitude elevada são capazes de prejudicar a alocação de recursos durante a execução orçamentária. Isso decorre do disposto no art. 11 da Lei de Responsabilidade Fiscal, que estabelece:

Art. 9º Se verificado, ao final de um bimestre, que a realização da receita poderá não comportar o cumprimento das metas de resultado primário ou nominal estabelecidas no Anexo de Metas Fiscais, os Poderes e o Ministério Público promoverão, por ato próprio e nos montantes necessários, nos trinta dias subseqüentes, limitação de empenho e movimentação financeira, segundo os critérios fixados pela lei de diretrizes orçamentárias. (BRASIL, 2000)

Portanto, a consequência imediata de um processo de previsão incapaz de gerar estimativas com acurácia é o contingenciamento de recursos, impactando sobremaneira a execução das políticas públicas.

Portanto, o objetivo deste trabalho é apresentar procedimentos de combinação que sejam capazes de melhorar a acurácia da previsão de ICMS no Estado de Minas Gerais.

Além desta introdução, esta monografia possui mais 5 capítulos. No capítulo 1 são apresentados os conceitos básicos relativos a séries temporais. No capítulo 2 são apresentados os métodos de previsão utilizados para gerar as estimativas, bem como procedimentos que podem ser utilizados para melhorar o ajuste dos modelos, como a análise de intervenção, e a acurácia das previsões geradas, como a combinação de previsão. No capítulo 3 são apresentadas as fontes das bases de dados e os softwares utilizados na monografia. No capítulo 4 são apresentados e discutidos os resultados encontrados com os ajustes dos modelos e as previsões geradas. Por fim, o último capítulo apresenta as conclusões e sugestões para estudos futuros.

1 Séries Temporais

Para realizarmos a previsão de séries temporais, a discussão de alguns conceitos básicos é essencial. O primeiro deles é o de processo estocástico. Um processo estocástico é um conjunto de variáveis aleatórias ordenadas no tempo. Se estamos diante de variáveis aleatórias contínuas, utilizamos a representação $Y(t)$, caso sejam discretas, utilizamos Y_t . Um processo estocástico pode então ser representado por

$$Y_t = \{Y_1, Y_2, Y_3, \dots, Y_n\}$$

Uma série temporal nada mais é do que uma realização única, das diversas possíveis, do processo estocástico em estudo. A partir desse ponto, a estratégia de análise será a mesma utilizada para dados transversais. Neste, utilizamos as informações de uma amostra para fazermos inferência sobre a população. Na análise de séries temporais, utilizamos as informações da realização para fazer inferência sobre o processo estocástico. (GUJARATI, 2004).

Nas próximas seções serão expostos os demais conceitos básicos que irão permitir a construção de modelos para previsões de séries temporais. Conceitos como estacionariedade, processo ruído branco e função de autocorrelação serão utilizados extensivamente no restante deste trabalho.

1.1 Estacionariedade

Para que seja possível fazer inferência sobre o processo estocástico, é necessário que o mesmo mantenha suas propriedades estatísticas constantes ao longo do tempo. Ou seja, o processo precisa estar em um estado de equilíbrio estatístico. (BOX; JENKINS; REINSEL, 2008). A essa característica dá-se o nome de estacionariedade. A justificativa para a condição necessária de estacionariedade para a previsão de séries temporais foi bem explicada por Pokorny citado por Gujarati (2004), ao dizer que:

The objective of B–J [Box–Jenkins] is to identify and estimate a statistical model which can be interpreted as having generated the sample data. If this estimated model is then to be used for forecasting we must assume that the features of this model are constant through time, and particularly over future time periods. Thus the simple reason for requiring stationary data is that any model which is inferred from these data can itself be interpreted as stationary or stable, therefore providing valid basis for forecasting. (GUJARATI, 2004, p. 840).

Para fins desse trabalho, e da grande maioria das aplicações, a estacionariedade fraca é suficiente. Um processo estocástico é dito fracamente estacionário se sua média e

variância são constantes ao longo do tempo e a covariância entre dois períodos depende apenas da distância entre eles, e não dos períodos específicos em que a covariância foi computada. Ou seja, se Y_t é um processo estocástico fracamente estacionário, então

$$\begin{aligned} E(Y_t) &= \mu \\ \text{Var}(Y_t) &= E[(Y_t - \mu)^2] = \sigma^2 \\ \text{Cov}(Y_t, Y_{t-k}) &= E[(Y_t - \mu)(Y_{t-k} - \mu)] = \gamma_k \end{aligned}$$

1.2 Operador de Defasagem

Ao trabalharmos com séries temporais, o operador de defasagem B é conveniente. Aplicado sobre uma série temporal Y_t , o operador “defasa” a mesma em um período, ou seja $BY_t = Y_{t-1}$. Generalizando para n aplicações temos que $B^n Y_t = Y_{t-n}$

Uma das vantagens do operador B é sua capacidade de expressar a operação de diferenciação, definida como

$$\Delta Y_t = Y_t - Y_{t-1} = (1 - B)Y_t$$

Em geral, uma diferenciação de ordem d pode ser expressa como

$$\Delta^d Y_t = (1 - B)^d Y_t$$

O operador de defasagem também permite uma notação sintética para o operador autoregressivo de ordem p e para o operador de médias móveis¹ de ordem q , representados respectivamente por

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q \end{aligned}$$

1.3 Exemplos de Processos Estocásticos

Nesta seção vamos apresentar exemplos de processos estocásticos.

Ruído Branco

Um dos processos estocásticos mais utilizados na análise de séries temporais é o chamado ruído branco. Ele denota uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância constante. Neste caso, o ruído branco é dito estritamente estacionário. Como neste trabalho somente será necessário a

¹ Os processos estocásticos autoregressivos e de médias móveis serão apresentados na seção 1.3.

estacionariedade fraca, podemos substituir o requisito de independência pelo de ausência de autocorrelação. Por fim, caso as variáveis aleatórias sigam a distribuição normal, diremos tratar-se de um ruído branco gaussiano, ou seja a_t é $NID(0, \sigma^2)$. Durante o processo de modelagem, um dos principais objetivos é encontrar resíduos que se comportem como um processo ruído branco.

Passeio Aleatório

Um processo estocástico é chamado de passeio aleatório caso possa ser expresso como

$$Y_t = Y_{t-1} + a_t$$

Onde a_t é um ruído branco. Ou seja, o valor de Y_t no período t é igual ao valor do período $t - 1$, mais um termo de erro aleatório. De forma geral temos que

$$\begin{aligned} Y_t &= Y_0 + \sum a_t \\ E(Y_t) &= Y_0 \\ Var(Y_t) &= t\sigma^2 \end{aligned}$$

Três pontos merecem destaque em relação ao passeio aleatório. Em primeiro lugar, é possível notar que Y_t é igual a um valor inicial mais uma soma de erros aleatórios, indicando que o impacto de um erro aleatório permanece para sempre na trajetória da série temporal em questão, isto é, ela possui memória infinita. Em segundo lugar, a variância do passeio aleatório aumenta com o tempo, indicando que trata-se de um processo estocástico não-estacionário. Por fim, apesar da não-estacionariedade do passeio aleatório, sua primeira diferença é estacionária, tendo em vista que $\Delta Y_t = Y_t - Y_{t-1} = a_t$.

Autoregressivo (AR)

Um processo estocástico é chamado de autoregressivo de ordem 1, $AR(1)$, caso possa ser expresso como

$$Y_t = \phi_1 Y_{t-1} + a_t$$

Onde a_t é um ruído branco. Ou seja, o valor de Y_t no período t é igual a uma proporção ϕ do seu valor no período $t - 1$ mais um termo de erro aleatório. Utilizando o operador de defasagem B , o mesmo processo pode ser expresso como

$$(1 - \phi_1 B)Y_t = a_t$$

Para que os modelos autoregressivos cumpram a condição de estacionariedade, é necessário que seus parâmetros fiquem restritos a intervalos específicos. Para um $AR(1)$, $-1 < \phi_1 < 1$, e para um $AR(2)$, $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

O processo $AR(1)$ pode ser generalizado para uma dada ordem p

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \cdots + \phi_p Y_{t-p} + a_t$$

Ou, utilizando o operador de defasagem, podemos escrever

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p) Y_t = a_t$$

$$\phi(B) Y_t = a_t$$

O nome autoregressivo é derivado da similaridade do processo $AR(p)$ com uma regressão múltipla, onde os valores defasados de Y_t fazem o papel de variáveis preditoras.

Médias Móveis (MA)

Um processo estocástico é chamado de médias móveis de ordem 1, $MA(1)$, caso possa ser expresso como

$$Y_t = a_t - \theta_1 a_{t-1}$$

Onde a_t é um ruído branco. Ou seja, o valor de Y_t no período t é igual a uma proporção do erro a no período $t - 1$ mais um erro a_t no período t . Utilizando o operador de defasagem B , o mesmo processo pode ser expresso como

$$Y_t = (1 - \theta_1 B) a_t$$

O processo $MA(1)$ pode ser generalizado para uma dada ordem q

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \cdots - \theta_q a_{t-q}$$

Ou, utilizando o operador de defasagem, podemos escrever

$$Y_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \cdots - \theta_q B^q) a_t$$

$$Y_t = \theta(B) a_t$$

Apesar do processo AR possuir uma interpretação mais simples, o processo MA também pode ser visto como uma regressão múltipla, onde o valor de Y_t é explicado pelos erros defasados.

Qualquer processo estacionário $AR(p)$ pode ser escrito como um processo $MA(\infty)$. O resultado contrário, ou seja, a possibilidade de escrever qualquer processo $MA(q)$ como um processo $AR(\infty)$ depende da condição de invertibilidade.

Para que os modelos de médias móveis cumpram a condição de invertibilidade, é necessário que seus parâmetros fiquem restritos a intervalos específicos. Para um $MA(1)$, $-1 < \theta_1 < 1$, e para um $MA(2)$, $-1 < \theta_2 < 1$, $\theta_1 + \theta_2 > -1$, $\theta_1 - \theta_2 < 1$.

Autoregressivo e Médias Móveis (ARMA)

Os processos *AR* e *MA* podem ser combinados gerando os chamados processos autoregressivos médias móveis (*ARMA*). Um processo estocástico autoregressivo de ordem 1 e médias móveis de ordem 1, *ARMA*(1, 1), pode ser expresso como

$$Y_t = \phi_1 Y_{t-1} + a_t - \theta_1 a_{t-1}$$

Ou, utilizando o operador de defasagem, podemos escrever

$$(1 - \phi_1 B)Y_t = (1 - \theta_1 B)a_t$$

O processo *ARMA*(1, 1) pode ser generalizado para uma dada ordem p e q

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Ou, utilizando o operador de defasagem, podemos escrever

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)a_t$$

$$\phi(B)Y_t = \theta(B)a_t$$

Autoregressivo e Médias Móveis Integrado (ARIMA)

Os modelos *AR* e *MA* apresentados são baseados na premissa de que as séries temporais em análise são fracamente estacionárias. No entanto, é comum, especialmente em séries econômicas e financeiras, a existência de séries temporais não-estacionárias, também chamadas de integradas.

Caso a série em questão possa ser transformada em estacionária por meio de uma diferenciação, como o passeio aleatório apresentado na seção 1.3, ela é dita integrada de ordem d , e representada por $I(d)$.

Portanto, se precisamos diferenciar uma série temporal d vezes para torná-la estacionária e depois utilizamos um modelo *ARMA*(p, q), dizemos que a série original é um *ARIMA*(p, d, q). Ou seja, é um processo autoregressivo integrado médias móveis de ordem p no componente autoregressivo, d no componente integrado e q no componente médias móveis.

Um processo estocástico *ARIMA*(1, 1, 1) pode ser expresso como

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + a_t - \theta_1 a_{t-1}$$

Ou, utilizando o operador de defasagem, podemos escrever

$$(1 - \phi_1 B)(1 - B)Y_t = (1 - \theta_1 B)a_t$$

O processo $ARIMA(1, 1, 1)$ pode ser generalizado para uma dada ordem p , d e q

$$\Delta^d Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \dots + \phi_p \Delta Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Ou, utilizando o operador de defasagem, podemos escrever

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t$$

$$\phi(B)(1 - B)^d Y_t = \theta(B) a_t$$

Autoregressivo e Médias Móveis Integrado Sazonal (SARIMA)

A complexidade final que pode ser introduzida aos modelos $ARIMA$ visa tratar o problema da sazonalidade. Isso ocorre quando existe autocorrelação significativa entre defasagens sazonais, ou seja, múltiplos de um período qualquer s . Neste caso, tem-se um modelo $ARIMA$ sazonal ($SARIMA$). Para expressarmos este tipo de modelo, precisamos estender a notação utilizada nos modelos $ARIMA(p, d, q)$ para $SARIMA(p, d, q)(P, D, Q)_s$, onde (p, d, q) representa a parte não sazonal, (P, D, Q) a parte sazonal e s uma sazonalidade de s períodos.

Utilizando o operador de defasagem, podemos expressar um processo estocástico $SARIMA(1, 1, 1)(1, 1, 1)_{12}$ como

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})a_t$$

1.4 Função de Autocorrelação e Autocorrelação Parcial

Na análise de séries temporais, nosso interesse principal é modelar a estrutura de dependência que naturalmente está presente em observações ordenadas no tempo. As funções de autocorrelação e autocorrelação parcial são utilizadas extensivamente para auxiliar na consecução desse objetivo. A correlação entre duas variáveis aleatórias Y_t e Y_{t-k} separadas por k períodos é chamada de autocorrelação de ordem k , e definida como

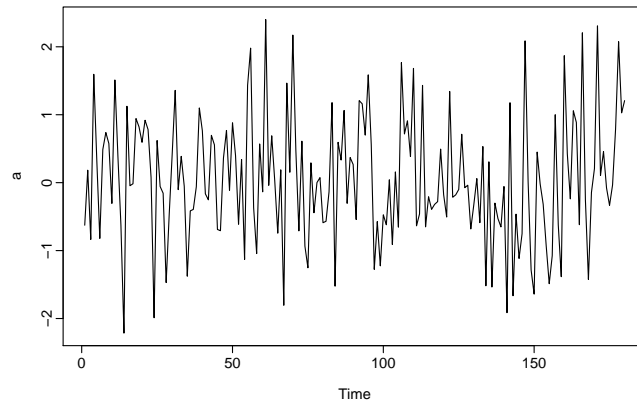
$$\rho_k = \frac{E[(Y_t - \mu)(Y_{t-k} - \mu)]}{\sqrt{E[(Y_t - \mu)^2]E[(Y_{t-k} - \mu)^2]}}$$

A partir da suposição de estacionariedade temos que $E[(Y_t - \mu)^2] = E[(Y_{t-k} - \mu)^2]$, portanto, $\sqrt{E[(Y_t - \mu)^2]E[(Y_{t-k} - \mu)^2]} = \sigma^2 = \gamma_0$, e a autocorrelação de ordem k pode ser expressa por

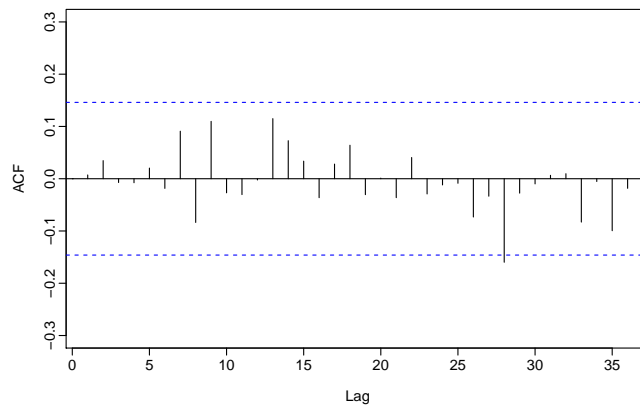
$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Se plotarmos em um gráfico ρ_k contra k , obtemos a função de autocorrelação (FAC), também chamado de correlograma. A figura 1 apresenta um processo ruído branco gaussiano e sua respectiva função de autocorrelação:

Figura 1 – Exemplo de um Ruído Branco Gaussiano



(a) Ruído Branco Gaussiano



(b) Função de Autocorrelação

Fonte: Elaboração do Autor

Até agora tratamos da autocorrelação teórica de um processo estocástico. Na realidade, a função de autocorrelação terá de ser estimada a partir dos dados. Para isso utilizamos a variância e covariância amostral. Ou seja

$$r_k = \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

Onde

$$\hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^{n-k} (y_t - \bar{y})(y_{t-k} - \bar{y}_{t-k})$$

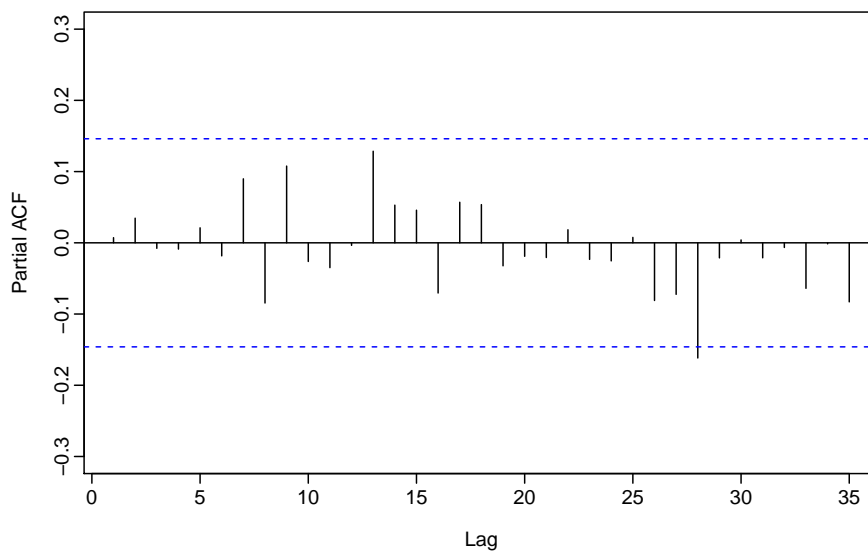
Para efetuar a análise da estrutura de dependência da série em estudo, as informações da função de autocorrelação precisam ser complementadas por um conceito associado, o de autocorrelação parcial. Lembre-se que a função de autocorrelação mostra a relação linear entre Y_t e Y_{t-k} para valores distintos de k . No entanto, se Y_t e Y_{t-1} são correlacionadas, Y_{t-1} e Y_{t-2} também o serão. Logo, Y_t e Y_{t-2} serão correlacionadas em virtude do fato de que ambas possuem associação com Y_{t-1} .

A autocorrelação parcial visa superar esse problema. Ela mensura a correlação linear entre Y_t e Y_{t-k} depois de removidos os efeitos das outras defasagens $1, 2, 3, \dots, k-1$. A autocorrelação parcial na primeira defasagem será igual a primeira autocorrelação, tendo em vista que não existem defasagens entre elas que necessitam ser removidas. As outras autocorrelações parciais podem ser estimadas a partir de

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_k Y_{t-k} \quad (1.1)$$

Onde a autocorrelação parcial de ordem k é simplesmente o coeficiente b_k na equação 1.1. Se plotarmos em um gráfico b_k contra k , obtemos a função de autocorrelação parcial (FACP). A figura 2 apresenta a função de autocorrelação parcial do processo ruído branco gaussiano mostrado na figura 1a:

Figura 2 – Função de Autocorrelação Parcial de um Ruído Branco Gaussiano



Fonte: Elaboração do Autor

1.4.1 Distribuição Amostral da Função de Autocorrelação

Nosso interesse em modelar a estrutura de dependência de uma série temporal precisa incluir a possibilidade de aleatoriedade, ou seja, queremos captar apenas as autocorrelações estatisticamente diferentes de zero. Para isso utilizamos a distribuição amostral da função de autocorrelação.

No caso de um processo ruído branco, e para valores suficientemente grandes de amostra, a função de autocorrelação tende a ser independente e normalmente distribuída com média 0 e variância $\frac{1}{N}$. A partir dessas informações é possível testar a hipótese de que a autocorrelação de ordem k é estatisticamente diferente de zero.

Estratégia distinta também pode ser usada para testar se as autorrelações são significativas. Ao invés de avaliar cada autocorrelação separadamente, podemos testar a hipótese de que as autocorrelações até determinada defasagem são estatisticamente diferentes de zero. Testes deste tipo são conhecidos como testes de *Portmanteau*. A primeira forma de realizar o teste de *Portmanteau* é com a estatística Q desenvolvida por Box-Pierce. Ela é definida como

$$Q = n \sum_{k=1}^h \hat{\rho}_k^2$$

Onde n representa o tamanho da amostra e h a defasagem. Em amostras grandes, a estatística Q segue aproximadamente uma distribuição qui-quadrado com $h - m$ graus de liberdade, onde m é o número de parâmetros do modelo estimado. A segunda forma de realizar o teste é por meio da estatística de Ljung-Box. Ela é definida como

$$LB = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n - k}$$

A estatística de Ljung-Box também segue uma distribuição qui-quadrado com $h - m$ graus de liberdade. No entanto, em amostras pequenas ela possui propriedades estatística mais desejáveis.

2 Métodos de Previsão

2.1 Suavização Exponencial

Os métodos de suavização exponencial são capazes de lidar com séries temporais que apresentam comportamentos variados, não deixando de serem conceitualmente claros, computacionalmente simples e possibilitarem a automatização do processo de previsão. Além disso, essa classe de métodos ainda possui um desempenho satisfatório em termos da acurácia das previsões geradas. Essas características justificam a elevada adoção destes métodos entre os profissionais da indústria de previsão, bem como o interesse de acadêmicos de refinarem a teoria estatística¹ por trás dos métodos de suavização exponencial.

As previsões desta classe de métodos compartilham o fato de serem geradas a partir de médias ponderadas de observações passadas. Os pesos utilizados para cada observação decaem exponencialmente com o tempo, fato que justifica o nome suavização exponencial. A racionalidade por trás da ponderação deriva do fato intuitivo de que quanto mais distante uma observação se encontra, menos informação ela deve possuir sobre o comportamento futuro da série.

Nesta seção serão apresentados três métodos adequados a séries temporais com padrões distintos. A Suavização Exponencial Simples (SES) para séries com nível local constante, Suavização Exponencial de Holt (SEH) para séries com tendência e Suavização Exponencial de Holt-Winters (SEHW) para séries com tendência e sazonalidade, seja a última aditiva ou multiplicativa.

2.1.1 Suavização Exponencial Simples (SES)

Conforme indicado no próprio nome, a suavização exponencial simples é o método mais simples de suavização exponencial. Sua utilização é adequada quando a série temporal em estudo não possui tendência ou sazonalidade. A previsão para o período $t + 1$, feita no período t pode ser expressa por

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \quad (2.1)$$

Ou seja, para fazer a previsão para o período $t + 1$, o método de suavização exponencial simples utiliza a previsão disponível para o último período, F_t , e ajusta esse valor com um percentual α do erro de previsão observado no período t , qual seja, $Y_t - F_t$. A consequência desse ajuste é que o erro de previsão passado corrige a previsão futura na

¹ HYNDMAN, R. J. et al. *Forecasting with Exponential Smoothing: The State Space Approach*. 1. ed. New York: Springer, 2008. 362 p.

direção oposta do erro observado, levando a um processo de previsão capaz de corrigir seus erros de forma automática. Para previsões com horizontes maiores que 1, assume-se que a função de previsão é horizontal, ou seja

$$F_{t+h} = F_{t+1}, \quad \text{para } h = 2, 3, \dots, n$$

As demais propriedades da suavização exponencial simples podem ser melhor compreendidas ao reescrevermos a equação 2.1 como

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t \quad (2.2)$$

e expandirmos recursivamente F_t em 2.2, obtendo

$$F_{t+1} = \alpha Y_t + \alpha(1 - \alpha)(F_{t-1}) + \alpha(1 - \alpha)^2(F_{t-2}) + \dots + \alpha(1 - \alpha)^{t-1}F_t + (1 - \alpha)^t F_1$$

Com essa representação fica claro que a previsão do método de suavização exponencial simples é uma média ponderada de todas as observações, com os pesos determinados pela constante de suavização α . A referida constante desempenha um papel fundamental nas previsões. Conforme observado por [Morettin e Tolo \(2006\)](#), valores pequenos de α geram previsões finais mais estáveis, tendo em vista que pesos maiores serão dados às observações passadas, fazendo com que flutuações aleatórias no presente exerçam menor impacto no cálculo das previsões.

Dois problemas principais decorrem da utilização do método de suavização exponencial simples, quais sejam, a determinação do valor ótimo para a constante de suavização α , e a determinação do valor de F_1 , conhecido como problema da inicialização.

Em relação a constante de suavização α , [Morettin e Tolo \(2006\)](#) argumentam que um procedimento objetivo de seleção é encontrar um valor de α que minimize

$$SQE = \sum_{t=1}^n (Y_t - F_t)^2 = \sum_{t=1}^n e_t^2$$

ou seja, o objetivo é minimizar a soma dos quadrados dos erros (SQE) da previsão.

Já em relação ao valor inicial F_1 , o problema existe quando o valor de α é pequeno e a série temporal é relativamente curta. Isso significa que o valor de inicialização vai influenciar de forma significativa e permanente as previsões posteriores.

2.1.2 Suavização Exponencial de Holt (SEH)

A suavização exponencial de Holt é uma extensão da suavização exponencial simples que permite a previsão de séries que apresentam tendência linear. Essa característica

é obtida por meio da utilização de outra constante de suavização, β , utilizada para suavização da tendência. As equações descritivas da suavização exponencial de Holt são

$$F_{t+h} = l_t + b_t h \quad (2.3)$$

$$l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.4)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (2.5)$$

onde F_{t+h} é a previsão para o período $t+h$, l_t é o nível suavizado no período t e b_t é a tendência suavizada no período t . Conforme observado em 2.3, a função de previsão não é mais horizontal, tendo em vista que possui uma tendência determinada pelo termo $b_t h$.

Os dois problemas encontrados na suavização exponencial simples persistem na suavização de Holt, quais sejam, a determinação dos valores ótimos para as constantes de suavização α e β , e o problema da inicialização, que agora deve determinar valores iniciais de l_1 e b_1 . Para a determinação das constantes de suavização, uma estratégia objetiva ainda é a minimização da soma dos quadrados dos erros de previsão. Já em relação aos valores iniciais de l_1 e b_1 , podem ser utilizados os parâmetros estimados do modelo de regressão linear simples aplicado as primeiras observações da série. (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).

2.1.3 Suavização Exponencial de Holt-Winters (SEHW)

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

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

onde F_{t+h} é a previsão para o período $t+h$, l_t é o nível suavizado no período t , b_t é a tendência suavizada no período t , s_t é a componente sazonal no período t e m representa o número de períodos da sazonalidade (e.g., número de meses em um ano).

As equações expostas acima tratam a sazonalidade de forma multiplicativa, ou seja, a sazonalidade é proporcional ao nível atual da série. Apesar de ser a forma mais usual

de sazonalidade, especialmente em séries econômicas e financeiras, também é possível modelar a sazonalidade aditiva, expressa como

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

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

$$\begin{aligned} l_m &= \frac{1}{m}(Y_1 + Y_2 + \dots + Y_m) \\ b_m &= \frac{1}{m}\left(\frac{Y_{m+1} - Y_1}{m} + \frac{Y_{m+2} - Y_2}{m} + \dots + \frac{Y_{m+m} - Y_m}{m}\right) \\ s_1 &= \frac{Y_1}{l_m}, s_2 = \frac{Y_2}{l_m}, \dots, s_m = \frac{Y_m}{l_m} \end{aligned}$$

2.2 Análise de Intervenção

Durante o processo de previsão de séries temporais é comum encontrarmos eventos singulares que provocam alterações no comportamento da série em estudo. Nesses casos, dois tipos principais de perguntas usualmente são feitas, quais sejam, assumindo a ocorrência de um evento em um instante t , existe evidência suficiente de que a mudança esperada ocorreu, e, em caso positivo, o que pode ser dito a respeito da natureza e da magnitude da mudança. (BOX; TIAO, 1975). Essas perguntas podem ser respondidas por meio da análise de intervenção.

Segundo Morettin e Tolo (2006), podemos entender uma intervenção como a ocorrência de um evento qualquer em um instante de tempo t conhecido. Essa ocorrência pode afetar temporária ou permanentemente a série em estudo. Usualmente, as intervenções são modeladas por meio de dois tipos principais de variáveis binárias, a chamada função degrau e a função impulso.

A função degrau é a forma mais simples de intervenção, onde é esperado que o evento vai gerar uma alteração súbita e permanente na série em estudo. Assumindo a ocorrência da intervenção no período u , podemos representar a variável binária como

$$X_t = \begin{cases} 0, & t < u \\ 1, & t \geq u \end{cases}$$

Por sua vez, a função impulso representa uma alteração súbita porém temporária no comportamento da série. Assumindo a ocorrência da intervenção no período u , a função impulso pode ser representada como

$$X_t = \begin{cases} 0, & t \neq u \\ 1, & t = u \end{cases}$$

2.3 Acurácia das Previsões

Existem na literatura diversas medidas de acurácia propostas para avaliar a performance de previsões². No entanto, o objetivo desta seção não é discutir os pontos positivos e negativos de cada uma delas, mas tão somente apresentar aquelas mais utilizadas para este fim.

Seja ϵ_t o erro de previsão definido como $\epsilon_t = Y_t - F_t$ e p_t o erro de previsão percentual definido como $p_t = 100(\epsilon_t/Y_t)$. Podemos definir então as seguintes medidas

$$\begin{aligned}\text{Erro Médio (ME)} &= \frac{\sum_{t=1}^n \epsilon_t}{n} \\ \text{Raiz do Erro Quadrático Médio (RMSE)} &= \sqrt{\frac{\sum_{t=1}^n \epsilon_t^2}{n}} \\ \text{Erro Percentual Médio (MPE)} &= \frac{\sum_{t=1}^n p_t}{n} \\ \text{Erro Absoluto Percentual Médio (MAPE)} &= \frac{\sum_{t=1}^n |p_t|}{n}\end{aligned}$$

Duas outras distinções em relação ao cálculo das medidas de acurácia merecem destaque. Em primeiro lugar, é usual a distinção entre acurácia calculada dentro da amostra e fora da amostra. O termo dentro da amostra significa que os dados utilizados para estimar o modelo também foram utilizados para determinar a magnitude dos mesmos. Já o termo fora da amostra significa que os dados não foram utilizados durante o processo de estimação, constituindo-se em previsão de fato. Pela própria natureza do processo de estimação, os erros dentro da amostra sempre são menores do que os fora da amostra.

Em segundo lugar, os horizontes de previsão para os quais as medidas de acurácia são calculadas podem variar, especialmente na previsão fora da amostra. Neste trabalho os erros de previsão foram calculados para o horizonte $t + 12$.

² HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, v. 22, p. 679–688, 2006.

2.4 Métodos Ingênuos de Previsão

As medidas apresentadas na seção 2.3 não são capazes de comparar qual o ganho relativo de acurácia obtido pela utilização de uma previsão em detrimento de outra. Makridakis, Wheelwright e Hyndman (1998) sugerem a utilização de um método de previsão extremamente simples que possa ser utilizado como um benchmarking para métodos e modelos mais completos. Este tipo de método é usualmente chamado de “ingênuo”. (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).

A racionalidade por trás desse tipo de comparação é medir quão melhor a previsão de um modelo complexo é em relação a previsão mais simples que poderia ser obtida pelos atores interessados. Ou seja, é possível responder a pergunta se os recursos investidos na elaboração da previsão mais complexa geraram os resultados desejados.

São dois os métodos ingênuos mais utilizados, usualmente chamados de Naive Forecast 1 (NF1) e Naive Forecast (NF2). O NF1 usa a observação mais recente como previsão, ou seja, assume que os dados seguem um passeio aleatório conforme exposto na seção 1.3. Já o NF2 admite a possibilidade de sazonalidade nos dados. Nesse caso, a previsão é a última observação do mesmo período do ano anterior.

Neste trabalho, tendo em vista a tendência de crescimento do ICMS observada durante a análise exploratória, será utilizada como previsão do método ingênuo um passeio aleatório com deslocamento. O termo deslocamento é utilizado pois o parâmetro de deslocamento δ permite o que a previsão para Y_t se desloque para cima ou para baixo ao longo do tempo. O passeio aleatório com deslocamento pode ser expresso por

$$Y_t = \delta + Y_{t-1} + a_t \quad (2.6)$$

A previsão para o período h é igual ao último valor disponível da série acrescido de h vezes o termo de deslocamento δ .

2.5 Critérios de Seleção de Modelos

Além das medidas de acurácia descritas na seção 2.3, diversos critérios de informação foram sugeridos para auxiliar no processo de escolha de modelos concorrentes. O problema básico que estes critérios de informação visam solucionar é o trade-off existente entre a qualidade de ajuste de um modelo a determinado conjunto de dados e sua complexidade associada, usualmente medida pelo número de parâmetros. Os dois critérios mais utilizados para este fim são o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC).

O AIC e o BIC podem ser expressos por

$$AIC = \log \hat{\sigma}_\epsilon^2 + \frac{2(p+1)}{T}$$
$$BIC = \log \hat{\sigma}_\epsilon^2 + \frac{(\log T)p}{T}$$

onde p representa o número de parâmetros estimados no modelo e T o tamanho da amostra. Para ambos os critérios quanto menor o valor obtido, melhor o modelo.

2.6 Combinação de Previsões

A combinação de previsões refere-se a agregação de estimativas concorrentes sobre o mesmo evento, com o objetivo de melhorar a acurácia das mesmas. Em uma extensa revisão de literatura, [Clemen \(1989\)](#) argumenta que:

Consider what we have learned about the combination of forecasts over the past twenty years. Models have been developed to find ‘optimal’ combinations of forecasts. Both simulation and empirical studies have been done to test the models. Bayesian interpretations have been presented. The results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy. This has been the result whether the forecasts are judgmental or statistical, econometric or extrapolation. Furthermore, in many cases one can make dramatic performance improvements by simply averaging the forecasts. ([CLEMEN, 1989](#), p. 559).

[Newbold e Harvey \(2004\)](#) apontam três fatores que explicam o interesse de acadêmicos e de profissionais na combinação de previsões: a atração intuitiva do conceito; a simplicidade de implementação da maior parte dos métodos utilizados para determinação de uma combinação específica e a vasta literatura empírica que sugere que a combinação funciona na prática. Cada um deles merece uma exposição mais detalhada.

Dois motivos principais explicam a atração intuitiva do conceito. Em primeiro lugar, é razoável afirmar que em face da impossibilidade de determinação do modelo verdadeiro, descartar as informações de um modelo válido parece contraproducente para o objetivo de acurácia. Em segundo lugar, a combinação de previsão é análoga ao princípio da diversificação na teoria do portfólio, onde o objetivo principal é a diversificação do risco. Neste, a diversificação ocorre entre ativos diferentes, já na combinação, entre previsões diferentes. ([NEWBOLD; HARVEY, 2004](#)).

Segundo [Clemen \(1989\)](#), o artigo de [Bates e Granger \(1969\)](#)³ é considerado a referência inicial sobre combinação de previsões, onde os autores desenvolveram e testaram uma série de técnicas para combinar previsões pontuais. [Newbold e Harvey \(2004\)](#) argumentam que as técnicas propostas por [Bates e Granger](#) requerem cálculos simples para determinação dos pesos das previsões individuais, contudo, diversos estudos concluirão que esses métodos simples geram previsões extremamente acuradas.

Os procedimentos para combinação variam em complexidade, no entanto, todos eles envolvem uma média ponderada de previsões, de forma que a complexidade é determinada pela forma de determinação dos pesos.

Seja F_{1t} e F_{2t} duas previsões para a mesma variável de interesse Y_t , uma combinação das previsões é denotada por

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

³ [BATES, J. M.; GRANGER, C. W. J. The Combination of Forecasts. *Operational Research Quarterly*, v. 20, p. 451–468, 1969.](#)

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

$$\epsilon_{ct} = Y_t - F_{ct} = (1 - \lambda)\epsilon_{1t} + \lambda\epsilon_{2t}$$

também é não-viesado e possui variância

$$Var(\epsilon_{ct}) = E(\epsilon_{ct}^2) = (1 - \lambda)^2\sigma_1^2 + \lambda^2\sigma_2^2 + 2\lambda(1 - \lambda)\rho\sigma_1\sigma_2 \quad (2.7)$$

Para minimizar a equação 2.7, temos de derivar em relação a λ e igualar a zero, obtendo o valor ótimo de λ igual a

$$\lambda = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (2.8)$$

2.6.1 Procedimentos de Combinação

Nesta seção serão detalhados os principais procedimentos para geração dos pesos da combinação.

Como na prática, os valores de σ_1 , σ_2 e ρ não são conhecidos, a equação 2.8 possui pouco valor operacional. No entanto, substituindo os parâmetros desconhecidos pelos seus estimadores amostrais usuais obtemos

$$\hat{\lambda} = \frac{\sum_{t=1}^T e_{1t}^2 - \sum_{t=1}^T e_{1t}e_{2t}}{\sum_{t=1}^T e_{1t}^2 + \sum_{t=1}^T e_{2t}^2 - 2\sum_{t=1}^T e_{1t}e_{2t}} \quad (2.9)$$

que representa o primeiro procedimento de combinação (Combinação Ótima - C1). A equação 2.9 pode ser generalizada para o caso de N previsões

$$\hat{\lambda} = \frac{\hat{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\hat{\Sigma}^{-1}\mathbf{1}} \quad (2.10)$$

onde $\hat{\Sigma}$ é a matrix de covariância amostral dos erros e $\mathbf{1}$ é o vetor coluna $(1, 1, \dots, 1)$.

No entanto, conforme apontado por [Newbold e Harvey \(2004\)](#), a equação 2.9 pode ser interpretada como o estimador de mínimos quadrados de λ com restrições nos coeficientes de que $\beta_0 = 0$ e $\beta_1 + \beta_2 = 1$. Portanto, um segundo procedimento de combinação (Regressão Linear - C2), sugerido por [Granger e Ramanathan \(1984\)](#), é generalizar 2.9 para

$$Y_t = \beta_0 + \beta_1 F_{1t} + \beta_2 F_{2t} + \epsilon_t \quad (2.11)$$

Um terceiro procedimento (Performance - C3) de combinação é ignorar a covariância amostral em 2.9 e atribuir pesos inversamente proporcionais ao erro quadrado médio, procedimento sugerido por [Bates e Granger \(1969\)](#) e expresso por

$$\hat{\lambda} = \frac{\sum_{t=1}^T e_{1t}^2}{\sum_{t=1}^T e_{1t}^2 + \sum_{t=1}^T e_{2t}^2} \quad (2.12)$$

A equação 2.12 pode ser generalizada por o caso de N previsões

$$\hat{\lambda}_i = \frac{(\sum_{t=1}^T e_{it}^2)^{-1}}{\sum_{n=1}^N (\sum_{t=1}^T e_{it}^2)^{-1}} \quad (2.13)$$

onde $\hat{\lambda}_i$ é o peso a ser atribuído a i -ésima previsão.

Além dos três procedimentos expostos acima, também serão utilizados procedimentos simples de combinação, que empiricamente possuem um desempenho satisfatório (CLEMEN, 1989). Serão utilizados como forma de determinação dos pesos a média simples (C4) e a mediana (C5).

3 Metodologia

Este capítulo visa apresentar as fontes das bases de dados e os softwares utilizados para elaboração desta monografia.

3.1 Base de Dados

As informações de arrecadação do ICMS utilizadas neste trabalho foram extraídas do site da Secretaria de Estado de Fazenda (SEF-MG). Foi utilizado somente a arrecadação de ICMS principal, ou seja, excluindo a parcela proveniente de multas, juros de mora e dívida ativa.

A base de dados original consiste em 132 observações mensais, no período de jan/2003 a dez/2013, do valor arrecadado de ICMS conforme registro dos sistemas corporativos do Estado de Minas Gerais. O período de jan/2003 a dez/2012 foi utilizado para ajuste dos modelos e determinação dos pesos de combinação, enquanto o período de jan/2013 a dez/2013 foi reservado para validação da acurácia das previsões.

Esses valores podem ser obtidos no endereço http://www.fazenda.mg.gov.br/governo/receita_estado/evolucaoreceita/.

Cabe ressaltar que existem diferenças entre esses valores e os publicados no relatórios contábeis do Estado de Minas Gerais. Conforme exposto no site da SEF-MG, essa diferença decorre principalmente de: valor de bens (não apenas dinheiro) recebidos pelo Estado como forma de pagamento de tributos; valores restituídos a contribuintes por terem sido recolhidos a maior ou indevidamente e diferenças oriundas de meses anteriores que não se referem ao mês em curso.

3.2 Software

Para elaboração deste trabalho foi utilizado o software R, na sua versão 3.0.2 (2013-09-25) e o pacote “forecast”.

O R é ao mesmo tempo uma linguagem de programação e um ambiente para estatística computacional. Ele possui uma série de técnicas estatísticas implementadas, bem como funcionalidade que suportam todas as fases do processo de análise de dados. Além disso, suas funcionalidade básicas são facilmente extendidas por meio da instalação de pacotes adicionais que cumprem propósitos específicos.

No endereço <http://www.r-project.org/> podem ser encontradas informações adicionais, bem como o instalador do R, que está licenciado em uma versão de software

livre, e possui versões para Windows, UNIX e MacOS.

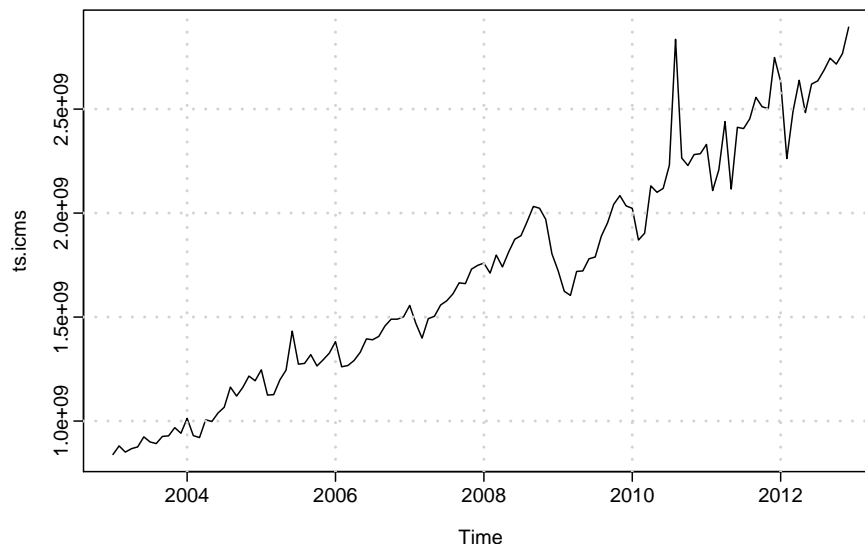
4 Resultados

Este capítulo está dividido em três seções. Na primeira seção será realizada uma análise exploratória da arrecadação do ICMS para aprofundar o conhecimento sobre suas características. Na segunda seção serão identificados e estimados os modelos originais que serão utilizadas para combinação de previsão. Por fim, na última seção serão apresentados os resultados da performance das previsões concorrentes.

4.1 Análise Exploratória

A figura 3 apresenta a arrecadação nominal do ICMS no estado de Minas Gerais no período de 2003 a 2012, compreendendo 120 observações mensais. É possível notar uma tendência de crescimento ao longo do período, corroborado pelo aumento das médias anuais de arrecadação (tabela 1). A taxa média de crescimento mensal foi de 1.05% e a anual de 12.67%. O valor mínimo de arrecadação foi 839 milhões em jan/2003 e o valor máximo 2,894 milhões em dez/2012.

Figura 3 – Evolução Nominal do ICMS no período 2003-2012



Fonte: Elaboração do Autor

Tendo sido identificada a componente tendência, a mesma será removida no restante da análise para facilitar a identificação das demais características da série, como sazonalidade e outliers. A tendência será removida ajustando-se uma regressão linear simples e utilizando os resíduos da mesma para o restante da análise.

Tabela 1 – Estatísticas Descritivas Arrecadação Nominal ICMS - Em Milhões

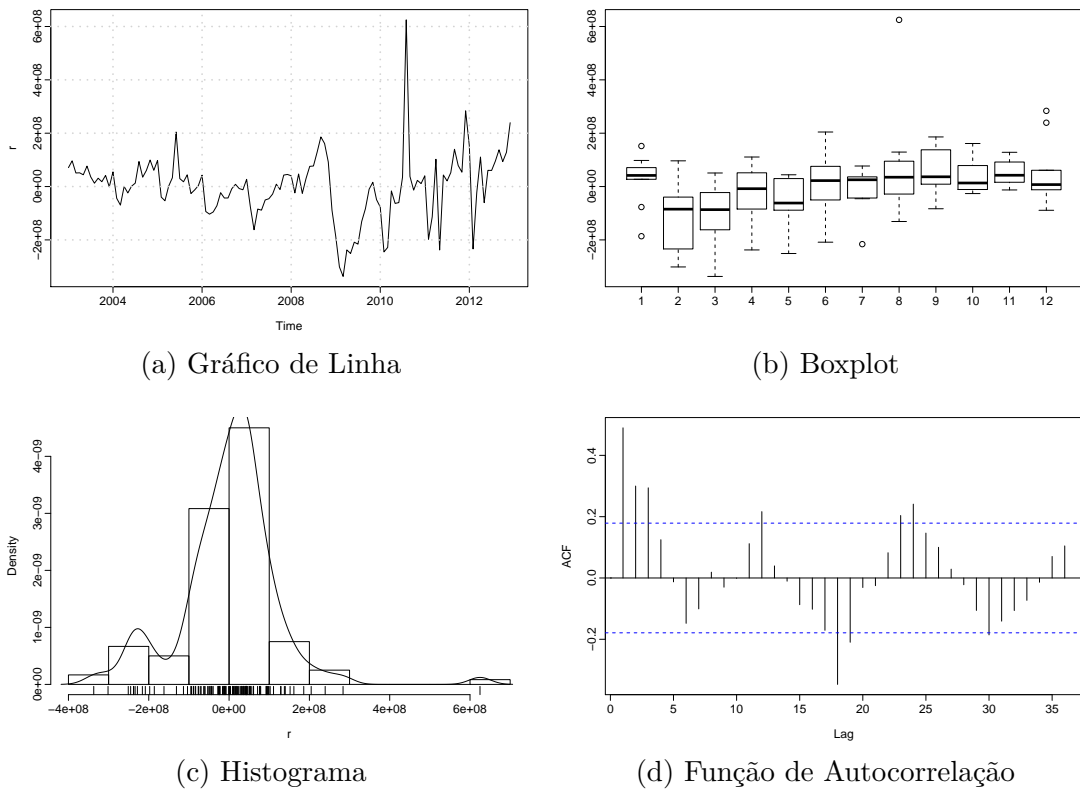
Ano	Média	Desv. Pad.	Coef. Var.
2003	899	39	0.043
2004	1,069	101	0.094
2005	1,261	85	0.067
2006	1,389	86	0.062
2007	1,581	107	0.068
2008	1,865	111	0.059
2009	1,830	166	0.091
2010	2,189	247	0.113
2011	2,399	186	0.078
2012	2,631	162	0.062

Fonte: Elaboração do Autor

Ao retirarmos a tendência, alguns pontos se destacaram na análise do gráfico 4a. Em primeiro lugar a observação de ago/2010 é maior do que a média local, indicando a possível presença de um outlier. Ao plotarmos a série padronizada (figura 5) essa suspeita é confirmada pelo fato de que essa observação apresentou um escore de 5.1. Em segundo lugar, no período entre nov/2008 e mar/2009, a arrecadação apresenta uma queda drástica. Todos estes eventos podem ser modelados por meio da análise de intervenção. A análise do gráfico 4c também permite a observação de grande assimetria nos dados e uma observação discrepante.

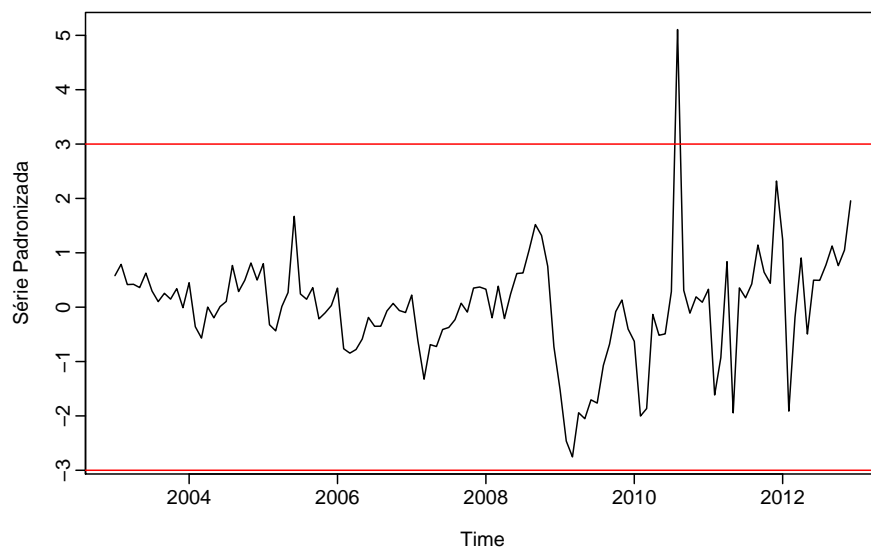
Sem a presença de tendência, a sazonalidade também se torna aparente. É possível notar claramente um padrão de sazonalidade (figuras 4b), merecendo destaque a queda observada nos meses de fevereiro, provavelmente decorrente do menor número de dias deste mês, fator que influencia diretamente a arrecadação do ICMS. A função de autocorrelação também apresenta as características de uma sazonalidade de 12 meses, com autocorrelações significativas nas defasagens 12 e 24.

Figura 4 – Arrecadação ICMS sem tendência



Fonte: Elaboração do Autor

Figura 5 – Arrecadação ICMS sem tendência - Série Padronizada



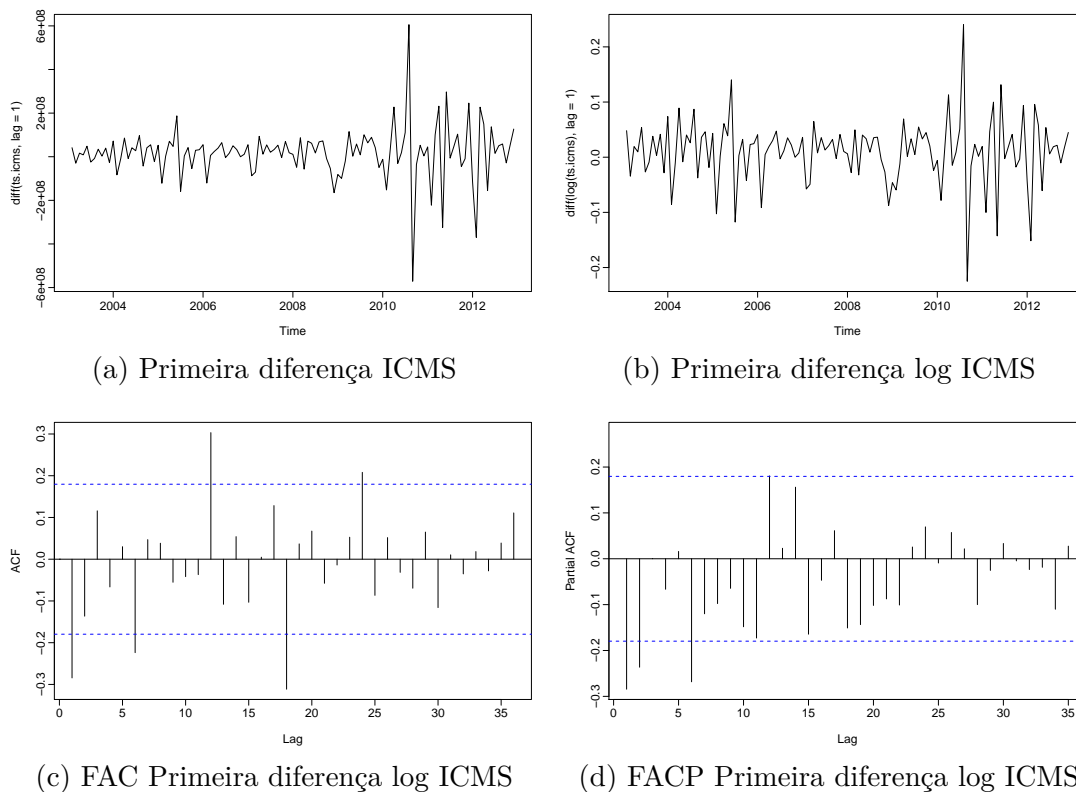
Fonte: Elaboração do Autor

4.2 Ajuste dos Modelos

A análise do gráfico 3 indica tratar-se de uma série não estacionária na média e na variância. Após a primeira diferença (figura 6a), é possível notar estacionariedade na média, mas não na variância. Como forma de estabilização da mesma, foi aplicada uma transformação logarítmica (figura 6b) nos dados originais seguida da primeira diferença, obtendo-se uma série estacionária.

A análise dos correlogramas (figuras 6c e 6d) não permitiu a identificação do comportamento teórico das autocorrelações, portanto, foram ajustados três modelos SARIMA de forma incremental, e, em todos eles, foram inseridos variáveis binárias para as observações atípicas identificados na análise exploratória, quais sejam, Nov/2008-Mar/2009 e ago/2010, bem como uma variável binária para jun/2005 tendo em vista que é de conhecimento prévio que houve um pico de arrecadação neste mês decorrente de anistias concedidas pelo Governo Mineiro. A análise de intervenção foi utilizada neste caso tendo em vista que foi capaz de melhorar a qualidade de ajuste dos modelos. Também foi ajustado um modelo de alisamento exponencial de Holt-Winters, em virtude de tratar-se de uma série com tendência e sazonalidade.

Figura 6 – Identificação dos Modelos



Fonte: Elaboração do Autor

A tabela 2 apresenta os parâmetros estimados dos modelos e a tabela 3 apresenta

as principais estatísticas de qualidade de ajuste. Todos os coeficientes foram significativos a um nível de significância de 0.05 e cumpriram as condições de estacionariedade e invertibilidade. O melhor ajuste foi obtido pelos modelos $SARIMA(2, 1, 0)(1, 0, 0)_{12}$ e $SARIMA(0, 1, 1)(1, 0, 0)_{12}$ pelos critérios AIC e BIC e pelo modelo $SARIMA(2, 1, 0)(0, 1, 1)_{12}$ pelos critérios RMSE e MAPE. Os resíduos de todos os modelos apresentaram ausência de autocorrelação e de heteroscedasticidade, no entanto, todos apresentaram desvios de normalidade (vide apêndice A), provavelmente em virtude de observações extremas presentes na amostra. Os quatro modelos também foram capazes de acompanhar o comportamento da série original (figura 7).

Tabela 2 – Estatísticas de Estimação dos Modelos (jan/2003 – dez/2012, n=120)

Modelo	Coefficientes	Estimativa	p-valor
$SARIMA(2, 1, 0)(1, 0, 0)_{12}$ – M1			
	ϕ_1	-0.400	0.000
	ϕ_2	-0.218	0.021
	Φ_1	0.504	0.000
	c	0.010	0.011
	x_1	0.236	0.000
	x_2	0.126	0.000
	x_3	-0.073	0.001
$SARIMA(0, 1, 1)(1, 0, 0)_{12}$ – M2			
	θ_1	-0.377	0.000
	Φ_1	0.507	0.000
	c	0.010	0.013
	x_1	0.244	0.000
	x_2	0.129	0.000
	x_3	-0.071	0.002
$SARIMA(2, 1, 0)(0, 1, 1)_{12}$ – M3			
	ϕ_1	-0.485	0.000
	ϕ_2	-0.334	0.000
	Θ_1	-0.707	0.013
	x_1	0.234	0.000
	x_2	0.125	0.000
	x_3	-0.077	0.002
Holt-Winters – M4			
	α	0.504	
	β	0.100	
	γ	0.000	

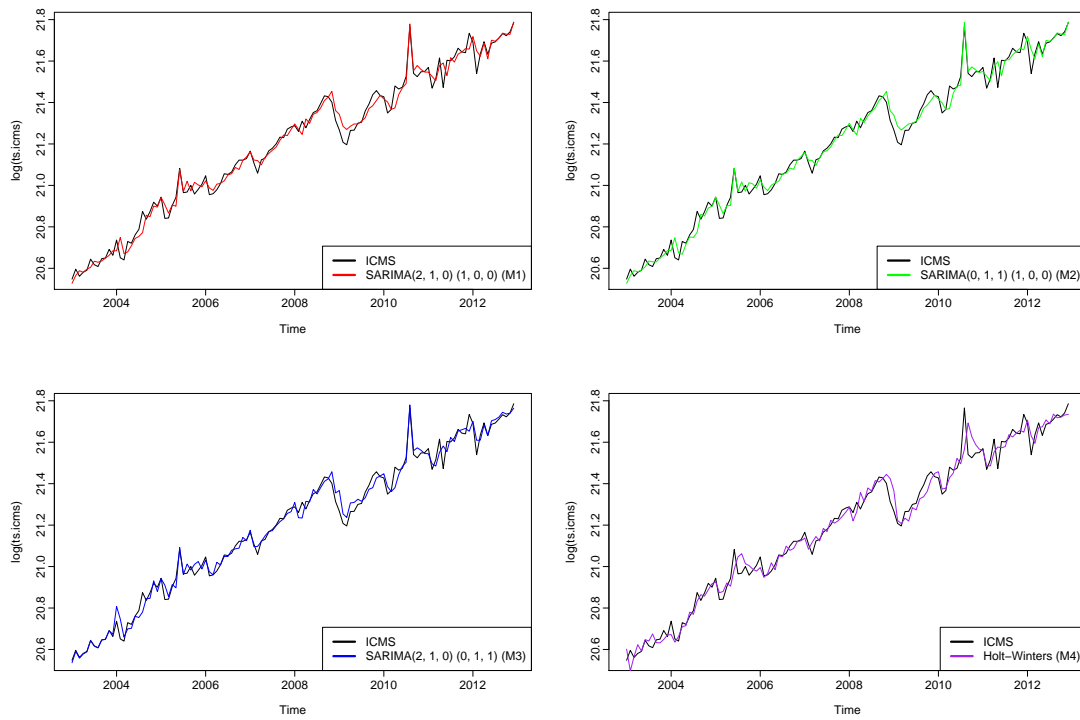
Fonte: Elaboração do Autor

Tabela 3 – Estatísticas de Ajuste dos Modelos (jan/2003 – dez/2012, n=120)

Modelo	AIC	BIC	ME	RMSE	MPE	MAPE
$SARIMA(2, 1, 0)(1, 0, 0)_{12}$ – M1	-422.16	-399.93	0.000	0.037	0.000	0.126
$SARIMA(0, 1, 1)(1, 0, 0)_{12}$ – M2	-421.57	-402.12	0.000	0.038	0.000	0.127
$SARIMA(2, 1, 0)(0, 1, 1)_{12}$ – M3	-340.43	-321.72	-0.003	0.034	-0.014	0.113
Holt-Winters – M4			0.000	0.046	0.000	0.159

Fonte: Elaboração do Autor

Figura 7 – Ajuste dos Modelos



Fonte: Elaboração do Autor

4.3 Comparação dos Modelos

Para comparação da acurácia das previsões dos modelos estimados com aquelas originárias das combinações de previsões foram separados da amostra inicial doze observações referentes ao período jan/2013 a dez/2013. Os pesos para os diferentes procedimentos de combinação apresentados na seção 2.6.1 foram calculados com base nos resíduos do período jan/2003 a dez/2012, ou seja, os pesos foram calculados dentro da amostra.

A tabela 4 apresenta as medidas de acurácia calculadas fora da amostra e os pesos estimados dentro da amostra para os quatro modelos originais, as cinco combinações e o modelo ingênuo utilizado para benchmarking. Merece destaque o fato de que a combinação ótima e a regressão linear geraram pesos extremamente próximos, justificando suas performances semelhantes.

De forma geral, as combinações foram superiores aos modelos originais. O RMSE médio das combinações foi de 106 milhões, enquanto o RMSE dos modelos originais - com a exclusão do modelo ingênuo - foi de 116 milhões, uma diferença superior a 10%. O MAPE médio das combinações foi de 2.91%, o dos modelos originais de 3.22%, diferença de 0.31 pp.

Ao compararmos as previsões pela acurácia dos doze meses, ou seja, a diferença entre a arrecadação efetiva no ano de 2013 e a previsão para o ano de 2013, permitindo o cancelamento de erros de direções opostas, as combinações mantêm um desempenho superior. Esses valores correspondem ao ME multiplicado por 12. Enquanto a média do erro dos modelos originais foi de 285 milhões, a média dos erros das combinações foi de 134 milhões. A procedimento de combinação média simples teve o melhor desempenho, com um erro de 27 milhões.

Além de medida de acurácia absoluta, como as apresentadas até agora, também podemos calcular medidas de acurácia relativa. A tabela 5 apresenta a razão entre o MAPE das previsões dos modelos originais e das combinações com o MAPE do método ingênuo. Valores maiores que 1 indicam que a previsão ingênua é superior e valores menores indicam o oposto, quantificando essa melhoria. Todas as previsões foram superiores ao método ingênuo, no entanto, das cinco melhores previsões, quatro são originárias das combinações de previsão. Merece destaque o procedimento de combinação ótimo, que leva em consideração a correlação entre os erros das previsões, que obteve um índice de 0.540, indicando que seu desempenho foi quase duas vezes superior ao método ingênuo.

Por fim, as figuras 8, 9 e 10 mostram os erros de previsão padronizados no período fora da amostra para as nove previsões em relação ao modelo ingênuo. É possível notar que o desempenho dos modelos foi relativamente semelhante em relação aos meses de maior desvio de previsão. Também é possível inferir que as combinações tiveram um desempenho melhor pois foram capazes de suavizar os meses de desvio elevado, como por exemplo o

Tabela 4 – Acurácia Fora da Amostra (jan/2013 – dez/2013, n=12) - Em Milhões

Modelo	ME	RMSE	MPE	MAPE	Pesos				
					Const	M1	M2	M3	M4
Modelos Originais									
Método Ingênuo	-69,4	189,1	-2.83%	5.23%	-	-	-	-	-
<i>SARIMA</i> (2, 1, 0)(1, 0, 0) ₁₂	11,1	125,3	0.10%	3.35%	-	1	-	-	-
<i>SARIMA</i> (0, 1, 1)(1, 0, 0) ₁₂	10,6	125,0	0.08%	3.34%	-	-	1	-	-
<i>SARIMA</i> (2, 1, 0)(0, 1, 1) ₁₂	30,3	103,4	0.87%	2.93%	-	-	-	1	-
Holt-Winters	-43,1	111,6	-1.61%	3.27%	-	-	-	-	1
Combinações									
Combinação Ótima – C1	14,6	100,1	0.33%	2.82%	-	0.436	-0.369	0.735	0.199
Regressão Linear – C2	20,3	101,2	0.52%	2.85%	0.039	0.438	-0.369	0.731	0.197
Performance – C3	7,9	109,2	0.05%	3.00%	-	0.263	0.257	0.313	0.168
Média Simples – C4	2,2	107,9	-0.14%	2.96%	-	0.250	0.250	0.250	0.250
Mediana – C5	10,9	109,4	0.14%	2.94%	-	-	-	-	-

Fonte: Elaboração do Autor

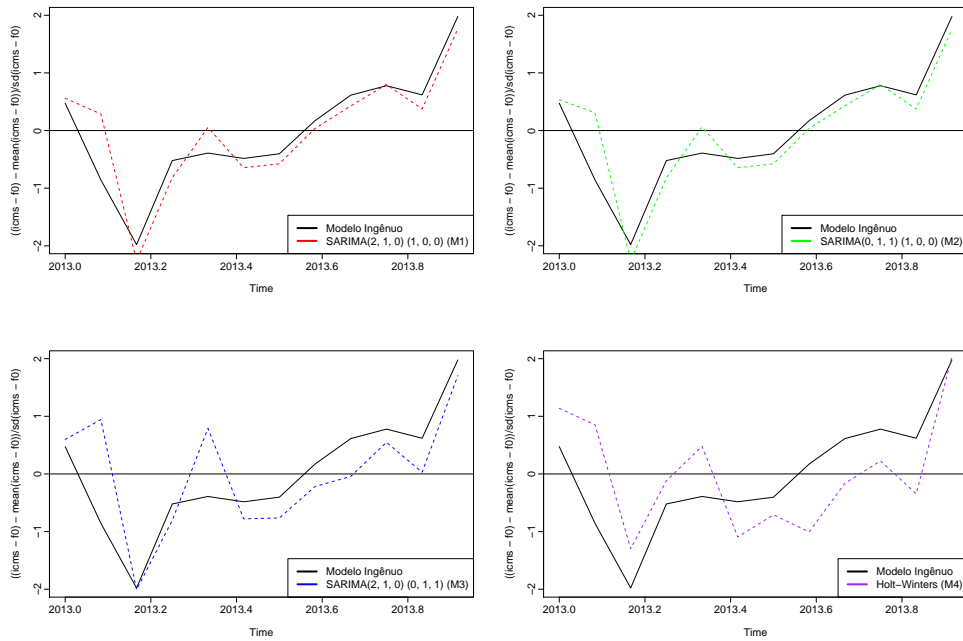
Tabela 5 – Razão entre MAPE e MAPE Método Ingênuo (jan/2013 – dez/2013, n=12)

Modelos Originais e Combinações	Razão MAPE
Combinação Ótima – C1	0.540
Regressão Linear – C2	0.545
<i>SARIMA</i> (2, 1, 0) (0, 1, 1) – M3	0.561
Mediana – C5	0.562
Média Simples – C4	0.565
Performance – C3	0.574
<i>SARIMA</i> (0, 1, 1) (1, 0, 0) – M2	0.625
<i>SARIMA</i> (2, 1, 0) (1, 0, 0) – M1	0.639
Holt-Winters – M4	0.641
Método Ingênuo – M0	1.000

Fonte: Elaboração do Autor

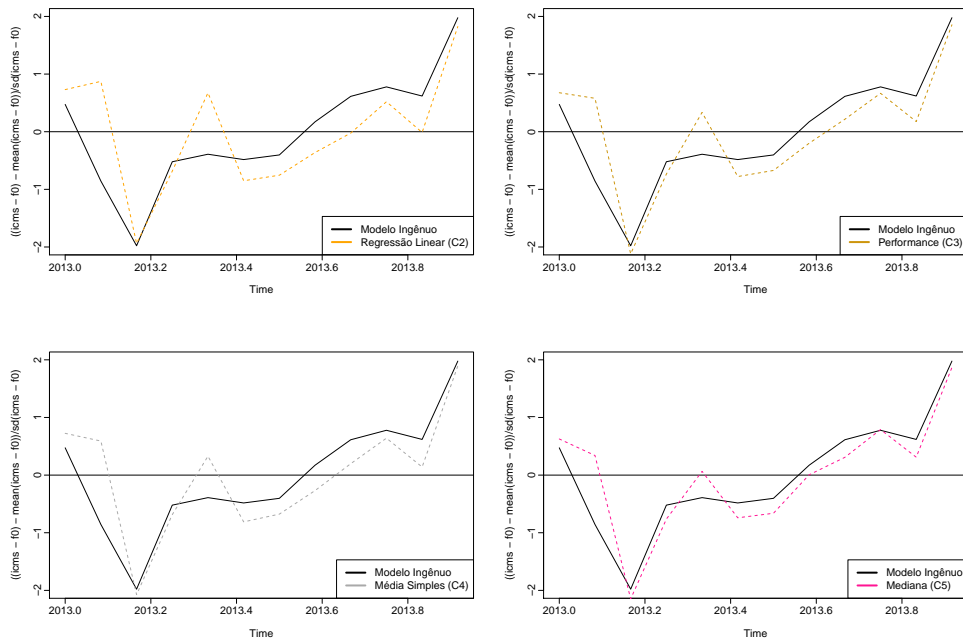
observado em março de 2013.

Figura 8 – Erros de Previsão Padronizados - Fora da Amostra - Modelos Originais



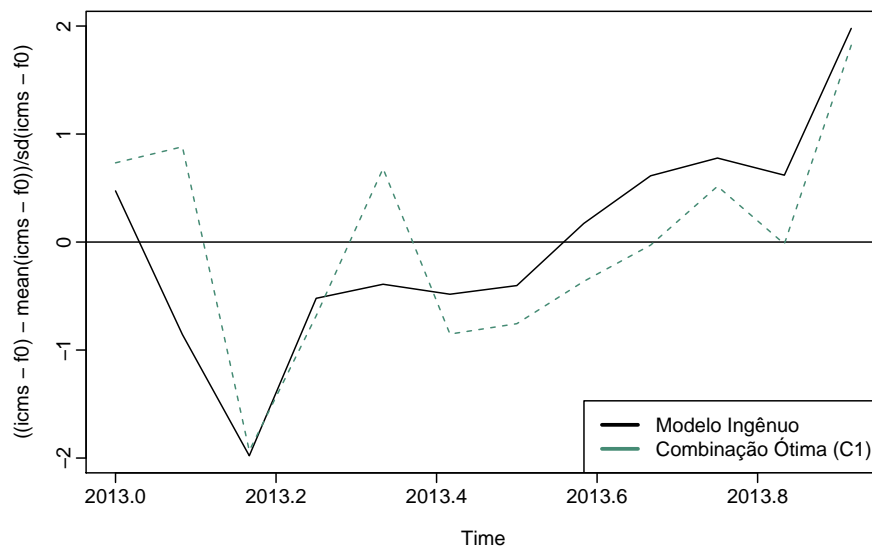
Fonte: Elaboração do Autor

Figura 9 – Erros de Previsão Padronizados - Fora da Amostra - Combinações



Fonte: Elaboração do Autor

Figura 10 – Erros de Previsão Padronizados - Fora da Amostra - Melhor Combinação



Fonte: Elaboração do Autor

Conclusão

A combinação de previsão é uma alternativa para melhorar a acurácia das previsões quando a identificação do “melhor” modelo não é possível ou extremamente complexa. O objetivo deste trabalho foi verificar empiricamente se a combinação de previsões é capaz de melhorar a acurácia das previsões para a arrecadação do ICMS em Minas Gerais.

Foi verificado que a combinação de previsão é capaz de melhorar a acurácia. Essa melhora foi observada mesmo para a combinação unicamente de modelos univariados de séries temporais bem como pela utilização de procedimentos simples de combinação, como a média aritmética. O RMSE médio foi reduzido de 116 milhões para os modelos originais para 106 milhões com as combinações. Melhora ainda superior foi obtida ao analisarmos o erro médio anual. Ele foi reduzido de 285 milhões para 134 milhões.

Neste estudo de caso, os procedimentos de combinação mais complexos tiveram um desempenho superior, especialmente a combinação ótima sugerida no artigo seminal de [Bates e Granger \(1969\)](#), que leva em consideração as covariâncias entre os erros das previsões para estimar os pesos. No entanto, cabe notar que se a medida de acurácia utilizada for o Erro Médio (ME), a média aritmética foi superior.

Sugere-se para trabalhos futuros a combinação de métodos que possuem natureza distinta, como modelos causais, multivariados e qualitativos, bem como a avaliação da estabilidade dos pesos quando eles são recalculados fora da amostra e para diferentes horizontes temporais.

Referências

- BATES, J. M.; GRANGER, C. W. J. The Combination of Forecasts. *Operational Research Quarterly*, v. 20, p. 451–468, 1969.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. 4. ed. New York: John Wiley & Sons, 2008. 784 p.
- BOX, G. E. P.; TIAO, G. Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*, v. 70, p. 70–79, 1975.
- BRASIL. Lei nº 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e da outras providências. *Diário Oficial da União*, Brasília, DF, 5 mai. 2000. Disponível em: <www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm>. Acesso em: 19 fev. 2014.
- CLEMEN, R. T. Combining forecasts : A review and annotated bibliography. *International Journal of Forecasting*, v. 5, p. 559–583, 1989.
- COSTA, E. A. de A. *Fatores Institucionais que Influenciam a Previsão das Receitas Orçamentárias: Um Estudo de Caso dos Governos Estaduais Brasileiros*. Dissertação (Mestrado) — Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2011. Disponível em: <repositorio.unb.br/bitstream/10482/9937/1/2011_EduardoAugustoAbreuCosta.pdf>.
- GRANGER, C.; NEWBOLD, P. *Forecasting Economic Time Series*. 2. ed. London: Academic Press, 1986. 338 p.
- GRANGER, C.; RAMANATHAN, R. Improved Methods of Combining Forecasts. *Journal of Forecasting*, v. 3, p. 197–204, 1984.
- GUJARATI, D. N. *Basic Econometrics*. 4. ed. New York: McGraw Hill, 2004. 1002 p.
- HYNDMAN, R. J. et al. *forecast: Forecasting functions for time series and linear models*. [S.l.], 2014. R package version 5.1. Disponível em: <CRAN.R-project.org/package=forecast>.
- HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, v. 22, p. 679–688, 2006.
- HYNDMAN, R. J. et al. *Forecasting with Exponential Smoothing: The State Space Approach*. 1. ed. New York: Springer, 2008. 362 p.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. J. *Forecasting: Methods and Applications*. 3. ed. New York: John Wiley & Sons, 1998. 656 p.
- MCQUARRIE, A. D. R.; TSAI, C.-L. *Regression and Time Series Model Selection*. 1. ed. New York: World Scientific, 1998. 455 p.

MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. 2. ed. São Paulo: Edgard Blücher, 2006. 784 p.

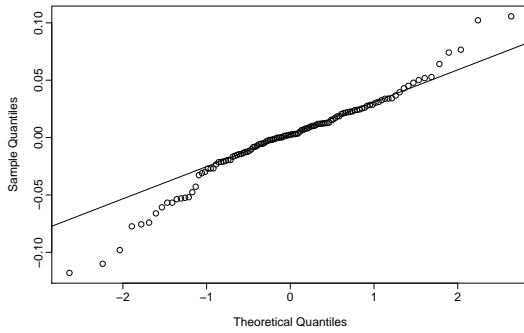
NEWBOLD, P.; HARVEY, D. Forecast Combination and Encompassing. In: CLEMENTS, M. P.; HENDRY, D. (Ed.). *A Companion to Economic Forecasting*. 1. ed. Oxford: Blackwell Publishing, 2004. cap. 12, p. 620.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: <www.R-project.org/>.

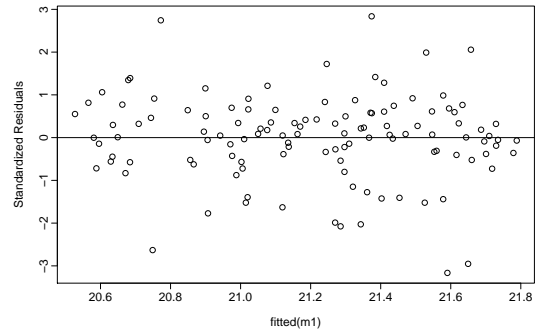
Apêndices

APÊNDICE A – Gráficos de Diagnóstico

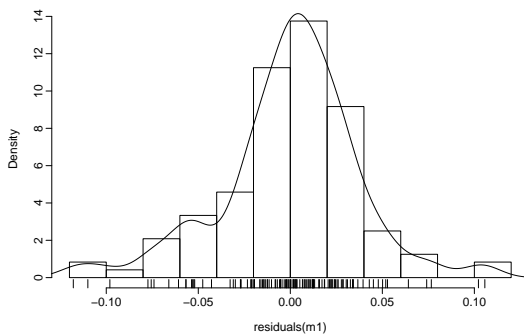
SARIMA(2, 1, 0) (1, 0, 0) – M1



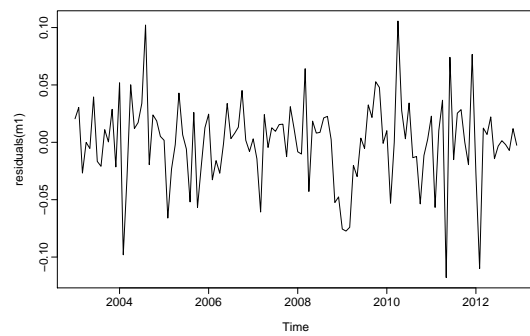
(a) QQ-Norm Resíduos



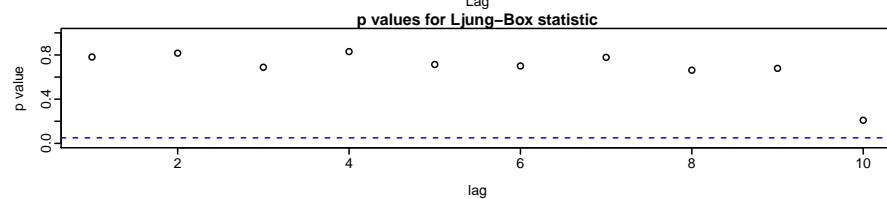
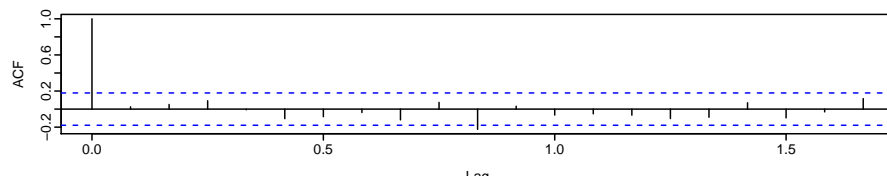
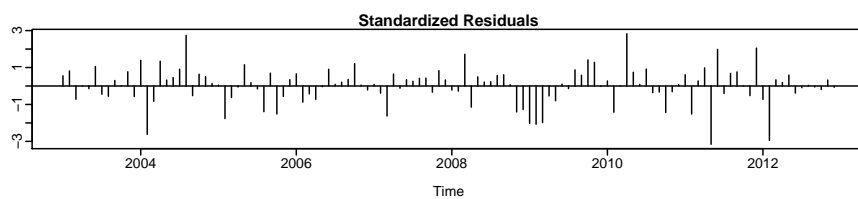
(b) Resíduos versus Estimados



(c) Histograma Resíduos



(d) Resíduos versus Tempo



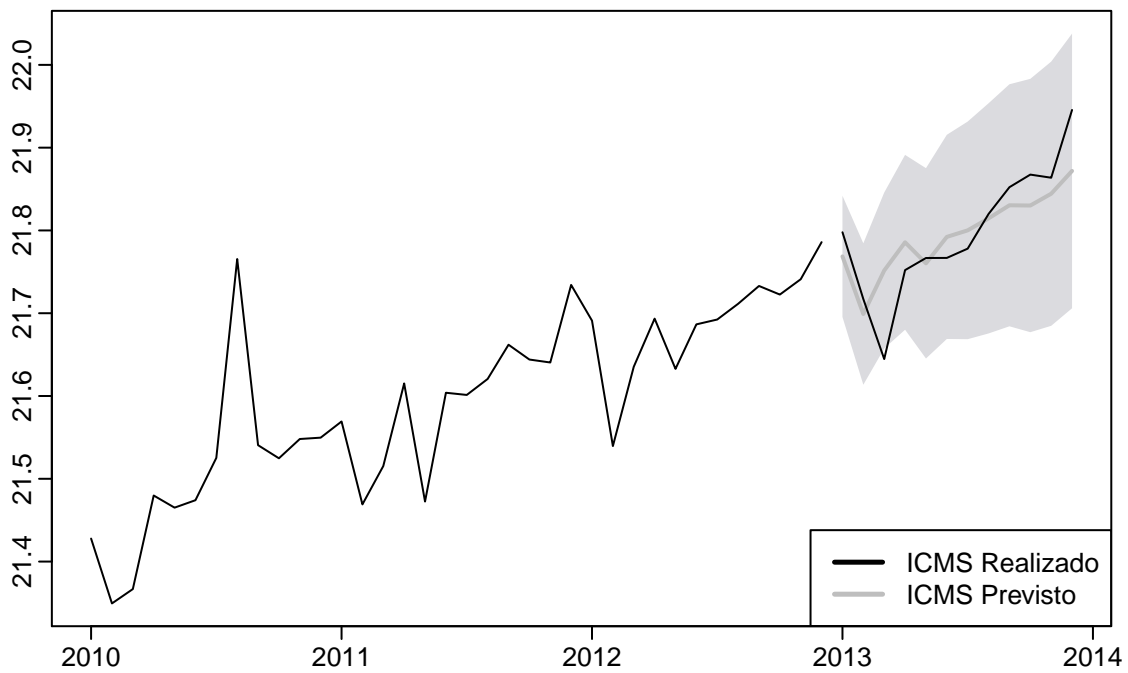


Figura 12 – Intervalo de Previsão de 95% - (Fora da Amostra)

Equação do Modelo SARIMA(2, 1, 0) (1, 0, 0) – M1

$$(1 + 0.400B + 0.218B^2)(1 - 0.504B^{12})(1 - B)Y_t = 0.010 + 0.236x_1 + 0.126x_2 - 0.073x_3 + a_t$$

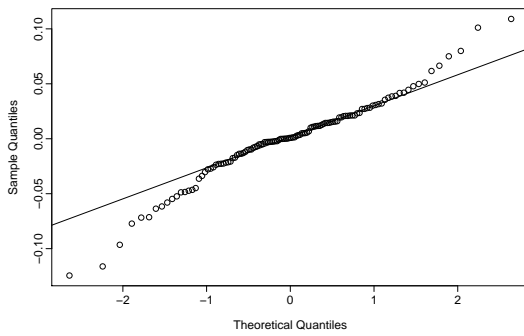
Teste de Normalidade dos Resíduos

Shapiro-Wilk normality test

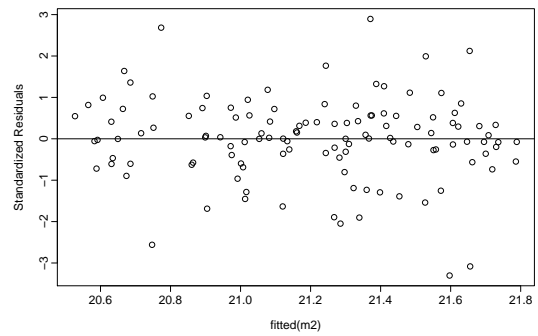
data: residuals(m1)

W = 0.9632, p-value = 0.002286

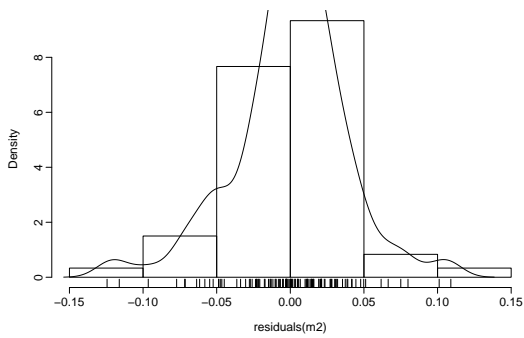
SARIMA(0, 1, 1) (1, 0, 0) – M2



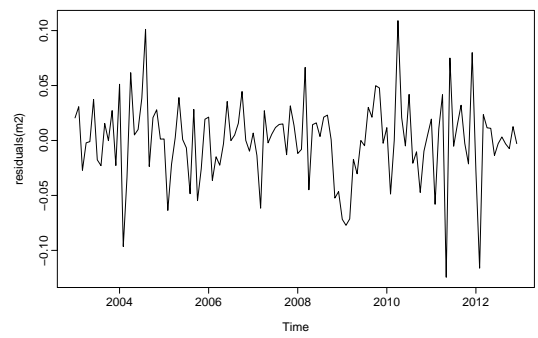
(a) QQ-Norm Resíduos



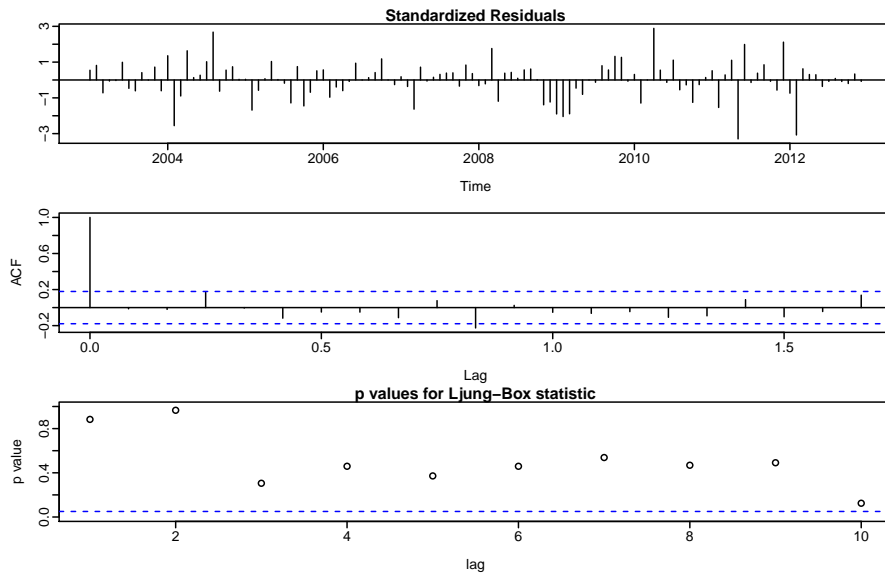
(b) Resíduos versus Estimados



(c) Histograma Resíduos



(d) Resíduos versus Tempo



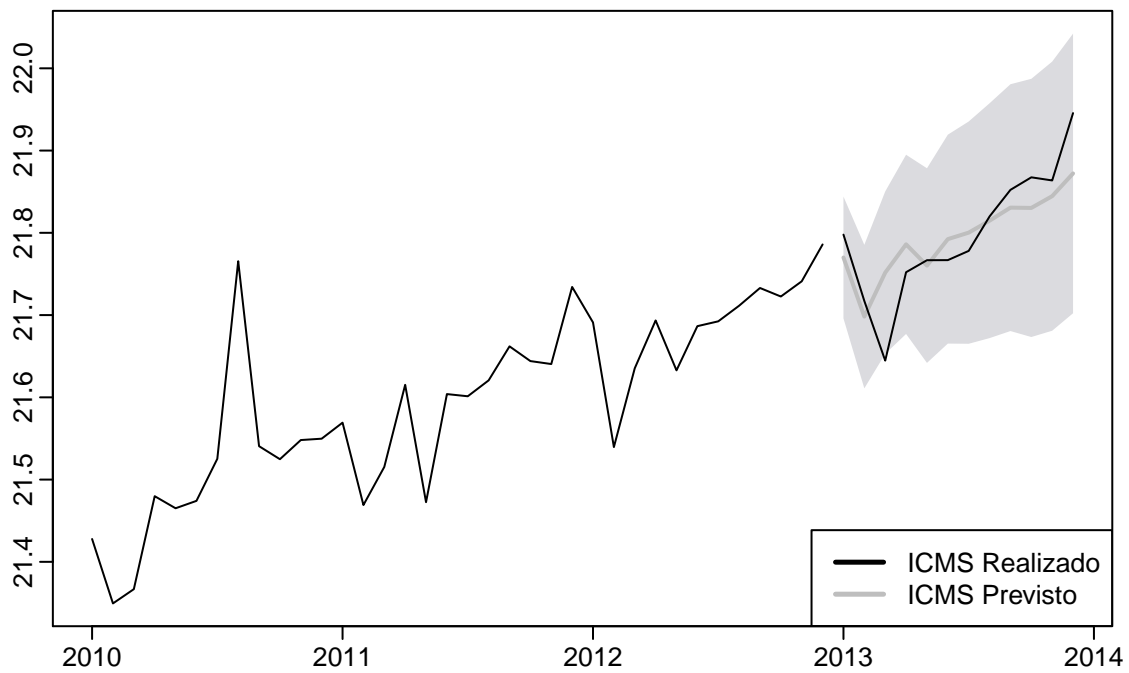


Figura 14 – Intervalo de Previsão de 95% - (Fora da Amostra)

Equação do Modelo SARIMA(0, 1, 1) (1, 0, 0) – M2

$$(1 - 0.507B^{12})(1 - B)Y_t = 0.010 + 0.244x_1 + 0.129x_2 - 0.071x_3 + (1 + 0.377B)a_t$$

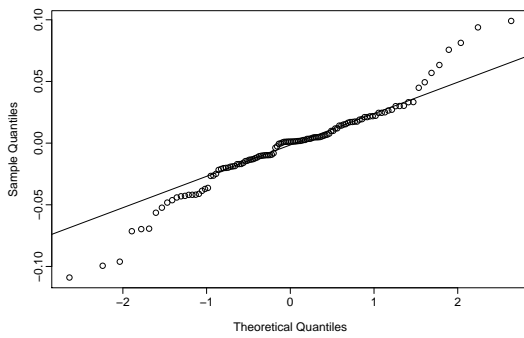
Teste de Normalidade dos Resíduos

Shapiro-Wilk normality test

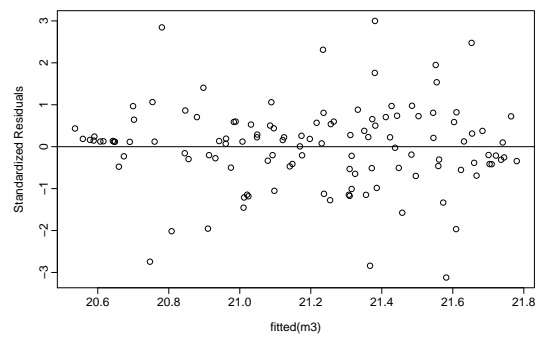
data: residuals(m2)

W = 0.9645, p-value = 0.002984

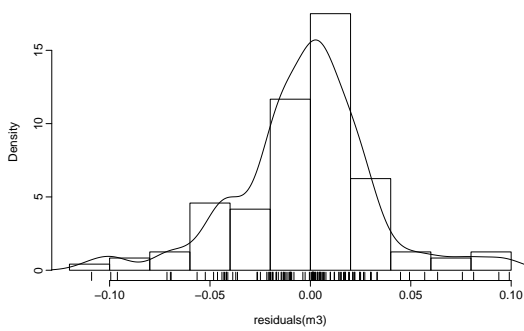
SARIMA(2, 1, 0) (0, 1, 1) – M3



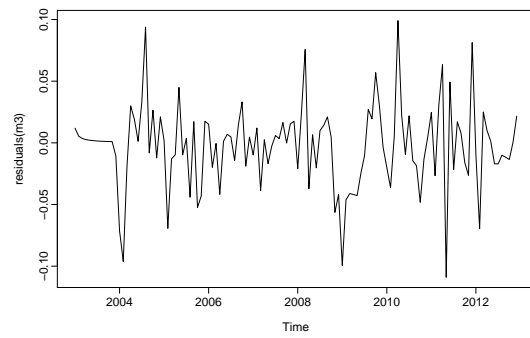
(a) QQ-Norm Resíduos



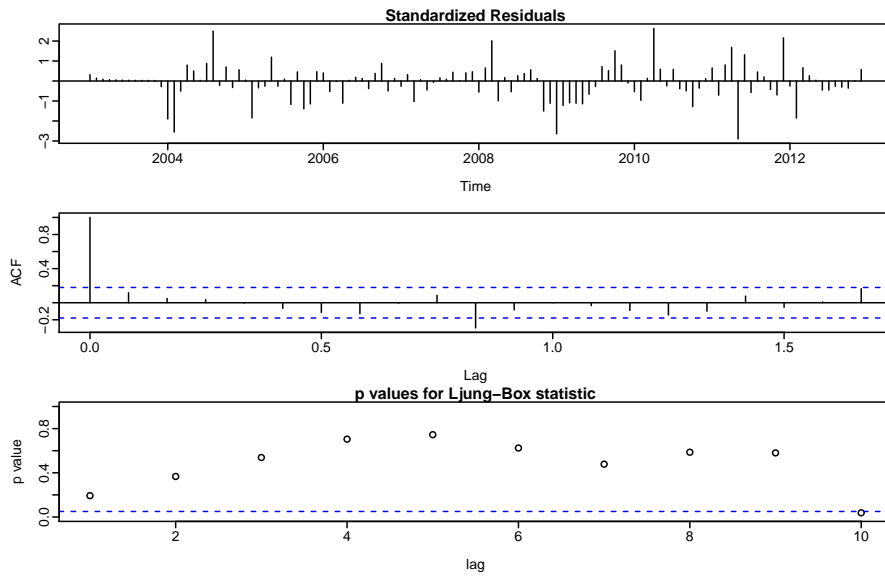
(b) Resíduos versus Estimados



(c) Histograma Resíduos



(d) Resíduos versus Tempo



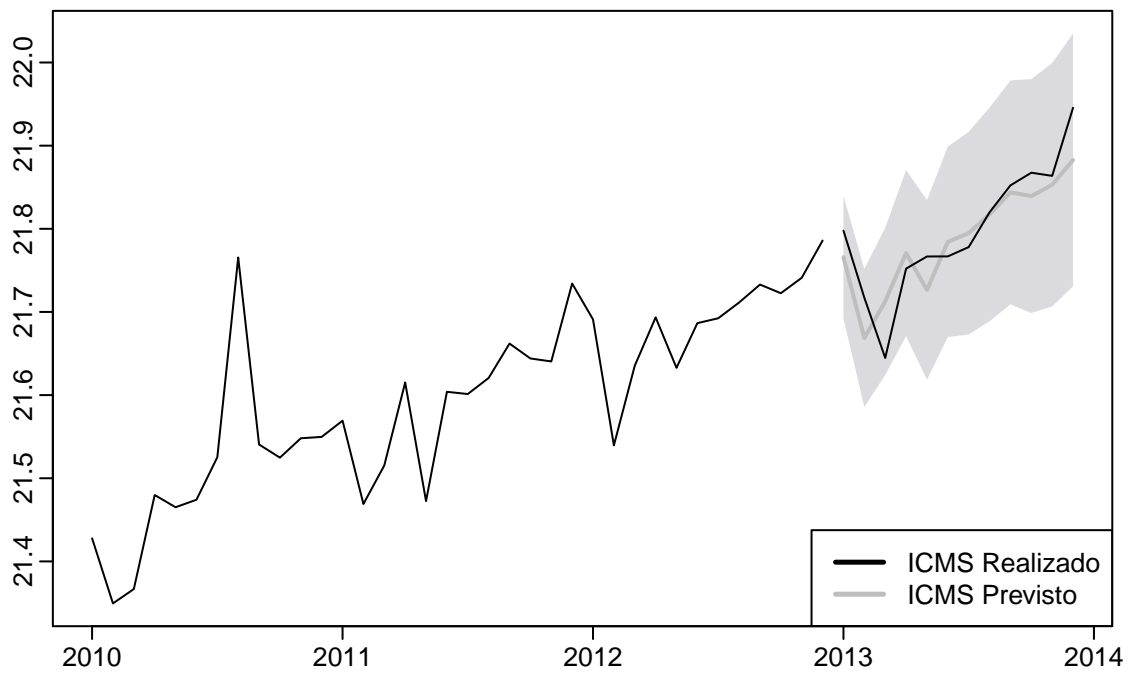


Figura 16 – Intervalo de Previsão de 95% - (Fora da Amostra)

Equação do Modelo SARIMA(2, 1, 0) (0, 1, 1) – M3

$$(1 + 0.485B + 0.334B^2)(1 - B)(1 - B^{12})Y_t = 0.234x_1 + 0.125x_2 - 0.077x_3 + (1 + 0.707B^{12})a_t$$

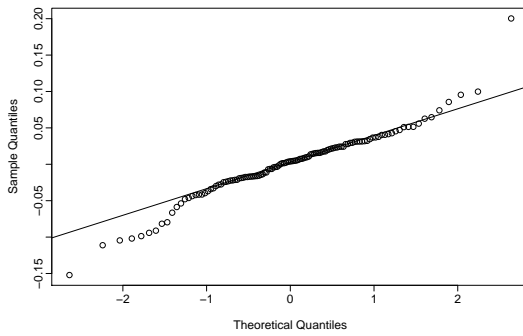
Teste de Normalidade dos Resíduos

Shapiro-Wilk normality test

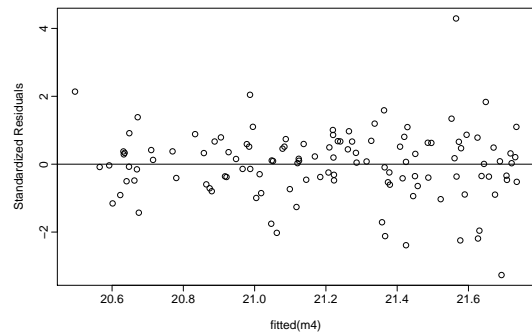
data: residuals(m3)

W = 0.9566, p-value = 0.0006803

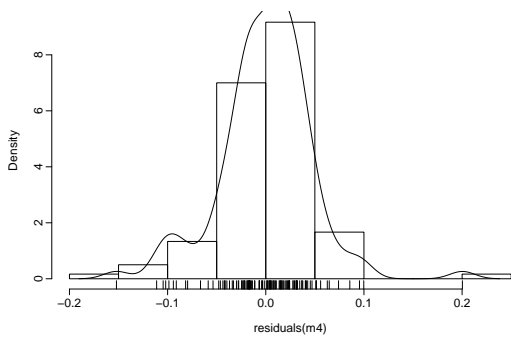
Holt-Winters – M4



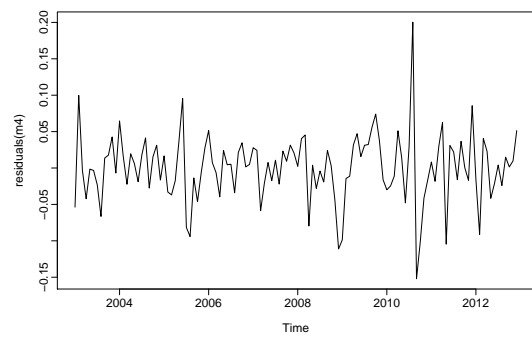
(a) QQ-Norm Resíduos



(b) Resíduos versus Estimados



(c) Histograma Resíduos



(d) Resíduos versus Tempo

APÊNDICE B – Script R

```
#####
# Monografia - Especialização em Estatística #
# Autor: Francisco Alves #
# Script para Análise dos Dados #
#####

#=====
# Leitura dos dados

rm(list=ls())

require(forecast)

dt <- read.table(file="base-icms.csv",
                 header=TRUE,
                 sep=";",
                 dec=".")

ts.icms <- ts(data=dt$icms[61:180],
             start=c(2003,1),
             end=c(2012,12),
             frequency=12)
# amostra de estimação

icms <- ts(data=dt$icms[181:192],
          start=c(2013,1),
          end=c(2013,12),
          frequency=12)
# amostra de validação

#=====
# Análise de Intervenção

it1 <- ifelse (time(ts.icms)>2010+((1/12)*6)
```

```
&
  time(ts.icms)<2010+((1/12)*8)
  , 1, 0)
# anistia - Ago/2010

it2 <- ifelse (time(ts.icms)==2005+((1/12)*5)
  , 1, 0)
# Anistia - Jun/2005

it3 <- ifelse (time(ts.icms)>=2008+((1/12)*11)
  &
  time(ts.icms)<=2009+((1/12)*2)
  , 1, 0)
# crise financeira - Nov/2008 ate Mar/2009

reg <- cbind(it1,it2,it3)
# variaveis binarias para estimacao dos modelos dentro da amostra (2003-2012)

it01 <- rep(0,12)
it02 <- rep(0,12)
it03 <- rep(0,12)

x <- cbind(it01,it02,it03)
# variaveis binarias para previsao dos modelos fora da amostra (2013)

#=====
# Estimacao dos Modelos

m0 <- rwf(ts.icms, h=12, drift=TRUE) # modelo ingenuo
f0 <- rwf(ts.icms, h=12, drift=TRUE)$mean

m1 <- Arima(x=log(ts.icms),
  order=c(2,1,0),
  seasonal=list(order=c(1,0,0), period=12),
  include.constant=TRUE,
  xreg=reg)
summary(m1)
y1 <- fitted(m1)
```

```
e1 <- residuals(m1)
f1 <- forecast.Arima(m1, level=0.95, lambda=0, xreg=x)$mean

m2 <- Arima(x=log(ts.icms),
            order=c(0,1,1),
            seasonal=list(order=c(1,0,0), period=12),
            include.constant=TRUE,
            xreg=reg)
summary(m2)
y2 <- fitted(m2)
e2 <- residuals(m2)
f2 <- forecast.Arima(m2, level=0.95, lambda=0, xreg=x)$mean

m3 <- Arima(x=log(ts.icms),
            order=c(2,1,0),
            seasonal=list(order=c(0,1,1), period=12),
            include.constant=TRUE,
            xreg=reg)
summary(m3)
y3 <- fitted(m3)
e3 <- residuals(m3)
f3 <- forecast.Arima(m3, level=0.95, lambda=0, xreg=x)$mean

m4 <- hw(log(ts.icms), h=12, seasonal="additive", damped=FALSE)
summary(m4)
y4 <- fitted(m4)
e4 <- residuals(m4)
f4 <- exp(forecast(m4)$mean)

y = cbind(y1, y2, y3, y4)
# data frame dos valores estimados - dentro da amostra - escala log

e = cbind(e1, e2, e3, e4)
# data frame dos erros - dentro da amostra - escala log
```

```
f = cbind(f1, f2, f3, f4)
# data frame das previsões - fora da amostra - escala original

#=====
# Metodos de Combinacao

# Combinacao Otima (C1) - Bates, Granger

S = cov(e)
Si = solve(S)
E = matrix(data = c(1, 1, 1, 1), nrow=4, ncol=1, byrow=TRUE )
Et = t(E)

p1 <- as.vector((Si %*% E) / ((Et %*% Si) %*% E)[1, 1])
# Pesos C1

c1 <- p1[1]*f[, 1] + p1[2]*f[, 2] + p1[3]*f[, 3] + p1[4]*f[, 4]
# Previsão C1

#=====
# Regressao Linear (C2) - Granger, Ramanathan

p2 <- as.vector(coefficients(lm(log(ts.icms) ~ y1 + y2 + y3 + y4)))
# Pesos C2

c2 <- exp(p2[1]) + p2[2]*f[, 1] + p2[3]*f[, 2] + p2[4]*f[, 3] + p2[5]*f[, 4]
# Previsão C2

#=====
# Performance (C3) - Bates, Granger

s1 = sum(e1^2)
s2 = sum(e2^2)
s3 = sum(e3^2)
s4 = sum(e4^2)

ss = (1 / s1) + (1 / s2) + (1 / s3) + (1 / s4)
```

```
p3 <- c((1 / s1) / ss, (1 / s2) / ss, (1 / s3) / ss, (1 / s4) / ss)
# Pesos C3

c3 <- p3[1]*f[, 1] + p3[2]*f[, 2] + p3[3]*f[, 3] + p3[4]*f[, 4]
# Previsão C3

#=====
# Media Simples (C4) e Mediana (C5)

c4 <- apply(f, 1, mean)
# Previsao C4

c5 <- apply(f, 1, median)
# Previsao C5

#=====
# Medidas de Acuracia

acuracia <- rbind(accuracy(f0, icms),
                 accuracy(f1, icms),
                 accuracy(f2, icms),
                 accuracy(f3, icms),
                 accuracy(f4, icms),
                 accuracy(c1, icms),
                 accuracy(c2, icms),
                 accuracy(c3, icms),
                 accuracy(c4, icms),
                 accuracy(c5, icms))[, c(1,2,4,5)]
# medidas de acuracia fora da amostra - escala natural
```