



Instituto de
MATEMÁTICA
E ESTATÍSTICA

UFRGS

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

**MODELOS DE SÉRIES TEMPORAIS:
UMA APLICAÇÃO NA PROJEÇÃO DE INADIMPLÊNCIA
DE UMA INSTITUIÇÃO FINANCEIRA**

ANGÉLICA SEGALA

Porto Alegre
2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

**MODELOS DE SÉRIES TEMPORAIS:
UMA APLICAÇÃO NA PROJEÇÃO DE INADIMPLÊNCIA
DE UMA INSTITUIÇÃO FINANCEIRA**

ANGÉLICA SEGALA

Porto Alegre

2017

ANGÉLICA SEGALA

**MODELOS DE SÉRIES TEMPORAIS:
UMA APLICAÇÃO NA PROJEÇÃO DE INADIMPLÊNCIA
DE UMA INSTITUIÇÃO FINANCEIRA**

Trabalho de Conclusão de Curso submetido como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Orientador Metodológico
Prof. Dr. Eduardo de Oliveira Horta
Coorientadora
Prof^ª. Dr^ª. Lisiane Priscila Roldão Selau

Porto Alegre
2017

Instituto de Matemática e Estatística
Departamento de Estatística

**MODELOS DE SÉRIES TEMPORAIS:
UMA APLICAÇÃO NA PROJEÇÃO DE INADIMPLÊNCIA
DE UMA INSTITUIÇÃO FINANCEIRA**
Angélica Segala

Banca examinadora:

Prof. Dr. Eduardo de Oliveira Horta
IME/UFRGS

Prof^ª. Dr^ª. Lisiane Priscila Roldão Selau
IME/UFRGS

Prof. Dr. Flávio Augusto Ziegelmann
IME/UFRGS

AGRADECIMENTOS

Agradeço primeiramente a Deus pela oportunidade de realizar este desafio que é a trajetória ao longo da graduação em Estatística. Aos meus pais, meu irmão, demais familiares e amigos pelo apoio e compreensão em todos os momentos que estive ausente. Aos professores que cruzaram o meu caminho e contribuíram para a minha formação. A Prof^ª. Lisi por todos os ensinamentos, conselhos e por sempre acreditar em mim. Por fim, agradeço imensamente ao meu orientador, Prof. Eduardo, por toda atenção, paciência e disponibilidade em todos os momentos que precisei.

“Se não puder se destacar pelo talento, vença pelo esforço.” (Dave Weinbaum)

Resumo

Este trabalho propõe estimar modelos econométricos de séries temporais para prever o comportamento da Taxa de Inadimplência e do Saldo da Carteira de Crédito Pessoal de uma instituição financeira. O objetivo é produzir estimativas mais acuradas em relação aos modelos atualmente praticados. A metodologia ARIMAX é utilizada para incorporar variáveis macroeconômicas na modelagem, a fim de captar possíveis mudanças no cenário econômico. Os modelos propostos são comparados com o modelo univariado autoregressivo de ordem 1 e com o modelo atualmente praticado pela instituição financeira. Para ambas as séries, os resultados mostram que a inclusão de variáveis exógenas na estimação dos modelos produz erros quadráticos médios e erros quantílicos médios menores. Entre os quatro modelos propostos para a série que representa o Saldo da Carteira de Crédito, dois deles apresentaram redução estatisticamente significativa dos erros de previsão em relação ao modelo atualmente praticado pela instituição. Já para a série da Taxa de Inadimplência a redução dos erros de previsão foi estatisticamente significativa em relação ao modelo atualmente praticado pela instituição, para os quatro modelos propostos neste trabalho.

Palavras-chave: Taxa de inadimplência, previsão, modelos ARIMAX.

Lista de Figuras

Figura 1 - Série IGPM, ACF e PACF.....	15
Figura 2 - Série Taxa DI, ACF e PACF	15
Figura 3 - Série Taxa Referencial, ACF e PACF	16
Figura 4 - Série IPCA, ACF e PACF	16
Figura 5 - Série IPCA com ajuste sazonal, ACF e PACF	16
Figura 6 - Evolução da série Taxa de Inadimplência	21
Figura 7 - ACF e PACF da série Taxa de Inadimplência.....	21
Figura 8 - Série de previsão do modelo ARMA(4,0,4)	22
Figura 9 - Série de previsão do modelo ARMA(3,0,4)	22
Figura 10 – Série de previsão do modelo ARMAX(7,0,5) com as variáveis exógenas IPCA (defasagem 7) e TR (defasagem 4).....	23
Figura 11 – Série de previsão do modelo ARMAX(6,0,7) com a variável exógena IPCA (defasagem 2 e 3).....	23
Figura 12 – Erros absolutos de previsão 12 passos à frente da série Taxa de Inadimplência ..	24
Figura 13 - Evolução do Saldo da Carteira de Crédito Pessoal PF	26
Figura 14 - ACF e PACF da série Saldo da Carteira de Crédito em nível e diferenciada.....	26
Figura 15 - Série de previsão do modelo SARIMA(3,1,4)(1,1,0) ₁₂	27
Figura 16 - Série de previsão do modelo SARIMA(0,1,1)(0,1,1) ₁₂	27
Figura 17 - Série de previsão do modelo SARIMAX(0,1,1)(1,1,0) ₁₂ com as variáveis exógenas IGPM (sem defasagem) e Taxa DI (defasagem 6)	28
Figura 18 - Série de previsão do modelo SARIMAX(0,1,1)(1,1,0) ₁₂ com as variáveis exógenas IPCA e IGPM (sem defasagem)	28
Figura 19 – Erros absolutos de previsão 12 passos à frente da série Saldo da Carteira de Crédito	29

Lista de Tabelas

Tabela 1 - Erros médios de previsão 12 passos à frente da série Taxa de Inadimplência.....	24
Tabela 2 - P-valores do Teste Diebold-Mariano para a série Taxa de Inadimplência.....	25
Tabela 3 - Erros médios de previsão 12 passos à frente da série Saldo da Carteira de Crédito	29
Tabela 4 - P-valores do Teste Diebold-Mariano para a série Saldo da Carteira de Crédito.....	30

Sumário

1.	Introdução.....	11
2.	Metodologia.....	13
2.1.	Descrição dos Dados	13
2.2.	Modelos Box-Jenkins	16
2.3.	Modelo ARMAX.....	18
2.4.	Erros de previsão	19
2.5.	Teste Diebold-Mariano.....	19
3.	Resultados.....	20
3.1.	Taxa de Inadimplência.....	20
3.2.	Saldo da Carteira de Crédito Pessoal Pessoa Física	25
4.	Considerações finais	30
	Referências Bibliográficas.....	32

1. Introdução

A inadimplência de crédito – objeto de estudo deste trabalho – é um assunto que muito interessa ao negócio bancário, por estar diretamente relacionada à mensuração do risco de crédito do sistema financeiro. A avaliação deste indicador é fundamental em diversos aspectos, como aprovação, precificação, avaliação de garantias e provisionamento (Zaniboni, 2012). Na área financeira, o controle de risco é o que permite a sobrevivência das instituições. Sem o controle do prejuízo e da inadimplência, seriam insustentáveis a concessão de crédito e a obtenção de lucros. Assim sendo, a gestão de riscos passou a ocupar, nos últimos tempos, posição de destaque na administração financeira, especialmente em consequência da expansão do crédito, do crescimento do mercado e da globalização. No Brasil, em particular, esse crescimento da gestão de riscos foi também motivado pela maior estabilidade econômica alcançada após a implantação do Plano Real e do controle da inflação (Bueno, 2003).

As três versões do Acordo de Capital de Basiléia, concebidas respectivamente em 1988, 2004 e 2010, são uma evidência da preocupação recorrente dos bancos centrais e da indústria bancária com o gerenciamento do risco de crédito do sistema (Marins e Neves, 2013). O Acordo foi concebido com o objetivo de criar exigências mínimas de capital para instituições financeiras como forma de fazer face ao risco de crédito (Fonte: Banco Central). Atualmente o acordo encontra-se na sua versão chamada de Basiléia III, e visa fortalecer e aprimorar as regras anteriores para promover um setor bancário mais resiliente. Entre as reformas destacam-se medidas para aumentar a capacidade de as instituições financeiras absorverem crises, melhorias de gestão e transparência no funcionamento e nas operações (Basel III, 2010).

O crédito sempre foi um instrumento de extrema importância na realização de negócios, viabilizando recursos para o crescimento e desenvolvimento econômico. Sob o aspecto financeiro, a concessão de crédito significa dispor a um tomador recursos para atender as suas mais diversas necessidades, mediante uma remuneração paga pelo uso do dinheiro no tempo, conhecida como juros (Bueno, 2003). Quando acontece uma concessão de recursos, a instituição financeira passa a possuir o chamado Risco de Crédito. Segundo Jorion (1997), este risco pode ser definido como sendo a possibilidade de a contraparte não cumprir com as obrigações monetárias contratuais relativas às transações financeiras. Esse não cumprimento das obrigações contratuais é chamado de inadimplência.

A intermediação financeira pode ser considerada a principal atividade bancária e, na maioria dos sistemas financeiros, o risco de crédito representa o principal risco a que as instituições estão expostas (Cihák, 2007). Neste sentido, sendo a taxa de inadimplência a medida mais comum para a mensuração do risco de crédito, a previsão da mesma é fundamental para nortear as políticas de crédito e, conseqüentemente, decisões acerca do controle da inadimplência. Logo, prever acuradamente este indicador torna-se importante para a maximização do lucro das instituições.

Estudos sobre a inadimplência ganharam um impulso considerável ao longo das últimas décadas. Diversas classes de modelos têm sido propostas para mensurar e prever o risco de crédito, destacando-se alguns trabalhos que procuram mostrar como o risco de crédito evolui ao longo do ciclo econômico. Por exemplo, Linardi (2008) utiliza técnicas de séries temporais para investigar a relação entre a taxa de inadimplência de empréstimos de bancos brasileiros e o cenário macroeconômico do Brasil no período de 2000 a 2007. Aplicando o modelo VAR para descrever a evolução da taxa de inadimplência, o autor encontrou indícios de que esta é influenciada pelo hiato do produto, pela taxa de juros nominais e pelo rendimento médio da população ocupada.

Sample (2013), por sua vez, estuda o comportamento da inadimplência através dos métodos FAVAR (Factor-Augmented Vector Autoregressive) e FAVECM (Factor Augmented Error Correction Models) e compara as projeções com modelos ARIMA e SARIMA. Para avaliação da eficácia preditiva o autor utilizou a metodologia MCS (Model Confidence Set), que permitiu observar a não supremacia das classes de modelos FAVAR e FAVECM em relação aos modelos univariados.

Por fim, Zaniboni (2013) apresenta modelos de previsão para a taxa de inadimplência do sistema financeiro no Brasil utilizando um conjunto abrangente de variáveis macroeconômicas. A base de dados mensais compreende o período de março de 2000 a dezembro de 2012 e foi obtida do banco de dados do Banco Central e do IPEA. Utilizou-se modelos de séries temporais ARIMA e ARIMAX para prever a inadimplência. Variáveis explicativas indicadoras do saldo da carteira de crédito por região, dívida pública interna e juros se mostraram estatisticamente significativas para explicar a inadimplência do sistema financeiro no Brasil.

Nesse contexto, o objetivo do presente trabalho é utilizar modelos econométricos de séries temporais para a previsão do comportamento da Taxa de Inadimplência e do Saldo da

Carteira de Crédito Pessoal de uma instituição financeira de atuação nacional. Através do modelo ARIMAX serão incorporadas variáveis macroeconômicas (exógenas¹) com o objetivo de captar possíveis mudanças no cenário econômico. A partir disso, serão construídas previsões fora da amostra para comparar a eficácia de projeção dos diferentes modelos. Para quantificar o desempenho de cada um dos modelos, será calculado o erro quadrático médio e o erro quantílico médio. A comparação das previsões geradas pelos modelos será realizada através do teste Diebold-Mariano (Diebold e Mariano, 1995).

Este trabalho está estruturado em quatro partes principais, incluindo a presente introdução. Inicialmente apresentam-se conceitos básicos relacionados à inadimplência de crédito e uma breve revisão bibliográfica do tema. A seção seguinte contempla a metodologia, onde serão apresentadas as variáveis em estudo, os modelos de séries temporais ARIMAX e os critérios utilizados para mensurar a capacidade preditiva dos mesmos. Posteriormente, apresentam-se os resultados e, por fim, as conclusões do trabalho.

2. Metodologia

Nesta seção serão apresentadas as variáveis em estudo e a metodologia utilizada para a construção dos modelos econométricos que pretendem prever o comportamento da Taxa de Inadimplência e do Saldo da Carteira de Crédito Pessoal de uma instituição financeira.

2.1. Descrição dos Dados

Para a realização deste trabalho foram utilizados dados reais da Carteira de Crédito Pessoal de uma instituição financeira de atuação nacional. A série mensal compreende o período de janeiro de 2012 a dezembro de 2016, totalizando 60 observações. Serão considerados modelos para descrever e prever o Saldo da Carteira de Crédito e a Taxa de Inadimplência da mesma.

Séries temporais de algumas variáveis macroeconômicas de interesse foram obtidas junto à instituição financeira, abrangendo o período de junho de 2011 a dezembro de 2016. Os meses anteriores a janeiro de 2012 foram utilizados para incorporar nos modelos variáveis exógenas defasadas. A seleção de covariáveis se restringiu às que foram fornecidas, pois tais

¹ Variáveis exógenas são aquelas definidas fora do modelo, introduzindo informações externas. Formalmente, são variáveis não correlacionadas com os termos de erro (inovações) do modelo.

variáveis já têm suas previsões calculadas periodicamente pela instituição financeira, facilitando a posterior implementação dos modelos preditivos aqui propostos. No entanto, para este trabalho as previsões das variáveis macroeconômicas foram obtidas através de um modelo AR(1), pois não foi possível obter em tempo hábil as previsões das mesmas junto à instituição.

A seguir detalham-se as variáveis utilizadas:

- *Composição da carteira de crédito*: Saldo de todos os contratos de crédito pessoal para pessoa física.
- *Percentual de inadimplência*: Relação entre o saldo em atraso entre 90 e 360 dias dos contratos de crédito pessoal, e o saldo de todos os contratos na data base.
- *IGP-M (Índice Geral de Preços do Mercado)*: Indicador que mede a taxa de inflação e foi instituído para ser usado no reajuste de operações financeiras, especialmente as de longo prazo.
- *IPCA (Índice de Preços ao Consumidor Amplo)*: Indicador que mede a taxa de inflação do custo de vida de segmentos da população.
- *Taxa Selic*: Taxa básica de juros da economia no Brasil, a Selic é a taxa média ajustada dos financiamentos diários apurados no Sistema Especial de Liquidação e de Custódia (Selic) para títulos Federais (Fonte: Banco Central).
- *Taxa DI*: Também conhecida como CDI (Certificado de Depósito Interbancário), a taxa DI é a principal referência para a rentabilidade dos investimentos de renda fixa e de fundos de investimento.
- *TJLP (Taxa de Juros de Longo Prazo)*: A TJLP é calculada com base em dois parâmetros: uma meta de inflação, que tem como índice padrão o IPCA; e um prêmio de risco.
- *TR (Taxa referencial)*: A TR foi criada no Plano Collor II para ser o principal índice brasileiro, uma taxa básica referencial dos juros a serem praticados no mês vigente e que não refletissem a inflação do mês anterior. É a taxa utilizada para corrigir o valor da Caderneta de Poupança, FGTS, Títulos Públicos e alguns financiamentos imobiliários.

- *Taxa de câmbio PTAX*: É uma taxa de câmbio calculada durante o dia pelo Banco Central. Consiste na taxa média do dia apurada com base nas operações realizadas no mercado intercmbiário.

Para todas as séries acima apresentadas realizaram-se a análise gráfica e o teste ADF para verificar a estacionariedade. Entre as variáveis macroeconômicas, IGPM e IPCA rejeitaram a hipótese nula de presença de raiz unitária na série. A inspeção visual, através da ACF, também sugere a estacionariedade destas séries. Sendo assim, não foi necessário tomar a diferença. Todas as demais séries macroeconômicas utilizadas como variáveis exógenas foram diferenciadas para atingir a estacionariedade.

Dentre todos os modelos estimados para a Taxa de Inadimplência, os que produziram menores erros foram os modelos que incorporaram as variáveis exógenas IPCA e Taxa Referencial na modelagem (ver Tabela 1). Já para a série que representa o Saldo da Carteira de Crédito, as variáveis exógenas que produziram modelos com previsões mais acuradas foram o IGPM, a Taxa DI e também o IPCA (ver Tabela 3). A seguir as Figuras 1, 2 e 3 exibem as séries do IGPM, Taxa DI e Taxa Referencial - variáveis exógenas que foram utilizadas nos modelos propostos neste trabalho - bem como suas funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

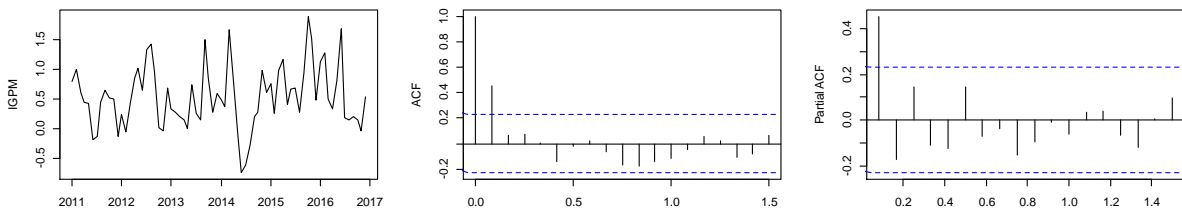


Figura 1 - Série IGPM, ACF e PACF

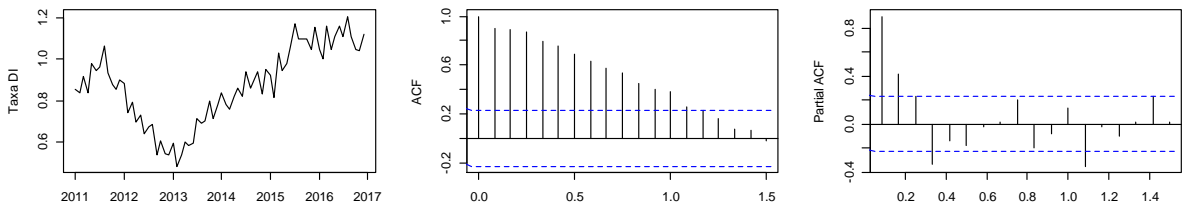


Figura 2 - Série Taxa DI, ACF e PACF

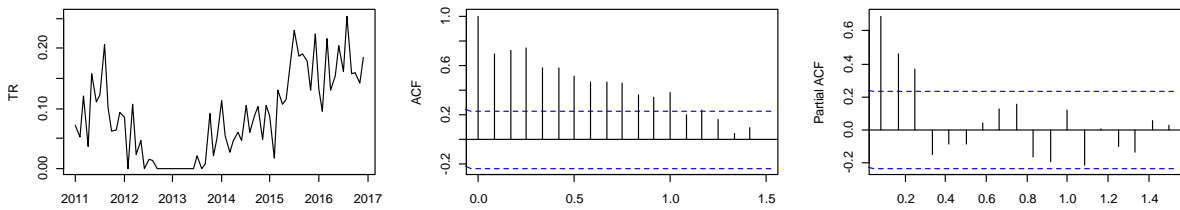


Figura 3 - Série Taxa Referencial, ACF e PACF

A Figura 4 exibe a série do IPCA, sua função de autocorrelação (ACF) e sua função de autocorrelação parcial (PACF). Através da ACF, visualiza-se picos que se repetem a cada 12 meses, sugerindo que há uma sazonalidade de ordem 12 na série. Assim, para esta série, realizou-se um ajuste sazonal através do método X-13ARIMA-SEATS implementado na função *seas* do pacote *seasonal* do *software R*. A Figura 5 exibe a série após o ajuste, onde através da ACF pode-se verificar que a sazonalidade foi removida.

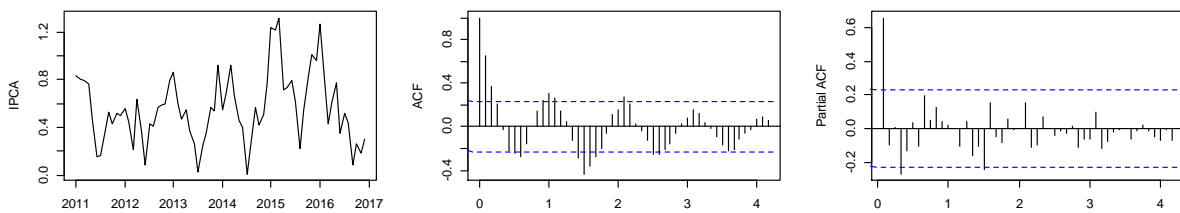


Figura 4 - Série IPCA, ACF e PACF

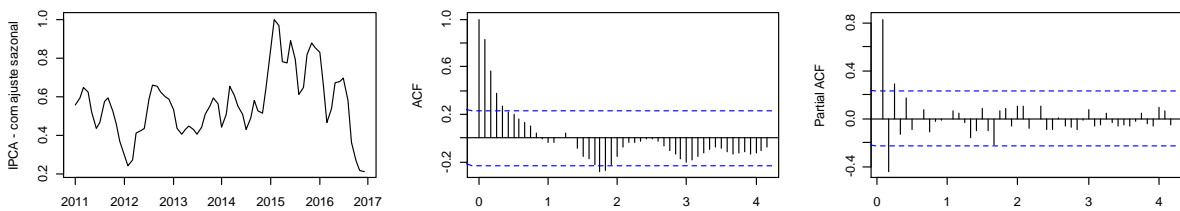


Figura 5 - Série IPCA com ajuste sazonal, ACF e PACF

2.2. Modelos Box-Jenkins

A abordagem proposta por Box e Jenkins (1976) é uma metodologia bastante utilizada para a análise de modelos paramétricos de séries temporais. Essa metodologia consiste no ajuste de modelos ao conjunto de dados. Morettin e Toloí (2004) recomendam um ciclo iterativo para estimar um modelo ARIMA, em que os estágios geralmente são dados por: identificação do modelo, estimação dos parâmetros e verificação da qualidade do modelo ajustado.

Diversos modelos amplamente explorados na literatura estão aninhados na classe de modelos ARIMA. Por exemplo, quando o valor atual de uma série temporal é explicado por

seus valores passados somados a um termo de erro, um modelo autoregressivo de ordem p , ou AR(p), pode ser utilizado. Já um processo de médias móveis de ordem q , ou MA(q), é indicado quando os valores passados dos erros explicam o valor atual da série. Quando um processo é autoregressivo e de médias móveis, ajusta-se o modelo por ARMA(p,q), que é dado pela expressão (1):

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (1)$$

onde,

ϕ_i são os coeficientes autoregressivos;

θ_j são os coeficientes de médias móveis;

ε_t é um processo ruído branco com média zero e variância σ_ε^2 constante.

Uma exigência desta metodologia é que a série temporal seja estacionária, ou seja, que sua média, variância e estrutura de correlação não mudem ao longo do tempo. Diante de uma série não estacionária deve-se incluir o termo relativo à integração I, ou seja, a diferenciação de ordem d das séries e neste caso tem-se um modelo ARIMA(p,d,q).

Portanto, pode-se descrever todos os modelos vistos anteriormente utilizando a nomenclatura ARIMA, isto é:

(i) ARIMA($p,0,0$) = AR(p);

(ii) ARIMA($0,0,q$) = MA(q);

(iii) ARIMA($p,0,q$) = ARMA(p,q).

E no caso de uma série não estacionária utiliza-se o modelo completo, ARIMA(p,d,q) com d diferenças na série original.

Após a modelagem podem ser calculadas previsões h passos à frente. As previsões apresentadas para modelos ARIMA podem ser obtidas conforme a equação (2), que corresponde a solução da equação de diferenças.

$$\hat{Z}_{t+h} = \phi_1 Z_{t+h-1} + \dots + \phi_{p+d} Z_{t+h-p-d} - \theta_1 \varepsilon_{t+h-1} - \dots - \theta_q \varepsilon_{t+h-q} + \varepsilon_{t+h} \quad (2)$$

A metodologia Box-Jenkins explora a autocorrelação entre os valores da série em diferentes instantes t sucessivos. Quando a autocorrelação é observada dentro de um período de tempo, considera-se uma estação de sazonalidade s e a série pode ser ajustada por um modelo descrito como SARIMA(p,d,q)(P,D,Q) $_s$. Os modelos SARIMA são definidos conforme a

equação (3) e as previsões h passos à frente são descritas pela equação (4), que corresponde a solução da equação de diferenças.

$$\Phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Z_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (3)$$

$$\Phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D \hat{Z}_t(h) = 0, \quad (4)$$

onde,

$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_p B^{sp}$ é o operador autoregressivo sazonal de ordem P;

$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ}$ é o operador de médias móveis sazonal de ordem Q;

$(1-B)^d$ é o operador diferença simples, com d indicando o número de diferenças simples;

$(1-B^s)^D$ é o operador diferença sazonal, com D indicando o número de diferenças sazonais;

$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é o operador autoregressivo de ordem p;

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o operador de médias móveis de ordem q;

s é o período sazonal;

ε_t é um processo ruído branco com média zero e variância σ_ε^2 constante.

2.3. Modelo ARMAX

O modelo ARMAX é uma variação da metodologia Box-Jenkins. Este método consiste em ajustar modelos autoregressivos de médias móveis incluindo variáveis exógenas ao conjunto de dados. Essa classe de modelos será utilizada para verificar a relação entre as variáveis macroeconômicas e as variáveis a serem preditas.

A notação ARMAX(p, q, w) refere-se ao modelo com p termos autoregressivos, q termos de médias móveis e w termos de variáveis exógenas. É dado pela expressão (5):

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \beta_{1,0} X_{1,t} + \beta_{1,1} X_{1,t-1} + \dots + \beta_{1,n_1} X_{1,t-n_1} + \dots + \beta_{w,0} X_{w,t} + \beta_{w,1} X_{w,t-1} + \dots + \beta_{w,n_w} X_{w,t-n_w}, \quad (5)$$

onde,

w é o número de variáveis exógenas;

n_k é o número de defasagens da k-ésima variável exógena;

$X_{k,t}$ é a k-ésima variável exógena no instante t;

$\beta_{k,t}$ é o coeficiente associado a k-ésima variável exógena no instante t.

2.4. Erros de previsão

A precisão da previsão refere-se ao quão próximos os valores preditos estão dos realizados (*ex post*), segundo algum critério ou função perda. Ao avaliar a eficiência preditiva de diferentes modelos, necessita-se de uma medida de acurácia, bem como de um critério de decisão para julgar qual previsão é mais eficiente. Neste trabalho será utilizada a técnica de validação cruzada para avaliar a capacidade de generalização dos modelos. O conjunto de dados será particionado em dois subconjuntos, sendo o período de janeiro de 2012 a dezembro de 2015 utilizado para a estimação dos parâmetros e o período de janeiro de 2016 a dezembro de 2016 utilizado para a validação do modelo. Após a estimação dos modelos executam-se as previsões e calcula-se duas medidas de acurácia para cada modelagem. Serão utilizadas duas funções perda neste trabalho: o erro quadrático médio e o erro quantílico médio, definidos da seguinte maneira: sendo o erro de previsão entre o valor real y e o valor predito \hat{y} representado por: $e_t = y_t - \hat{y}_t$, o erro quadrático médio é dado pela soma do quadrado dos erros de previsão, dividido pelo número de previsões, conforme a expressão (6):

$$EQM = \frac{1}{n} \sum_{t=1}^n e_t^2. \quad (6)$$

Enquanto o EQM minimiza os erros de previsão da média condicional, o erro quantílico médio minimiza os erros que prevêm os quantis da distribuição. Este critério foi utilizado, pois é de interesse da instituição financeira que o modelo não subestime as previsões. Sendo assim optou-se por minimizar os erros referentes ao quantil 60. Conforme a expressão (7):

$$EQtM = \frac{1}{n} \sum_{t=1}^n \rho_{\tau}(e_t), \quad (7)$$

$$\rho_{\tau}(x) = (\tau - I[x < 0])x, \text{ onde } \tau = 0,6$$

2.5. Teste Diebold-Mariano

O teste Diebold-Mariano (Diebold e Mariano, 1995) compara previsões geradas por dois modelos concorrentes, para testar se as previsões fora da amostra de ambos os modelos são estatisticamente iguais. O teste é feito a partir de uma função perda (L) associada a um erro de previsão (e_t). A perda diferencial entre as duas previsões pode ser definida como:

$d_t = L(e_{1t}) - L(e_{2t})$. Para verificar se as previsões são equivalentes testa-se se $E(d_t) = 0$, ou seja, a hipótese nula é de que as duas previsões têm a mesma capacidade preditiva. Usualmente, utiliza-se o erro quadrático médio ou o erro médio absoluto de previsão para calcular a função perda. Neste trabalho, utilizaremos o erro quadrático médio e o erro quantílico médio.

3. Resultados

Nesta seção apresenta-se os resultados da estimação de modelos de séries temporais para a previsão da Taxa de Inadimplência e do Saldo da Carteira de Crédito Pessoal, bem como as comparações da capacidade preditiva dos mesmos.

Após a identificação da ordem dos modelos, a estimação dos parâmetros é feita pelo método de máxima verossimilhança, utilizando a função *arima* do pacote *stats* do *software R*. O próximo passo será verificar, através dos resíduos, se o modelo representa adequadamente os dados. O teste de Box e Pierce (1970) será aplicado para avaliar a autocorrelação residual. A hipótese nula do teste é de que os resíduos são um processo de ruído branco. Se a hipótese nula não for rejeitada, temos o indicativo de que os resíduos constituem um processo de ruído branco e o modelo é adequado. Diante do modelo escolhido, o interesse é a previsão para um horizonte de 12 meses, que será realizada utilizando a equação de diferenças através da função *predict* do pacote *stats* do *software R*.

3.1. Taxa de Inadimplência

A primeira fase da estimação de um modelo Box-Jenkins é a avaliação das características da série temporal. A Figura 6 apresenta a evolução da série da taxa de inadimplência para o período de ajuste (jan/2012 a dez/2015) e para o período de validação (jan/2016 a dez/2016). A Figura 7, apresenta a função de autocorrelação (ACF) e autocorrelação parcial (PACF). Através do teste de Dickey-Fuller (ADF), obteve-se um p-valor de 0,37, não rejeitando a hipótese nula do teste. Neste teste, a hipótese nula é de que a série contém pelo menos uma raiz unitária, ou seja, pelo teste conclui-se que a série pode não ser estacionária. No entanto, a inspeção gráfica da ACF mostra um decaimento rápido para valores próximos de zero, sendo este um indicativo de que a série é estacionária. Como o teste ADF possui baixo poder, principalmente e para amostras pequenas, tratou-se a série como estacionária.

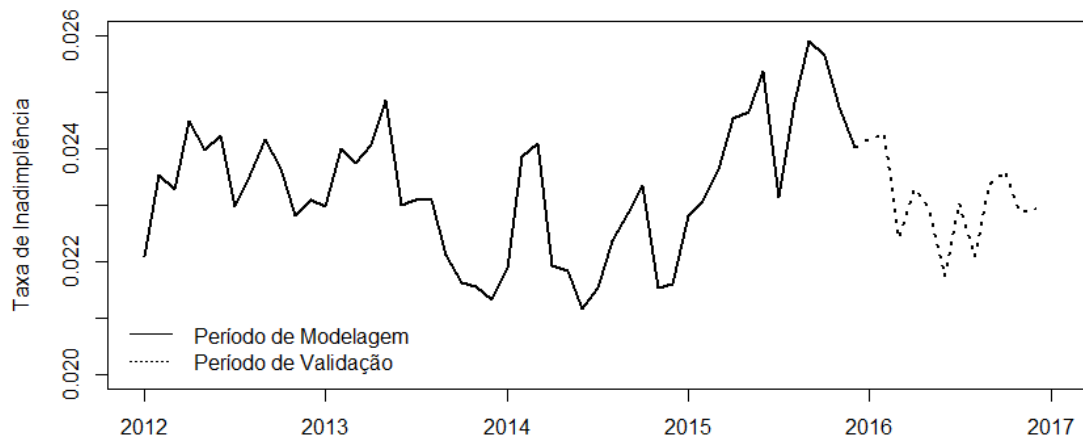


Figura 6 - Evolução da série Taxa de Inadimplência

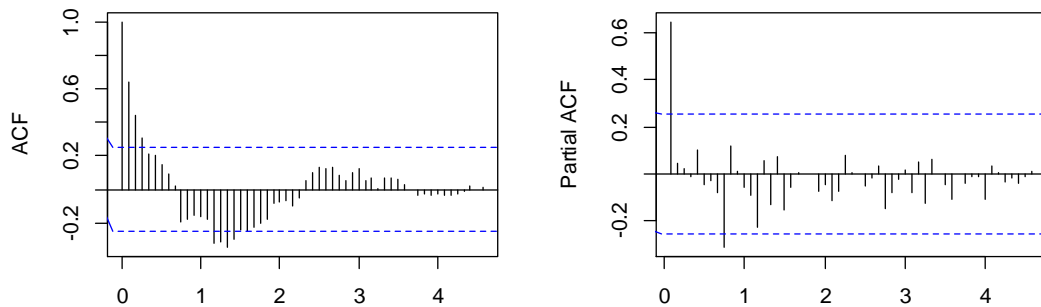


Figura 7 - ACF e PACF da série Taxa de Inadimplência

Inicialmente estimaram-se os parâmetros de modelos $ARMA(p, q)$ com $p, q = 1, \dots, 8$. Os parâmetros de cada modelo foram estimados e as previsões dos próximos doze meses foram comparadas com os valores realizados (*ex post*) da série, calculando-se as duas medidas de acurácia para comparação do poder preditivo de cada modelo.

Pela medida do erro quadrático médio foi selecionado o modelo $ARMA(4,0,4)$ e pelo erro quantílico médio o melhor modelo foi o $ARMA(3,0,4)$. Os ajustes foram avaliados através do teste Box-Pierce e ambos não rejeitaram a hipótese nula de que os resíduos são ruído branco. As Figuras 8 e 9 exibem as séries de previsões realizadas por cada um dos modelos e seus respectivos intervalos de confiança.

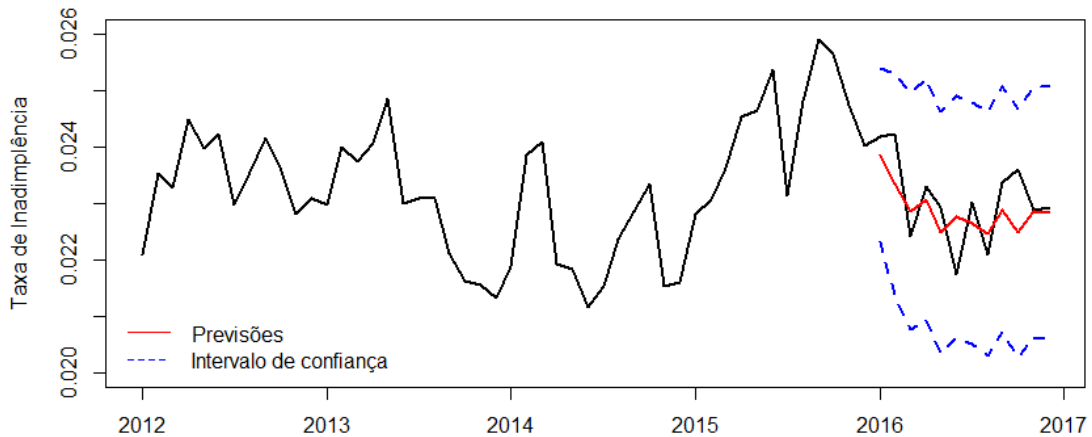


Figura 8 - Série de previsão do modelo ARMA(4,0,4)

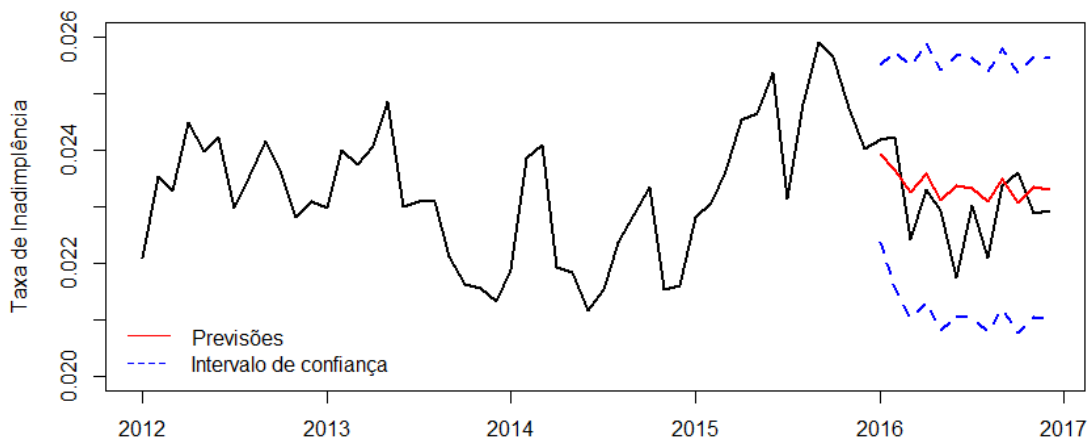


Figura 9 - Série de previsão do modelo ARMA(3,0,4)

Após a análise univariada, estimaram-se parâmetros de modelos ARMAX(p, q) com $p, q = 0, \dots, 8$ com a inclusão das variáveis exógenas individualmente, com defasagens de 0 a 7, resultando em um total de 5.832 modelos. Na próxima etapa, foi realizada a estimação dos parâmetros de modelos ARMAX(p, q) com $p, q = 0, \dots, 8$ com a inclusão das variáveis exógenas combinadas duas a duas, incluindo combinações de 0 a 7 defasagens. Nesta etapa da modelagem obteve-se um total de 55.272 modelos.

O modelo que apresentou o menor erro quadrático médio foi o ARMAX(7,0,5) com as variáveis exógenas IPCA na defasagem 7 e Taxa Referencial na defasagem 4. Pela medida do erro quantílico foi selecionado o modelo ARMAX(6,0,7) com a variável exógena IPCA com 2 e 3 defasagens, ou seja, foram utilizados valores do IPCA nos tempos $t-2$ e $t-3$ para explicar o valor da série da Taxa de Inadimplência no tempo t . A autocorrelação residual foi avaliada através do teste Box-Pierce e ambos os modelos não rejeitaram a hipótese de que os resíduos

são um processo de ruído branco. As Figuras 10 e 11 exibem as séries de previsões realizadas por cada um dos modelos ARMAX selecionados e seus respectivos intervalos de confiança.

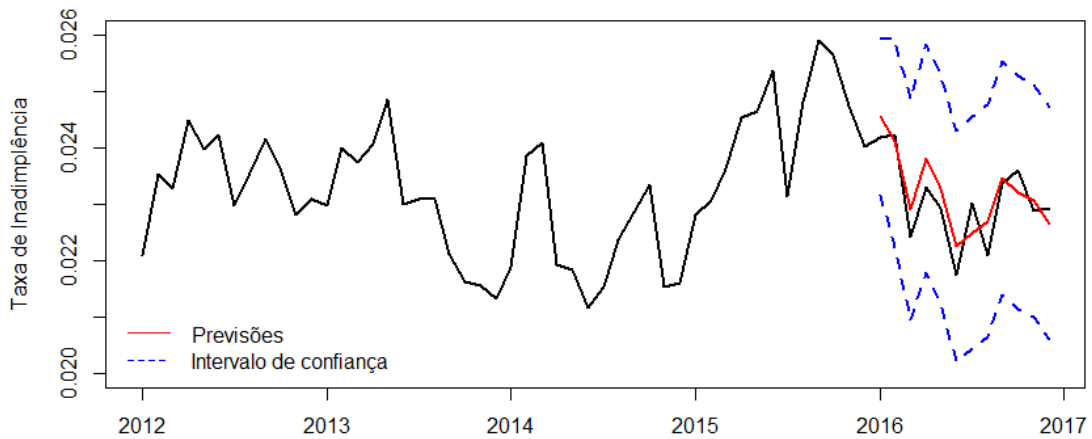


Figura 10 – Série de previsão do modelo ARMAX(7,0,5) com as variáveis exógenas IPCA (defasagem 7) e TR (defasagem 4)

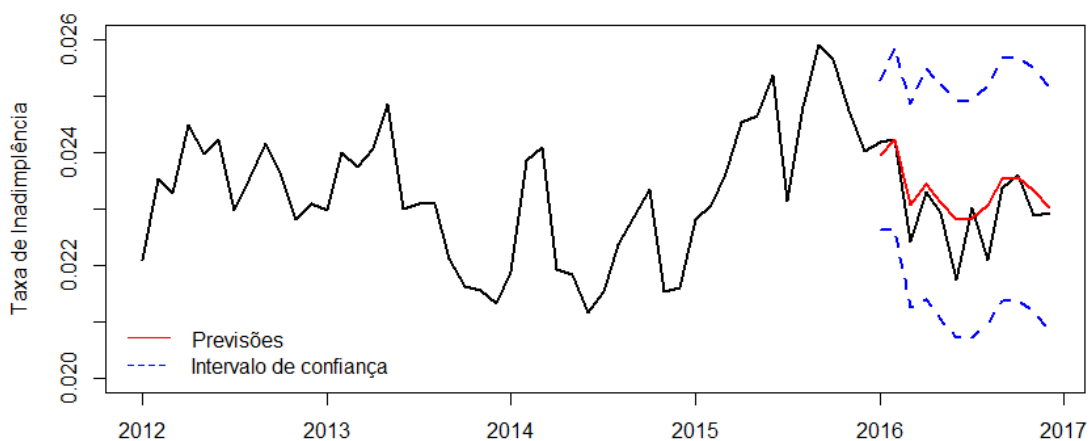


Figura 11 – Série de previsão do modelo ARMAX(6,0,7) com a variável exógena IPCA (defasagem 2 e 3)

A Tabela 1 apresenta os erros médios de previsão calculados para cada modelo, considerando as previsões para um horizonte de 12 meses. Pelos erros médios observa-se que a inclusão das variáveis exógenas reduz os erros de previsão em relação aos demais modelos.

Tabela 1 - Erros médios de previsão 12 passos à frente da série Taxa de Inadimplência

Modelo	EQM	EQ-60
Modelo atual praticado	0,004548	0,001737
AR(1)	0,000731	0,000254
ARMA(4,0,4)	0,000584	0,000258
ARMA(3,0,4)	0,000686	0,000241
ARMAX(7,0,5) - IPCA e TR	0,000401	0,000167
ARMAX(6,0,7) - IPCA	0,000492	0,000146

Os erros de previsão foram comparados com o modelo atualmente praticado pela instituição financeira e com o modelo univariado AR(1). A Figura 12 exibe um comparativo dos erros absolutos de previsão de todos os modelos. Podemos observar uma supremacia dos modelos AR(1), ARMA e ARMAX em relação ao modelo atualmente praticado em todos os horizontes de tempo. Em relação ao modelo AR(1), visualiza-se superioridade preditiva dos modelos ARMA e ARMAX, principalmente em horizontes de $t+1$ até $t+8$, já para horizontes de $t+9$ até $t+12$ os erros de previsão são semelhantes. Dentre os 4 modelos propostos, pode-se destacar o modelo ARMAX(6,0,7) com as covariáveis IPCA com 2 e 3 defasagens, o qual possui uma menor variação dos erros de previsão em todos os horizontes de tempo. Além disso, subestima os valores em apenas 3 horizontes de tempo, com erros muito pequenos, o que do ponto de vista prático é importante.

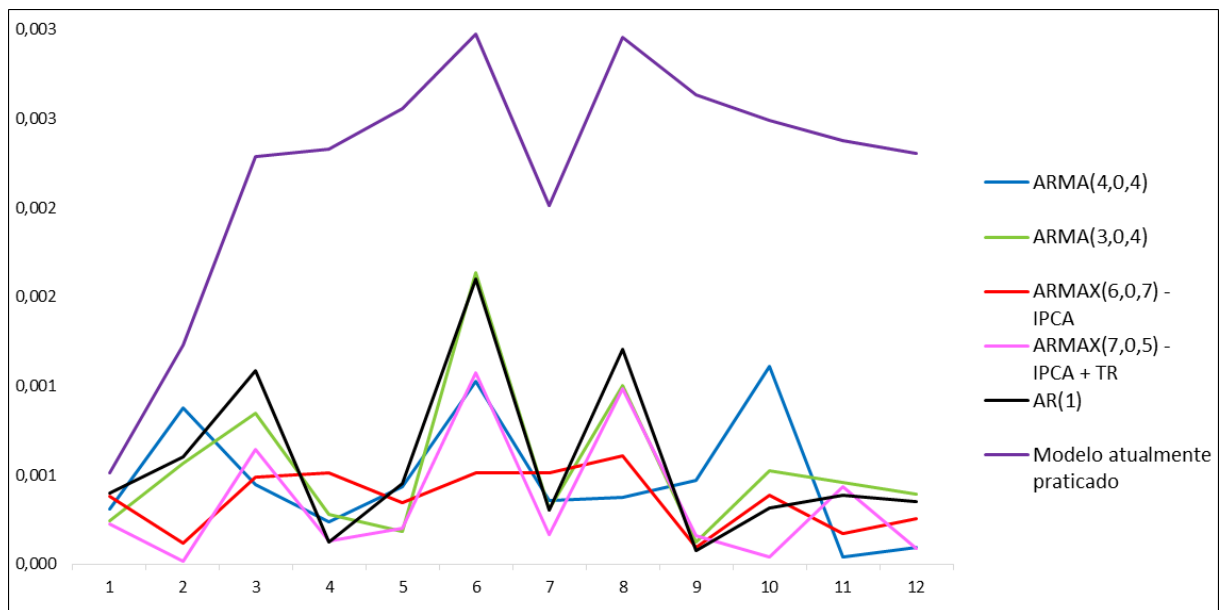


Figura 12 – Erros absolutos de previsão 12 passos à frente da série Taxa de Inadimplência

Para verificar se o poder preditivo dos modelos difere significativamente, foi realizado o teste Diebold-Mariano (1995). Para aplicação deste teste, foram realizadas 12 previsões um passo à frente para cada modelo. A hipótese nula do teste é de que os modelos comparados possuem a mesma capacidade preditiva. A Tabela 2 apresenta os p-valores do teste em que verifica-se que as previsões dos modelos ARMA e ARMAX não diferem significativamente do modelo AR(1). Já em comparação com o modelo atualmente praticado pela instituição, há superioridade preditiva significativa dos modelos ARMA e ARMAX propostos neste trabalho.

Tabela 2 - P-valores do Teste Diebold-Mariano para a série Taxa de Inadimplência

Modelo	EQM		EQ-60	
	AR(1)	Modelo atualmente praticado	AR(1)	Modelo atualmente praticado
ARMA(4,0,4)	0,2046	0,0027	0,6678	0,0028
ARMA(3,0,4)	0,9917	0,0019	0,6513	0,0003
ARMAX(7,0,5) - IPCA e TR	0,6669	0,0080	0,6727	0,0191
ARMAX(6,0,7) - IPCA	0,5697	0,0033	0,8585	0,0029

3.2. Saldo da Carteira de Crédito Pessoal

Na segunda modelagem ajustaram-se modelos para prever a evolução da série que representa o Saldo da Carteira de Crédito Pessoal. A Figura 13 mostra a evolução da série para o período de ajuste (jan/2012 a dez/2015) e para o período de validação (jan/2016 a dez/2016). A figura 14, apresenta a função de autocorrelação (ACF) e autocorrelação parcial (PACF) para a série original e para a série diferenciada. Graficamente visualiza-se um decaimento lento da ACF, o que é um indicativo de não estacionariedade da série. Através do teste ADF, obteve-se um p-valor de 0,296, ou seja, não rejeitamos a hipótese nula de presença de raiz unitária. Sendo assim, foi tomada a primeira diferença da série para fins de modelagem. Após a diferenciação, pela análise gráfica visualiza-se picos que se repetem a cada 12 meses; esta característica é indicativo de uma sazonalidade de ordem 12 na série. Assim, um modelo SARIMA foi ajustado.

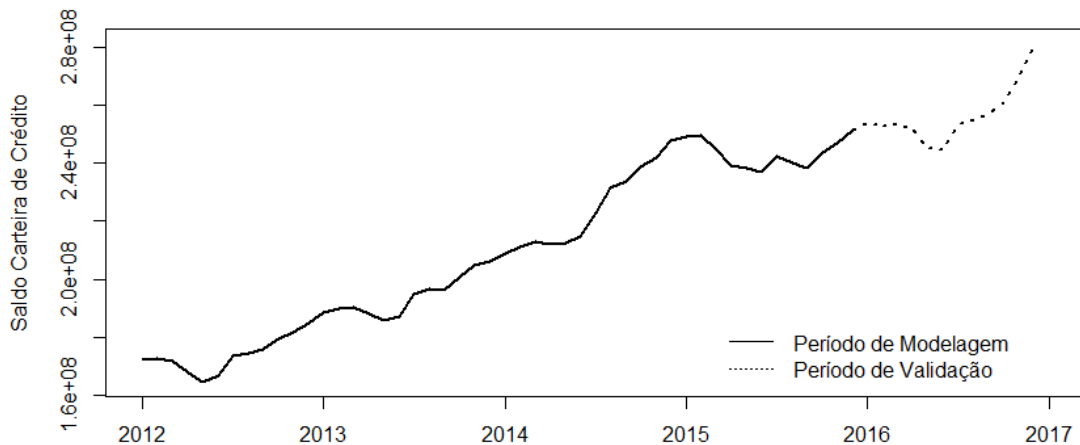


Figura 13 - Evolução do Saldo da Carteira de Crédito Pessoal

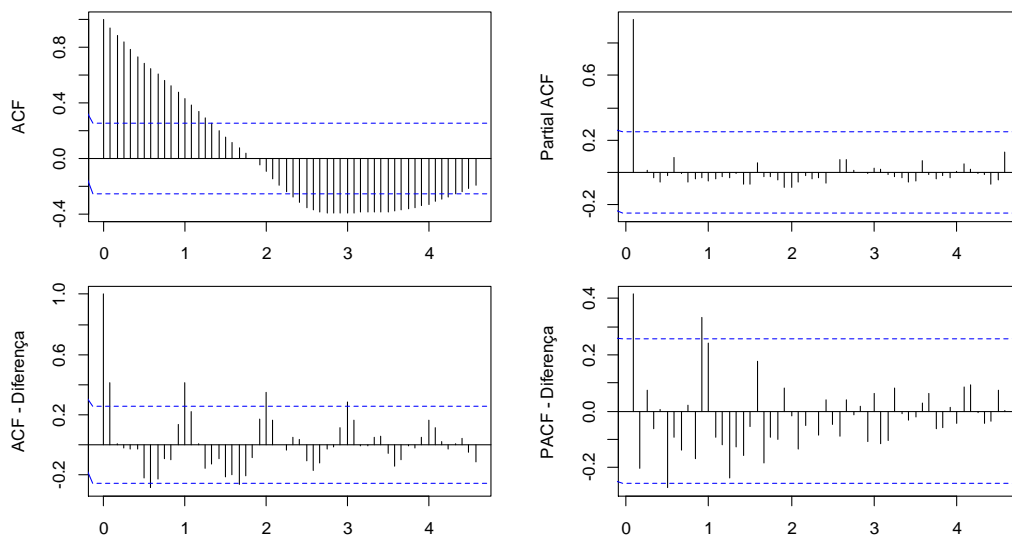


Figura 14 - ACF e PACF da série Saldo da Carteira de Crédito em nível e diferenciada

Para esta série, a estimação dos parâmetros se deu em três etapas. Primeiramente estimou-se parâmetros de modelos SARIMA(p,d,q)(P,D,Q)_s com $p, q = 0, \dots, 4$ e $P, Q = 0, \dots, 1$, sem a inclusão de variáveis exógenas. Para cada modelo foram geradas previsões para os próximos doze meses, e estas foram comparadas com os valores realizados da série.

Pela medida do erro quadrático médio foi selecionado o modelo SARIMA (3,1,4)(1,1,0)₁₂. O ajuste foi avaliado através do teste Box-Pierce, que obteve um p-valor de 0,824, não rejeitando a hipótese nula de que os resíduos são ruído branco. A significância dos parâmetros foi avaliada, e apesar do modelo apresentar dois parâmetros não significativos manteve-se a escolha, pois apresenta um melhor poder preditivo em relação aos demais. Pelo erro quantílico médio, o modelo SARIMA (0,1,1)(0,1,1)₁₂ foi selecionado. O teste Box-Pierce

apresentou p-valor de 0,691 e todos os parâmetros foram significativos. As Figuras 11 e 12 exibem as séries de previsões realizadas por cada um dos modelos e seus respectivos intervalos de confiança.

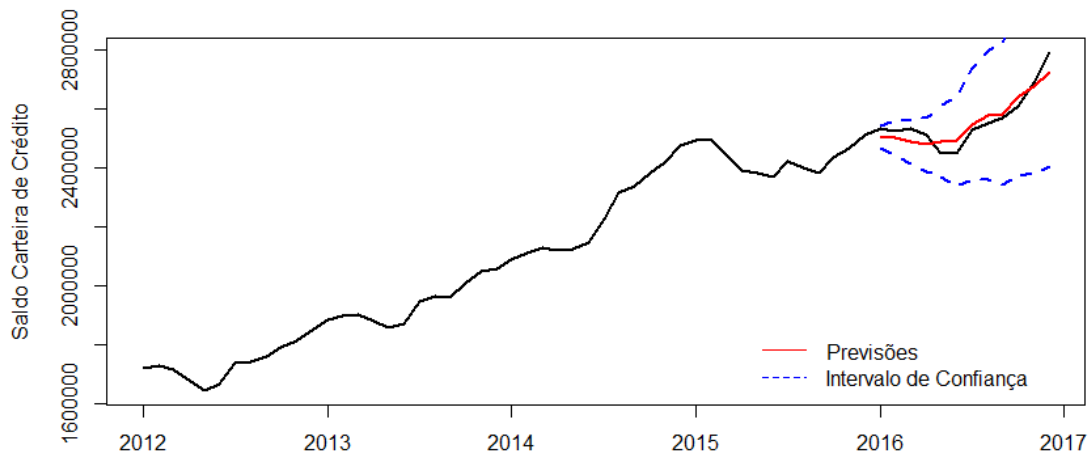


Figura 15 - Série de previsão do modelo SARIMA(3,1,4)(1,1,0)₁₂

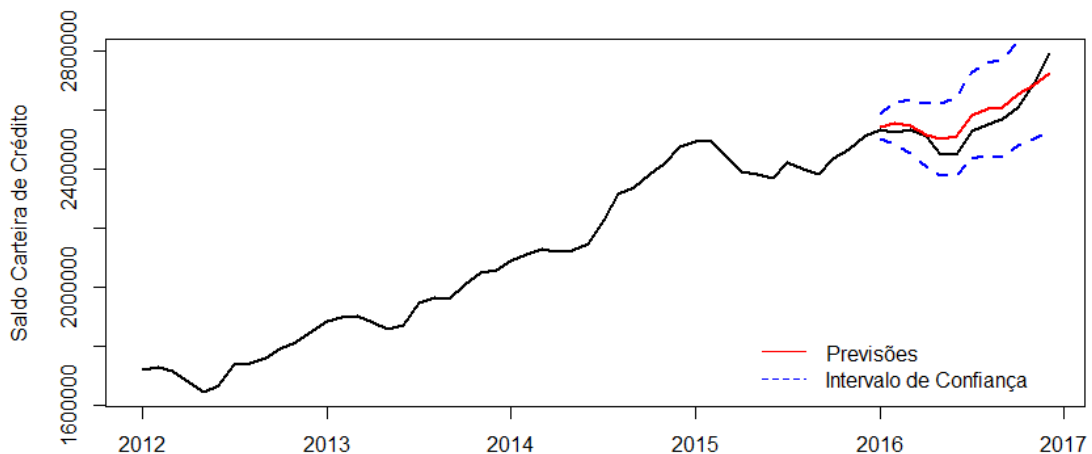


Figura 16 - Série de previsão do modelo SARIMA(0,1,1)(0,1,1)₁₂

A próxima etapa foi estimar parâmetros de modelos SARIMAX(p,d,q)(P,D,Q)_s com $p, q = 0, \dots, 3$ e $P, Q = 0, \dots, 1$, com a inclusão das variáveis exógenas individualmente. Foram utilizadas diferentes defasagens, isto é, estimou-se modelos utilizando lags variando de 0 a 7 para todas as variáveis exógenas. Ainda, estimou-se parâmetros de modelos SARIMAX(p,d,q)(P,D,Q)_s com $p, q = 0, \dots, 1$ e $P, Q = 0, \dots, 1$, incluindo variáveis exógenas combinadas duas a duas e combinando diferentes defasagens. Aqui a variação de parâmetros p, q restringiu-se em 0 ou 1 devido ao grande número de possíveis combinações, gerando um total de 40.896 modelos.

Pela medida do erro quadrático médio foi selecionado o modelo SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IGPM sem nenhuma defasagem e Taxa DI na defasagem 6. O teste Box-Pierce apresentou p-valor de 0,489, indicando que os resíduos são ruído branco. Apesar de a inclusão das variáveis exógenas ter reduzido o erro quadrático médio em relação aos modelos sem as variáveis exógenas, seus parâmetros não foram significativos. Pelo erro quantílico médio, o modelo selecionado foi o SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IPCA e IGPM sem nenhuma defasagem. O teste Box-Pierce indicou que os resíduos são ruído branco (p-valor = 0,487). Neste modelo os parâmetros das variáveis exógenas também não foram significativos, apesar de reduzirem o erro quantílico médio. As Figuras 17 e 18 exibem as séries de previsões realizadas por cada um dos modelos e seus respectivos intervalos de confiança.

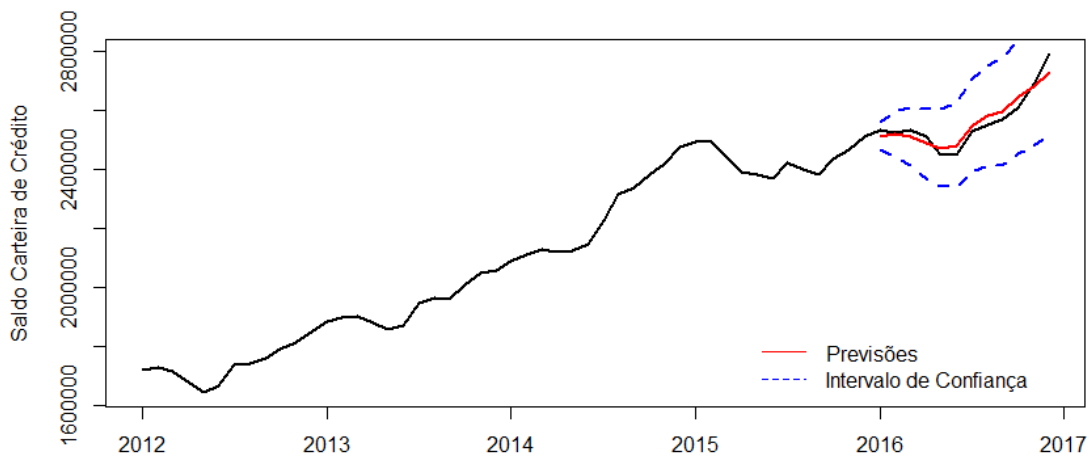


Figura 17 - Série de previsão do modelo SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IGPM (sem defasagem) e Taxa DI (defasagem 6)

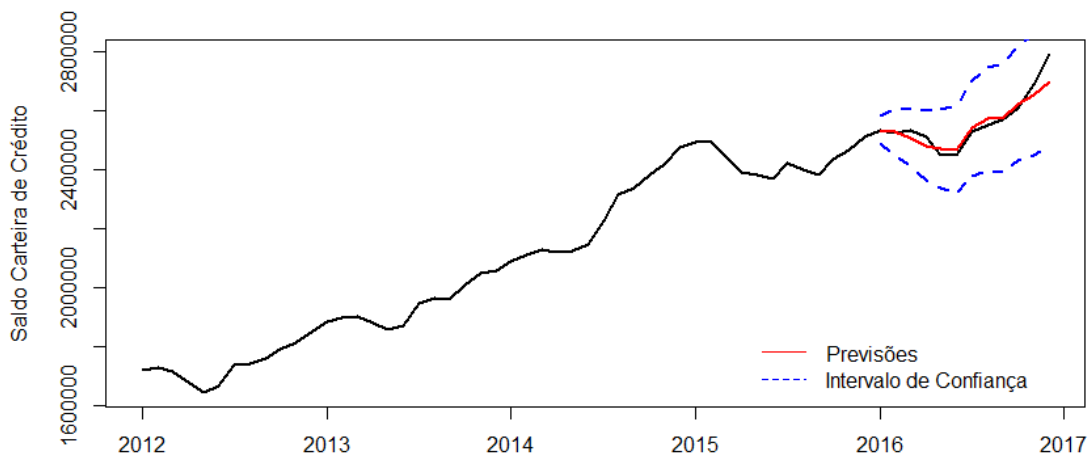


Figura 18 - Série de previsão do modelo SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IPCA e IGPM (sem defasagem)

A Tabela 3 apresenta os erros médios de previsão calculados para cada modelo, considerando previsões para um horizonte de 12 meses. Pelos erros médios observa-se supremacia dos modelos SARIMA e SARIMAX em relação ao modelo AR(1) e ao modelo atualmente praticado pela instituição.

Tabela 3 - Erros médios de previsão 12 passos à frente da série Saldo da Carteira de Crédito

Modelo	EQM	EQ-60
Modelo atual praticado	116.677	55.692
AR(1)	114.520	45.258
SARIMA(3,1,4)(1,1,0)	35.540	16.444
SARIMA(0,1,1)(0,1,1)	42.850	15.676
SARIMAX(0,1,1)(1,1,0) ₁₂ - IGPM + Taxa DI	30.043	12.799
SARIMAX(0,1,1)(1,1,0) ₁₂ - IPCA + IGPM	33.769	12.441

A Figura 19 mostra os erros absolutos de previsão dos seis modelos. Através deste comparativo, observamos que para horizontes de t+1 até t+6 os erros de previsão são semelhantes para todos os modelos. No entanto, para horizontes de t+7 até t+12 os modelos SARIMA produzem erros menores em relação ao modelo AR(1) e ao modelo atualmente praticado. Dentre todos os modelos podemos destacar o modelo SARIMAX(0,1,1)(1,1,0)₁₂ com IPCA e IGPM, que apresenta menor erro para 8 horizontes de previsão em relação aos demais.

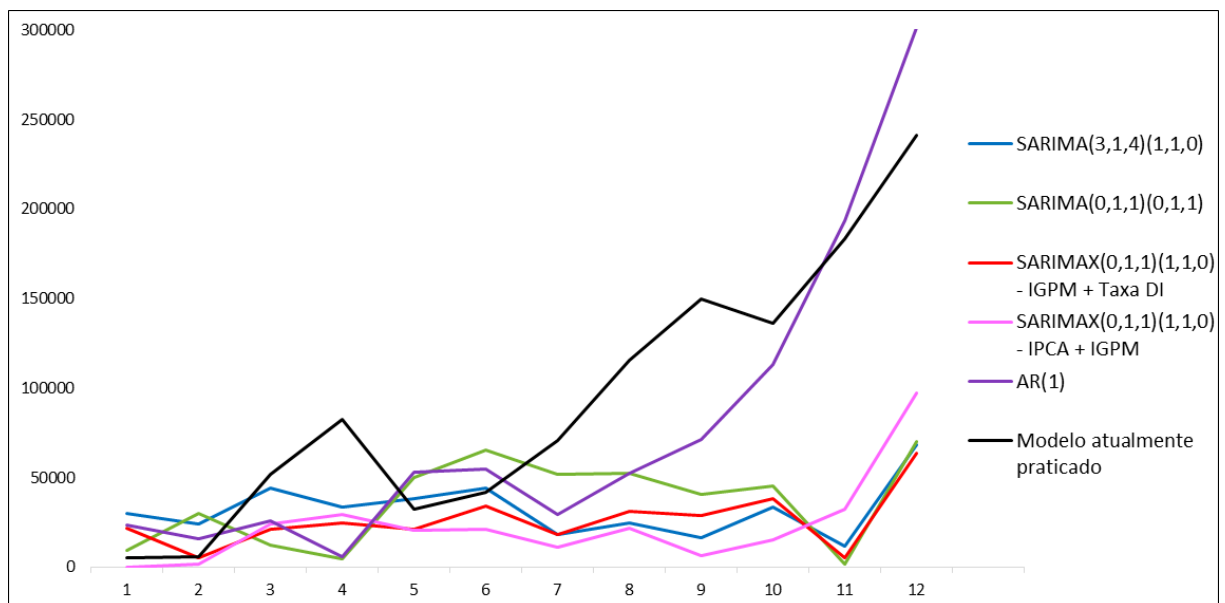


Figura 19 – Erros absolutos de previsão 12 passos à frente da série Saldo da Carteira de Crédito

Para verificar se o poder preditivo dos modelos difere significativamente, foi realizado o teste Diebold-Mariano. Para aplicação deste teste, foram calculadas 12 previsões um passo à frente para cada modelo. A hipótese nula do teste é de que os modelos comparados possuem a mesma capacidade preditiva. A Tabela 4 apresenta os p-valores do teste em que verifica-se que as previsões dos modelos SARIMA e SARIMAX não diferem significativamente do modelo univariado AR(1). Já em comparação com o modelo atualmente praticado pela instituição, há uma superioridade preditiva estatisticamente significativa do modelo SARIMA(0,1,1)(0,1,1)₁₂. O modelo SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IGPM e Taxa DI apresentam superioridade preditiva significativa quando utiliza-se a o erro quantílico como função perda para a comparação pelo teste Diebold-Mariano.

Tabela 4 - P-valores do Teste Diebold-Mariano para a série Saldo da Carteira de Crédito

Modelo	EQM		EQ-60	
	AR(1)	Modelo atualmente praticado	AR(1)	Modelo atualmente praticado
SARIMA(3,1,4)(1,1,0) ₁₂	0,2646	0,3120	0,2769	0,1549
SARIMA(0,1,1)(0,1,1) ₁₂	0,1061	0,0433	0,0686	0,0198
SARIMAX(0,1,1)(1,1,0) ₁₂ - IGPM + Taxa DI	0,3066	0,1261	0,1145	0,0472
SARIMAX(0,1,1)(1,1,0) ₁₂ - IPCA + IGPM	0,2521	0,3936	0,1355	0,0886

4. Considerações finais

O principal objetivo deste trabalho foi estimar modelos de séries temporais com o propósito de reduzir os erros de previsão em relação aos modelos atualmente praticados pela instituição financeira. A metodologia Box-Jenkins foi explorada, estimando-se modelos ARMA, ARMAX, SARIMA E SARIMAX para prever o Saldo da Carteira de Crédito Pessoal e o comportamento da Taxa de Inadimplência da mesma. Possíveis mudanças no cenário econômico foram consideradas incorporando-se variáveis macroeconômicas aos modelos. As covariáveis foram exploradas individualmente e combinadas duas a duas. Considerando-se que o efeito econômico de hoje pode afetar a Taxa de Inadimplência e a evolução da Carteira de Crédito em um período futuro, incorporaram-se todas as combinações 2-a-2 de variáveis exógenas, considerando-se de 0 a 7 defasagens das variáveis macroeconômicas.

Para a Taxa de Inadimplência, o modelo ARMAX(7,0,5) com as variáveis exógenas IPCA (defasagem 7) e TR (defasagem 4) apresentou o menor erro quadrático médio.

Considerando o erro quantílico médio, o modelo ARMAX(6,0,7) com a variável exógena IPCA (defasagem 2 e 3) apresentou o melhor desempenho. O teste Diebold-Mariano indicou que ambos os modelos possuem superioridade preditiva em relação ao modelo atualmente praticado pela instituição financeira. Porém quando comparados com o modelo AR(1) as diferenças não apresentam-se significativas. Dentre os modelos propostos para esta série, destaca-se o modelo ARMAX(6,0,7) com as covariáveis IPCA com duas e três defasagens, o qual possui uma menor variação dos erros de previsão em todos os horizontes de tempo. Além disso, este modelo produziu estimativas que subestimam os valores realizados em apenas três horizontes de tempo e com erros pequenos, o que do ponto de vista prático é relevante.

Entre os modelos propostos para prever o Saldo da Carteira de Crédito, o SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IGPM e Taxa DI (defasagens 0 e 6) foi o que apresentou o menor erro quadrático médio. O teste Diebold-Mariano indicou que este modelo possui superioridade preditiva em relação ao modelo atualmente praticado pela instituição. O modelo SARIMAX(0,1,1)(1,1,0)₁₂ com as variáveis exógenas IPCA e IGPM (sem defasagem) foi o que apresentou o menor erro quantílico médio. No entanto, as diferenças não foram significativas. Entre os modelos propostos sem variáveis exógenas, o SARIMA(0,1,1)(0,1,1)₁₂ apresentou superioridade preditiva significativa em relação ao modelo atualmente praticado. Já em comparação com o modelo AR(1) as diferenças não foram significativas para nenhum dos modelos.

Este trabalho contribuiu, do ponto de vista prático, para auxiliar na solução de uma demanda real propondo modelos que produzissem estimativas mais eficientes em relação às praticadas atualmente. Do ponto de vista acadêmico, contribuiu por ter explorado diferentes modelos de séries temporais.

Referências Bibliográficas

MARINS, J. T. M.; NEVES, M. B. E.; 2013. BCB. Banco Central do Brasil. *Inadimplência de Crédito e Ciclo Econômico: um exame da relação no mercado brasileiro de crédito corporativo*. Disponível em: <https://www.bcb.gov.br/pec/wps/port/TD304.pdf>. Acesso em 19/03/2017.

ZANIBONI, N. C.; *A Inadimplência do Sistema Financeiro no Brasil explicada por meio de fatores Macroeconômicos*. Dissertação (Mestrado em Administração), Programa de Pós-Graduação em Administração, Universidade de São Paulo, 2013.

ANTUNES, A.; RIBEIRO, N.; ANTÃO, P. *Estimating Probabilities of Default Under Macroeconomic Scenarios*. Banco de Portugal, 2005.

LINARDI, F. D. M. *Avaliação dos Determinantes Macroeconômicos da Inadimplência Bancária no Brasil*. XXXVI Encontro Nacional de Economia. Rio de Janeiro, 2008.

BUENO, V. F. F. *Avaliação de risco na concessão de crédito bancário para micros e pequenas empresas*. Dissertação (Mestrado em Engenharia da Produção), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 2003.

BB. Banco do Brasil. *Acordo de Basiléia*. Disponível em: <http://www.bb.com.br/portalbb/page51,136,3696,0,0,1,8.bb?codigoNoticia=7724>. Acesso em 28/02/2017.

BCB. Banco Central do Brasil. *Recomendações de Basiléia*. Disponível em: <http://www.bcb.gov.br/fis/supervisao/basileia.asp>. Acesso em 25/03/2017.

BASEL COMMITTEE ON BANKING SUPERVISION (2010). *Basel III: A global regulatory framework for more resilient banks and banking systems*, BIS. Disponível em: <http://www.bis.org/publ/bcbs189.pdf>. Acesso em 01/05/2017.

BCB. Banco Central do Brasil. *Conceitos e Metodologias*. Disponível em: http://www.bcb.gov.br/htms/estabilidade/2016_04/refMet.pdf. Acesso em 01/05/2017.

SEMPLE, P.A. *Previsão da Inadimplência Bancária no Brasil através dos métodos FAVAR E FAVECM*. Dissertação (Mestrado em Economia), Fundação Getúlio Vargas, Escola de Economia de São Paulo, 2013.

CIHAK, M. “*Introduction to applied stress testing*”, IMF Working Paper, 2007, WP/07/59.

JORION, P. *Value at risk: the new benchmark for controlling market risk*. Irwin Professional Pub., 1997.

DIEBOLD, F.X., MARIANO, R.S. *Comparing Predictive Accuracy*. Journal of Business and Economic Statistics, 1995, 13: 253-63.

MORETTIN, P. A.; TOLOI, C. M. C., *Análise de Séries Temporais*, 2. Ed., São Paulo: Editora Edgard Blucher, 2004.

ENDERS, WALTER, *Applied Econometric Time Series*. University of Alabama, 2015, 4. ed.

SHUMWAY. R. H.; STOFFER, D.S., *Times Series Analysis and Its Applications With R Examples*. Springer, 2011.

R DOCUMENTATION. *Seasonal Adjustment With X-13ARIMA-SEATS*. Disponível em: <https://www.rdocumentation.org/packages/seasonal/versions/1.6.1/topics/seas>. Data de acesso: 20/05/2017.

R DOCUMENTATION. *ARIMA Modelling of Time Series*. Disponível em: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/arima.html>. Data de acesso: 20/08/2016.

R DOCUMENTATION. *Forecast from ARIMA fits*. Disponível em: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/predict.arima.html>. Data de acesso: 20/08/2016.