

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO**

**MODELO BETA AUTORREGRESSIVO DE
MÉDIAS MÓVEIS: CRITÉRIOS DE SELEÇÃO
E APLICAÇÕES**

DISSERTAÇÃO DE MESTRADO

Renata Rojas Guerra

Santa Maria, RS, Brasil

2015

MODELO BETA AUTORREGRESSIVO DE MÉDIAS MÓVEIS: CRITÉRIOS DE SELEÇÃO E APLICAÇÕES

Renata Rojas Guerra

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção (PPGEP) da Universidade Federal de Santa Maria (UFSM, RS),
como requisito parcial para obtenção do título de
Mestre em Engenharia de Produção

Orientador: Prof. Dr. Fábio Mariano Bayer

Santa Maria, RS, Brasil

2015

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Rojas Guerra, Renata
Modelo beta autorregressivo de médias móveis:
critérios de seleção e aplicações / Renata Rojas Guerra.-
2015.
72 p. ; 30cm

Orientador: Fábio Mariano Bayer
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Tecnologia, Programa de Pós-Graduação em
Engenharia de Produção, RS, 2015

1. Crédito consignado 2. Critérios de seleção de
modelos 3. Inadimplência de crédito 4. Modelo BETA ARMA
5. Simulação de Monte Carlo I. Mariano Bayer, Fábio II.
Título.

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**MODELO BETA AUTORREGRESSIVO DE MÉDIAS MÓVEIS:
CRITÉRIOS DE SELEÇÃO E APLICAÇÕES**

elaborada por
Renata Rojas Guerra

como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

COMISSÃO EXAMINADORA:

Fábio Mariano Bayer, Dr.
(Presidente/Orientador)

Clailton Ataídes de Freitas, Dr. (DE-UFSM)

Augusto Maciel da Silva, Dr. (DE-UFSM)

Santa Maria, 27 de fevereiro de 2015.

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

MODELO BETA AUTORREGRESSIVO DE MÉDIAS MÓVEIS: CRITÉRIOS DE SELEÇÃO E APLICAÇÕES

AUTORA: RENATA ROJAS GUERRA

ORIENTADOR: FÁBIO MARIANO BAYER

Local da Defesa e Data: Santa Maria, 27 de fevereiro de 2015.

A modelagem e a previsão de séries temporais é um campo de ampla aplicabilidade em diversas áreas científicas e tecnológicas. No âmbito específico de variáveis restritas ao intervalo $(0, 1)$, como taxas e proporções, a utilização de modelos clássicos, que supõem normalidade da variável de interesse, pode não ser adequada. Neste contexto, Rocha e Cribari-Neto (2009) propuseram o modelo beta autorregressivo de médias móveis (β ARMA). Por assumir que a variável de interesse possui distribuição beta, que é uma distribuição mais flexível que a normal e com suporte restrito ao intervalo $(0, 1)$, o β ARMA possibilita modelagens e previsões mais condizentes com a natureza desses dados. Contudo, apenas a escolha do modelo paramétrico mais adequado não garante a acurácia do modelo ajustado. A identificação das defasagens a serem incluídas também exerce um papel de relevância neste sentido. É neste propósito que foram desenvolvidos os critérios de seleção de modelos, ou critérios de informação. Estes comparam as capacidades de explicação entre um grupo de modelos candidatos e selecionam, dentro deste grupo, o modelo que minimiza a perda de informações. Diante do exposto, este trabalho tem o objetivo de avaliar, via simulações de Monte Carlo, o desempenho de diferentes critérios de seleção no modelo β ARMA. Por meio de um extenso estudo de simulação, considerando diversos cenários e tamanhos amostrais, foram avaliados os desempenhos em amostras de tamanho finito dos critérios AIC, BIC, HQ, AICc, BICc e HQc. Como resultados numéricos gerais, destaca-se que os critérios HQ, BICc e HQc foram os que alcançaram os melhores níveis de identificação do modelo verdadeiro. Utilizando os critérios de seleção sugeridos no estudo de simulação também foram ajustados modelos β ARMA a dados reais. Para isso, foram considerados o índice de inadimplência de crédito e a relação entre o crédito consignado e o crédito total pessoa física, ambos do Sistema Financeiro Nacional. Também foram ajustados os clássicos modelos ARIMA comparativamente ao modelo β ARMA na realização de previsões e posterior comparação entre os resultados de ambas as aplicações. Para as duas variáveis há um grau razoável de proximidade entre os dados originais e previstos, com superioridade do β ARMA tanto dentro quanto fora do conjunto de observações utilizado para estimação dos modelos.

Palavras-chave: Crédito consignado. Critérios de seleção de modelos. Inadimplência de crédito. Modelo β ARMA. Simulação de Monte Carlo.

ABSTRACT

Master's Dissertation
Post-Graduate Program in production engineering
Federal University of Santa Maria

DISSERTATION TITLE

AUTHOR: RENATA ROJAS GUERRA

ADVISOR: FÁBIO MARIANO BAYER

Defense Place and Date: Santa Maria, February 27st, 2015.

Time series modeling and forecasting has many applicability in scientific and technological researches. Specifically about variables restricted to the interval $(0, 1)$, which includes rates and proportions, the classical regression models could not be suitable because they assume normality. In this context, Rocha and Cribari-Neto (2009) proposed the beta autoregressive moving average (β ARMA) model. It admits that the variable of interest is beta distributed. The beta distribution is more flexible than the normal distribution and also assumes that the dependent variable is restricted to the interval $(0, 1)$. Through β ARMA is possible to obtain results closer to the nature of the data. But just choose the better parametric model does not guarantee the accuracy of the fitted model. To identify the lags is also relevant to ensure the accuracy of the adjusted model. It is in this purpose that the model selection criteria, or information criteria, were developed. They compare the explanatory capacity of a group of models and select, among this group, the model which minimizes the information loss. In this context, this paper aims to evaluate by Monte Carlo simulations the performance of different selection criteria in β ARMA model. Considering several scenarios and sample sizes, the selection criteria evaluated was AIC, BIC, HQ, AICc, BICc and HQc. The results indicate that BICc, HQ and HQc had the better performance identifying the true model among the candidate models. Using the selection criteria indicated by the simulation study, were also adjusted β ARMA models to real data. It were considered the credit delinquency and the relationship between payroll loan and individual credit, both variables are from national financial system. It was adjusted the classical ARIMA models too. This models were compared with β ARMA in applications. For both variables was found a reasonable proximity between the original data and the predicted by the models, with advantage for β ARMA, as much inside as outside the sample.

Keywords: β ARMA model. Credit delinquency. Model selection criteria. Monte Carlo Simulation. Payroll loan.

LISTA DE FIGURAS

Figura 2.1 – Densidades beta para diferentes valores de μ (indicados no gráfico) e ϕ	18
Figura 3.1 – Frequência (em %) de identificação dos modelos com precisão $\phi = 20$	26
Figura 3.2 – Frequência (em %) de identificação dos modelos com precisão $\phi = 120$	27
Figura 4.1 – Evolução da participação crédito/PIB no Brasil	31
Figura 4.2 – Índice de inadimplência do SFN no período de março de 2011 a dezembro de 2013	34
Figura 4.3 – Histograma e densidades estimadas do índice de inadimplência do SFN	35
Figura 4.4 – Correlogramas do índice de inadimplência em nível e em primeira diferença	36
Figura 4.5 – Série de inadimplência de crédito em primeira diferença	37
Figura 4.6 – Valores originais e previstos dentro e fora da amostra para a série de inadimplência de crédito do SFN	41
Figura 4.7 – Resíduos padronizados dos modelos para a inadimplência de crédito	41
Figura 4.8 – Correlogramas dos resíduos padronizados dos modelos ARIMA e β ARMA da inadimplência de crédito	42
Figura 4.9 – Crédito consignado como proporção do crédito total PF no período de março de 2007 a agosto de 2014	44
Figura 4.10 – Histograma e densidades estimadas do índice de CC/Crédito PF	45
Figura 4.11 – Correlogramas do CC/Crédito PF em nível e em primeira diferença	46
Figura 4.12 – Série de CC/Crédito PF em primeira diferença	46
Figura 4.13 – Valores originais e previstos dentro e fora da amostra para a relação de CC/Crédito PF	49
Figura 4.14 – Resíduos padronizados dos modelos para o CC/Crédito PF	50
Figura 4.15 – Correlogramas dos resíduos padronizados dos modelos ARIMA e β ARMA do CC/Crédito PF	51

LISTA DE TABELAS

Tabela 3.1 – Descrição dos cenários de acordo com as classes de modelos candidatos e o modelo verdadeiro	24
Tabela 3.2 – Frequência (em %) de identificação no Cenário 3 com precisão $\phi = 20$	29
Tabela 3.3 – Frequência (em %) de identificação no Cenário 3 com precisão $\phi = 120$	30
Tabela 4.1 – Medidas descritivas para a inadimplência de crédito.....	33
Tabela 4.2 – Critérios para identificação do modelo ARIMA para a inadimplência de crédito.....	38
Tabela 4.3 – Critérios para identificação do modelo β ARMA para a inadimplência de crédito.....	39
Tabela 4.4 – Estimativas do modelo ARIMA selecionado para a inadimplência de crédito	39
Tabela 4.5 – Estimativas do modelo β ARMA selecionado para a inadimplência de crédito	39
Tabela 4.6 – Previsões para o índice de inadimplência de crédito para meses fora da amostra.....	40
Tabela 4.7 – Medidas descritivas para o crédito consignado como proporção do crédito total PF.....	45
Tabela 4.8 – Critérios para identificação do modelo ARIMA para índice de CC/Crédito PF	47
Tabela 4.9 – Critérios para identificação do modelo β ARMA para índice de CC/Crédito PF.....	48
Tabela 4.10 – Estimativas do modelo ARIMA selecionado para índice de CC/Crédito PF ..	48
Tabela 4.11 – Estimativas do modelo β ARMA selecionado para índice de CC/Crédito PF .	48
Tabela 4.12 – Previsões para a relação CC/Crédito PF para meses fora da amostra.....	50
Tabela A.1 – Frequência (em %) de identificação no Cenário 1 com $\phi = 20$	67
Tabela A.2 – Frequência (em %) de identificação no Cenário 2 com $\phi = 20$	67
Tabela A.3 – Frequência (em %) de identificação no Cenário 4 com $\phi = 20$	68
Tabela A.4 – Frequência (em %) de identificação no Cenário 5 com $\phi = 20$	69
Tabela A.5 – Frequência (em %) de identificação no Cenário 1 com $\phi = 120$	70
Tabela A.6 – Frequência (em %) de identificação no Cenário 2 com $\phi = 120$	70
Tabela A.7 – Frequência (em %) de identificação no Cenário 4 com $\phi = 120$	71
Tabela A.8 – Frequência (em %) de identificação no Cenário 5 com $\phi = 120$	72

LISTA DE APÊNDICES

APÊNDICE A – Tabelas de frequência (em %) de identificação dos modelos	66
---	-----------

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Justificativa	13
1.2 Objetivos	14
1.2.1 Objetivo geral	14
1.2.2 Objetivos específicos	14
2 REFERENCIAL TEÓRICO	15
2.1 O modelo ARIMA	15
2.2 O modelo βARMA	17
2.3 Critérios de seleção	21
3 AVALIAÇÃO NUMÉRICA DE CRITÉRIOS DE SELEÇÃO	24
4 APLICAÇÕES A DADOS REAIS	31
4.1 Modelando o índice de inadimplência	32
4.1.1 Modelagem e discussão	37
4.2 Modelando o crédito consignado	43
4.2.1 Modelagem e discussão	47
5 CONCLUSÃO	52
REFERÊNCIAS	54
APÊNDICES	65

1 INTRODUÇÃO

Modelos estatísticos que supõem normalidade dos dados, como o modelo de regressão linear (GUJARATI; PORTER, 2011) e os modelos autorregressivos, integrados e de médias móveis (ARIMA) (BOX; JENKINS, 1976), são amplamente utilizados em diversas áreas científicas e tecnológicas, especialmente em aplicações empíricas no campo da engenharia de produção (EDIGER; AKAR, 2007; BABAI et al., 2013; RAMOS; SANTOS; REBELO, 2015). Dentre esta variada gama de aplicações, a taxa de desemprego (VICENTE; LOPEZ-MENENDEZ; PEREZ, 2015), utilização de capacidade produtiva (CASARIN; DALLA VALLE; LEINSEN, 2012), proporção de participação de um produto no mercado e índice de inadimplência (LOUZIS; VOULDIS; METAXAS, 2012) são destacáveis como exemplos de variáveis duplamente limitadas comumente modeladas. Entretanto, quando a variável de interesse está contida no intervalo $(0, 1)$, como nestes últimos casos, a utilização de modelos clássicos que supõem normalidade muitas vezes torna-se inapropriada, uma vez que este pode prever valores da variável resposta fora do intervalo unitário padrão (CRIBARI-NETO; ZEILEIS, 2010).

Uma das soluções mais utilizadas para corrigir este problema é a transformação da variável resposta. Todavia, este procedimento dificulta a interpretação dos parâmetros estimados em relação à variável original e não considera as possíveis assimetrias dos dados (BAYER, 2011). Assim, no intuito de dirimir os prejuízos acarretados por esta problemática, Ferrari e Cribari-Neto (2004) propuseram o modelo de regressão beta. Este considera que a variável resposta de interesse possui distribuição beta e, com isso, permite uma modelagem mais adequada a dados do tipo taxas e proporções que são restritas a intervalos duplamente limitados. Seguindo esta linha, Rocha e Cribari-Neto (2009) introduziram o modelo beta autorregressivo de médias móveis (β ARMA). Este consiste em uma extensão do modelo de regressão beta para previsões de séries temporais e tem forma semelhante aos modelos ARIMA (BOX; JENKINS, 1976). Porém, em contraponto com o ARIMA, que supõe normalidade dos dados, o β ARMA assume que a variável de interesse possui distribuição beta, sendo útil para modelagem e previsão de séries temporais que assumem valores no intervalo $(0, 1)$.

Neste contexto, há um crescente interesse no estudo de extensões, melhoramentos e aplicações desses modelos que assumem distribuição beta para a variável de interesse. Cribari-Neto e Zeileis (2010) apresentam a modelagem de regressão beta no sistema R (R Core Team, 2013). Ainda para o modelo de regressão beta, há trabalhos que abordam melhoramentos inferenci-

ais (OSPINA; CRIBARI-NETO; VASCONCELOS KLAUS, 2006; BAYER; CRIBARI-NETO, 2013), modelagens e generalizações (SIMAS; SOUZA; ROCHA, 2010; OSPINA; FERRARI, 2011; PEREIRA, 2010), técnicas de diagnóstico (ESPINHEIRA; FERRARI; CRIBARI-NETO, 2008a,b) e aplicações (SILVA; SOUZA, 2014; ALMEIDA; SOUZA, 2015). No âmbito de séries temporais, Amisano e Casarin (2007), Billio e Casarin (2011) e Casarin, Dalla Vale e Leinsen (2012), utilizam abordagens bayesianas para realizar inferências em processos beta autoregressivos (β AR) de primeira ordem. Guolo e Varin (2014) também contribuem para o tema ao modelar a incidência do vírus da gripe no Canadá por meio de uma cópula gaussiana com matriz de correlação correspondente a processos autorregressivos e de médias móveis com distribuição beta. Essas são algumas das principais inovações em termos de modelos de previsão em séries temporais que assumem distribuição beta. Entretanto, há um campo considerável de estudos a serem desenvolvidos em termos do β ARMA, principalmente no âmbito de pesquisas metodológicas.

Dentre os tópicos em destaque, cabe ressaltar a proposição, a utilização e a avaliação de critérios de seleção de modelos. Torres (2005), Bayer (2011) e Bayer e Cribari-Neto (2015a; 2015b), que realizam tal análise no contexto de regressão beta, ressaltam que o processo de escolha do conjunto de covariáveis úteis a partir de um conjunto mais amplo de regressores é parte reconhecidamente fundamental do processo de modelagem e constitui uma constante preocupação dos pesquisadores. Desta forma, este é um tema recorrente também em outras classes de modelos (HANNAN; QUINN, 1979; PAN, 2001; KLETTING; GLATTING, 2009; MARHUENDA; MORALES; PARDO, 2014) e em pequenas amostras (ISHIGURO; SAKAMOTO; KITAGAWA, 1997; BENGTTSSON; CAVANAUGH, 2006; SEGHOUANE, 2010).

O importante papel dos critérios de seleção de modelos é ainda mais evidente ao se tratar de modelos compostos por termos autorregressivos e de médias móveis. Conforme Gujarati e Porter (2011), estes tipos de modelos são por vezes denominados ateoréticos pois não são derivados de nenhuma teoria econômica. Estes utilizam procedimentos predominantemente estatísticos, como os critérios de seleção, para determinar a ordem de defasagens dos termos autorregressivos e de médias móveis do modelo.

Diante do exposto, insere-se o seguinte questionamento: como selecionar a ordem de defasagens que propiciará o melhor balanceamento entre a complexidade e capacidade explicativa nos modelos β ARMA? Para responder a essa pergunta serão avaliados, por meio de um estudo de simulação de Monte Carlo, critérios de seleção de modelos originalmente desenvolvidos para

regressão linear, modelos autorregressivos e modelos lineares generalizados (MLG) (MCCULLAGH; NELDER, 1989), os quais podem ser diretamente estendidos aos modelos β ARMA. Além disso, também serão apresentadas aplicações do β ARMA e dos critérios de seleção em séries temporais relacionadas ao volume de crédito do Sistema Financeiro Nacional (SFN).

No intuito de realizar previsões mais acuradas, serão considerados comparativamente os modelos β ARMA e ARIMA para a previsão do comportamento destas séries temporais reais. Com isso, torna-se possível verificar se as projeções obtidas através do β ARMA são mais acuradas que as obtidas por meio de modelos ARIMA. As variáveis utilizadas nas aplicações são o índice de inadimplência do total de crédito disponível e a relação entre o crédito consignado e o crédito total pessoa física, ambos para o SFN. A escolha destas séries temporais é embasada em seus importantes papéis na dinâmica de crédito do país e, por conseguinte, no desenvolvimento da economia brasileira (ALMEIDA, 2013).

1.1 Justificativa

A construção de projeções eficazes para o comportamento de séries temporais é de suma importância não só na engenharia de produção, mas também na economia, ciências sociais, climatologia e diversas outras áreas do conhecimento. Com isso, a estimação de modelos que não representam satisfatoriamente o comportamento da variável poderá incorrer em diversos prejuízos às instituições que utilizarem tais informações em seu processo de tomada de decisão.

No âmbito específico de séries temporais restritas ao intervalo $(0, 1)$, a estimação de modelos ARIMA torna-se inadequada uma vez que supõe que a variável de interesse é normalmente distribuída. Cribari-Neto e Zeileis (2010) argumentam que, nesses casos, modelos estatísticos que assumem distribuição normal podem prever valores fora do intervalo de possíveis ocorrências da variável. Ademais, ao supor normalidade o pesquisador não estará considerando o possível comportamento assimétrico dos dados, característica comum em variáveis do tipo taxas e proporções (SOUZA; CRIBARI-NETO, 2013). Já a utilização do β ARMA possibilita a realização de previsões mais condizentes com a natureza dos dados, pois a distribuição beta possui essas características, ou seja, tem formato de caráter variado e suporte restrito ao intervalo unitário padrão.

Entretanto, apenas escolha do modelo paramétrico mais adequado, neste caso via modelos β ARMA, não garante a capacidade preditiva do modelo estimado. O processo de identificação das defasagens a serem incluídas nos termos autorregressivos (AR) e de médias móveis

(MA) também desempenha papel particularmente importante neste sentido. Isto ocorre principalmente porque os modelos que possuem mais parâmetros nem sempre são os que melhor se adaptam aos dados. A inclusão excessiva de parâmetros pode apenas tornar a estimação mais custosa e a análise mais complexa, devendo ser buscado o equilíbrio entre a melhora no ajuste e a complexidade do modelo (EMILIANO, 2009).

Conforme Giampaoli, Raposo e Torres (2009), os critérios de seleção foram criados para auxiliar no processo de decisão do modelo mais apropriado dentre um conjunto de modelos candidatos. Eles comparam as respectivas capacidades de explicação de cada modelo candidato e selecionam aquele que minimiza a perda de informações. Entretanto, um critério de seleção nem sempre é melhor que o outro (MCQUARRIE; TSAI, 1998). Eles podem obter comportamentos diferentes de acordo com as especificidades do modelo utilizado (EMILIANO et al., 2009). Com isso, é necessário avaliar o desempenho dos diversos critérios para cada classe de modelos, bem como identificar aquele que é mais adequado à natureza dos dados.

1.2 Objetivos

1.2.1 Objetivo geral

Avaliar o desempenho de diferentes critérios de seleção no modelo β ARMA e realizar aplicações a dados reais.

1.2.2 Objetivos específicos

- Destacar os critérios de seleção mais usuais na literatura científica;
- Utilizar simulações de Monte Carlo para verificar quais critérios são melhores para a seleção de modelos β ARMA em amostras de tamanho finito;
- Apresentar aplicações do β ARMA, e dos critérios de seleção avaliados, para modelagem do índice de inadimplência e da relação entre o crédito consignado e o crédito total pessoa física do SFN;
- Comparar o desempenho dos modelos ARIMA e β ARMA nas aplicações realizadas.

2 REFERENCIAL TEÓRICO

Nesta seção são apresentados os modelos ARIMA e β ARMA. Ambos são utilizados para identificação, estimação e previsão de séries temporais e têm suas estruturas de regressão compostas em termos autorregressivos e de médias móveis. O ARIMA é útil para variáveis que assumem distribuição normal. Já o β ARMA assume distribuição beta para a variável de interesse, sendo uma generalização do modelo de regressão beta (FERRARI; CRIBARI-NETO, 2004) especificamente para a modelagem de variáveis do tipo taxas e proporções ao longo do tempo.

2.1 O modelo ARIMA

Por fornecer informações relevantes na elaboração de estratégias de gestão e planejamento, uma das principais características consideradas ao modelar séries temporais consiste na capacidade preditiva do modelo a ser estimado. Foi nesse propósito que Box e Jenkins (1976) propuseram a classe de modelos ARIMA. Este parte do pressuposto que a análise probabilística das propriedades da série temporal fornece ferramentas suficientes para a previsão de seu próprio comportamento.

O ARIMA assume que a variável de interesse y possui distribuição normal, com densidade definida por:

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}, -\infty < \mu < \infty, \sigma^2 > 0, \quad (2.1)$$

a qual é destacável por seu formato simétrico, com achatamento e posicionamento relacionados aos valores dos parâmetros μ e σ^2 e gráfico em forma de sino. Neste caso, temos que

$$E(y) = \mu$$

e

$$Var(y) = \sigma^2.$$

Além da suposição de normalidade, um aspecto frequentemente mencionado em análises de previsão diz respeito à estacionaridade das séries. Se uma série temporal apresentar média e variância constantes ao longo do tempo e se sua covariância entre dois períodos depender apenas da defasagem entre os mesmos, e não da época em que for calculada, então esta

atende os requisitos da chamada estacionaridade fraca (GRANGER; NEWBOLD, 1986). A importância da verificação desta característica justifica-se pelo fato de que se uma série é não estacionária os choques aleatórios que ocorrerem tendem a se perpetuar, tornando impossível estimar todos os períodos da série e fazendo com que as inferências obtidas a partir destas informações não sejam válidas (POKORNY, 1987).

Quando a série temporal é estacionária a estrutura de regressão dos modelos de previsão pode ocorrer através de apenas termos AR, apenas termos MA, ou ainda através de modelos com termos autoregressivos e de médias móveis, chamados modelos ARMA (WHEELWRIGHT; MAKRIDAKIS, 1985). Assim, seja y_t , $t = 1, \dots, n$, um conjunto de n ocorrências da variável de interesse y gerada por um processo ARMA estacionário, em que cada y_t possui densidade definida em (2.1), com média μ_t , variância constante σ^2 , a especificação do modelo ARMA é dada por:

$$\mu_t = \alpha + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j r_{t-j}. \quad (2.2)$$

Desta forma o valor esperado da variável dependente, $E(y_t) = \mu_t$, é explicado pelos seus p valores passados y_{t-i} (termo $AR(p)$) e pelas q médias móveis r_{t-j} (termo $MA(q)$), em que

$$r_t = y_t - \mu_t, \quad (2.3)$$

e $r_t \sim N(0, \sigma^2)$. O termo α corresponde ao intercepto, enquanto que $\varphi = (\varphi_1, \dots, \varphi_p)$ e $\theta = (\theta_1, \dots, \theta_q)$ são os vetores que representam os parâmetros dos termos AR e MA, respectivamente.

O método usualmente utilizado para estimação do vetor paramétrico $\xi = (\alpha, \varphi, \theta, \sigma^2)$ é o de máxima verossimilhança. Assim, partindo da equação (2.3), temos que o processo $ARMA(p, q)$ pode ser escrito como:

$$r_t = y_t - \left(\alpha + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \theta_j r_{t-j} \right).$$

Neste caso, a função de verossimilhança é dada por:

$$L(y; \xi) = (2\pi)^{-n/2} (\sigma)^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - \mu_t)^2 \right\}, \quad (2.4)$$

em que μ_t em (2.4) é dado pela estrutura definida em (2.2). A log-verossimilhança condicional à escolha de y_t e r_t é escrita da seguinte forma:

$$\ell(y; \xi) = \sum_{t=1}^n \ell_t(\mu_t, \sigma^2), \quad (2.5)$$

em que,

$$\ell_t(\mu; \sigma^2) = -n \log \sigma - \frac{S_*(\alpha, \varphi, \theta)}{2\sigma^2},$$

onde $S_*(\alpha, \varphi, \theta)$ é a chamada soma de quadrados condicional (MORETTIN; TOLOI, 2006). Os estimadores de máxima verossimilhança (EMV's) do vetor ξ serão os respectivos valores que maximizam a log-verossimilhança dada em (2.5). Detalhes sobre função escore, matriz de informação de Fisher e demais procedimentos inferenciais nos modelos ARIMA podem ser encontrados em Box e Jenkins (1976), Box, Jenkins e Reinsel (2008) e Morettin e ToloI (2006).

Se a série temporal for gerada por um processo AR, então o vetor de parâmetros θ será igual a zero. Analogamente, se a série temporal for gerada por um processo MA, então o vetor de parâmetros φ será igual a zero. Caso a série de dados analisada seja não estacionária, uma das alternativas para solucionar o problema é a sua diferenciação tantas vezes quanto necessário para estacionarizá-la (MORETTIN; TOLOI, 2006). Quando é preciso diferenciar uma série d vezes para torná-la estacionária, diz-se que esta é integrada de ordem d , ou $I(d)$. Foi atentando para este fato que Box e Jenkins (1976) incluíram o termo I na definição dos modelos ARMA. Com isso, se um modelo ARMA com p termos autorregressivos e q termos de médias móveis é baseado em uma série $I(d)$, então este será um ARIMA(p, d, q).

Também podem ser encontradas outras variações e generalizações desse modelo: os modelos SARIMA acrescentam um componente sazonal (BOX; JENKINS; REINSEL, 2008); os modelos ARFIMA consideram que a ordem de integração do modelo não necessariamente corresponde a períodos inteiros (HOSKING, 1981); os GARMA são modelos dinâmicos úteis para modelar variáveis com diferentes distribuições da família exponencial e baseados nos MLG (BENJAMIN; RIGBY; STASINOPOULOS, 2003); por fim, o modelo β ARMA representa os casos em que a variável resposta está restrita ao intervalo $(0, 1)$.

2.2 O modelo β ARMA

Foi com o objetivo de definir um modelo dinâmico para séries temporais do tipo taxa e/ou proporção que Rocha e Cribari-Neto (2009) desenvolveram o modelo β ARMA. O modelo assume que a variável de interesse y possui distribuição beta com densidade definida por:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{(\mu\phi-1)}(1-y)^{(1-\mu)\phi-1}, 0 < y < 1, \quad (2.6)$$

em que

$$E(y) = \mu$$

e

$$Var(y) = \frac{V(\mu)}{1 + \phi},$$

sabendo que $V(\mu) = \mu(1 + \mu)$ é a função de variância, onde μ é o parâmetro que representa a média e ϕ pode ser interpretado como um parâmetro de precisão. Quanto maior for o valor de ϕ , menor será a variância de y fixando-se μ . Esta é uma reparametrização da densidade beta original, proposta por Ferrari e Cribari-Neto (2004) na definição do modelo de regressão beta.

Em contraponto com a curva normal, a distribuição beta tem a vantagem de ser mais flexível, podendo assumir formato de caráter variado, de acordo com os valores dos dois parâmetros que indexam a função densidade. Com isso, a distribuição beta pode assumir formas simétricas, assimétricas à esquerda ou assimétricas à direita, entre outras (SOUZA; CRIBARI-NETO, 2013). A Figura 2.1 ilustra o comportamento da curva da densidade beta conforme se alteram os valores dos parâmetros ϕ e μ .

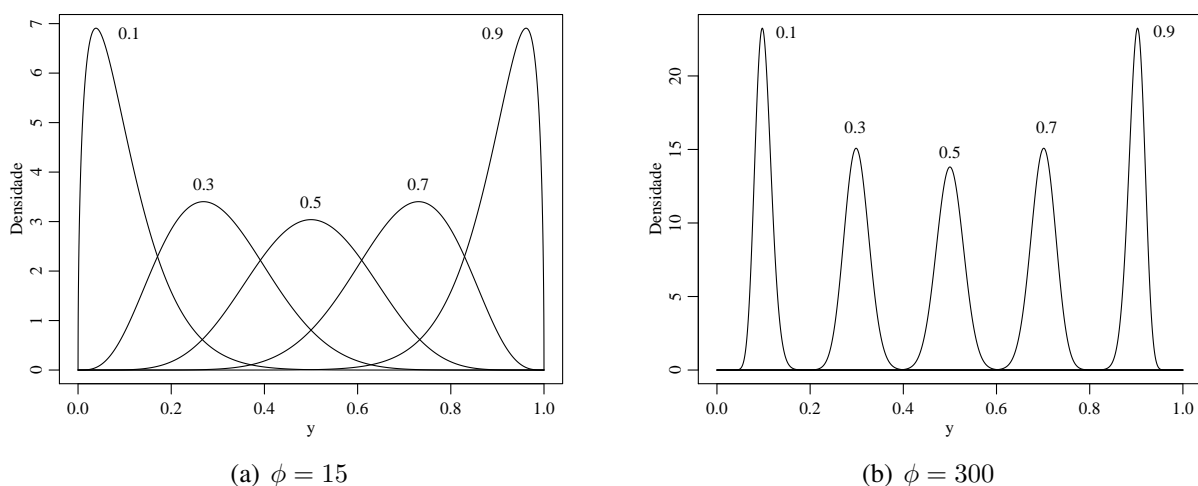


Figura 2.1 – Densidades beta para diferentes valores de μ (indicados no gráfico) e ϕ

O suporte da distribuição beta é o intervalo unitário contínuo $(0, 1)$ e o β ARMA pressupõe estacionaridade dos dados (ROCHA; CRIBARI-NETO, 2009). Entretanto, se a série temporal apresentar padrões típicos de ausência de estacionaridade, a aplicação do modelo ainda é viável mediante a diferenciação da série. Quando uma série temporal é diferenciada ocorre uma alteração no intervalo de restrição passando, por exemplo, para $(-1, 1)$ na primeira diferença, $(-2, 2)$ na segunda diferença e $(-d, d)$ na d -ésima diferença. Conforme salientam Rocha e Cribari-Neto (2009), o β ARMA também é válido se a variável resposta é restrita a um intervalo duplamente limitado em (a, b) , onde a e b são escalares e $a < b$. Nestes casos, pode-se aplicar a transformação $(y - a)/(b - a)$ para modelar y_t dentro do intervalo unitário padrão.

Dessa forma, sendo y_t , $t = 1, \dots, n$, um conjunto de n ocorrências da variável de interesse y , em que cada y_t possui densidade dada por (2.6), com média μ_t e precisão constante ϕ , o modelo β ARMA é definido por:

$$g(\mu_t) = \alpha + \sum_{i=1}^p \varphi_i g(y_{t-i}) + \sum_{j=1}^q \theta_j r_{t-j}, \quad (2.7)$$

em que o termo α é o intercepto, p e q são as ordens autorregressiva e de médias móveis do modelo, cujos vetores paramétricos são φ e θ , respectivamente. Além disso, r_{t-j} representa os vetores do termo de erro na $t - j$ defasagem e $g(\cdot)$ representa uma função de ligação monótona e duas vezes diferenciável. Funções de ligação usuais para dados no intervalo $(0, 1)$ são as funções logit e probit. Para maiores informações sobre funções de ligação ver McCullagh e Nelder (1989).

Para estimação do vetor paramétrico $\gamma = (\alpha, \phi, \varphi^\top, \theta^\top)^\top$, onde $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p)^\top$ e $\theta = (\theta_1, \theta_2, \dots, \theta_q)^\top$, é utilizado o método de máxima verossimilhança. Assim, partindo da Equação (2.7), tem-se que a função de log-verossimilhança condicional às primeiras m observações, onde $m = \max\{p, q\}$, do vetor de parâmetros γ pode ser escrita como:

$$\ell(y; \gamma) = \sum_{t=1}^n \ell_t(\mu_t, \phi),$$

em que

$$\begin{aligned} \ell_t(\mu, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1 - \mu_t) \phi) \\ &+ (\mu_t \phi - 1) \log y_t + \{(1 - \mu_t) \phi - 1\} \log(1 - y_t). \end{aligned}$$

Já o vetor escore $U(\gamma)$ é construído a partir das derivadas parciais da log-verossimilhança condicional em relação a cada um dos parâmetros desconhecidos α , ϕ , φ e θ . Assim, temos:

$$U_\alpha(\gamma) = \phi \mathbf{1}^\top T(y^* - \mu^*),$$

$$U_\phi(\gamma) = \sum_{t=m+1}^n \{\mu(y^* - \mu^*) + \log(1 - y_t) - \psi((1 - \mu_t) \phi) + \psi(\phi)\},$$

$$U_\varphi(\gamma) = \phi P^\top T(y^* - \mu^*),$$

e

$$U_\theta(\gamma) = \phi R^\top T(y^* - \mu^*),$$

em que $y_t^* = \log y_t / (1 - y_t)$ e $\mu_t^* = \psi(\mu_t \phi) - \psi((1 - \mu_t)\phi)$, $y^* = (y_{m+1}^*, \dots, y_n^*)^\top$, $\mu^* = (\mu_{m+1}^*, \dots, \mu_n^*)^\top$, $\mathbf{1}$ é um vetor $(n-m) \times 1$ com números um e $T = \text{diag}(1/g(\mu_{m+1}), \dots, 1/g(\mu_n))$, P é uma matriz $(n-m) \times p$ cujo elemento (i, j) é dado por $g(y_{i+m-j})$, $\psi(\cdot)$ é a função dígamma e R é uma matriz $(n-m) \times q$, em que o elemento da i -ésima linha e j -ésima coluna é dado por r_{i+m-j} .

A obtenção dos EMV's ocorre mediante a resolução do seguinte sistema:

$$\begin{cases} U_\alpha = 0 \\ U_\phi = 0 \\ U_\varphi = 0 \\ U_\theta = 0 \end{cases}.$$

Como a solução desse sistema não possui forma fechada, é necessária a utilização de um algoritmo de otimização não-linear para a determinação dos EMV's. Neste trabalho o algoritmo quasi-Newton BFGS (PRESS et al., 2007; NOCEDAL; WRIGHT, 1999; MITTELHAMMER; JUDGE; MILLER, 2000) é utilizado para obtenção dos EMV's.

Para inferências intervalares e testes de hipóteses em grandes amostras é necessário o conhecimento da matriz de informação de Fisher. A matriz de informação de Fisher conjunta pode ser expressa matricialmente por:

$$K = K(\gamma) = \begin{bmatrix} K_{\alpha\alpha} & K_{\alpha\phi} & K_{\alpha\varphi} & K_{\alpha\theta} \\ K_{\phi\alpha} & K_{\phi\phi} & K_{\phi\varphi} & K_{\phi\theta} \\ K_{\varphi\alpha} & K_{\varphi\phi} & K_{\varphi\varphi} & K_{\varphi\theta} \\ K_{\theta\alpha} & K_{\theta\phi} & K_{\theta\varphi} & K_{\theta\theta} \end{bmatrix},$$

em que $K_{\alpha\alpha} = \phi \text{tr}(W)$, $K_{\alpha\phi} = K_{\phi\alpha}^\top = \mathbf{1}^\top T c$, $K_{\alpha\varphi} = K_{\varphi\alpha}^\top = \phi P^\top W \mathbf{1}$, $K_{\alpha\theta} = K_{\theta\alpha}^\top = \phi R^\top W \mathbf{1}$, $K_{\phi\phi} = \text{tr}(D)$, $K_{\phi\varphi} = K_{\varphi\phi}^\top = P^\top T c$, $K_{\phi\theta} = K_{\theta\phi}^\top = R^\top T c$, $K_{\varphi\varphi} = \phi P^\top W P$, $K_{\varphi\theta} = K_{\theta\varphi}^\top = \phi R^\top W P$, $K_{\theta\theta} = \phi R^\top W R$. O termo W é representado por $W = \text{diag}\{w_{m+1}, \dots, w_n\}$, com

$$w_t = \phi \frac{\{\psi(\mu_t \phi) + \psi^\top((1 - \mu_t)\phi)\}}{g^\top(\mu_t)^2},$$

$c = \{c_{m+1}, \dots, c_n\}$, em que $c_t = \phi \{\psi^\top(\mu_t \phi) \mu_t - \psi^\top((1 - \mu_t)\phi)(1 - \mu_t)\}$, e $D = \text{diag}\{d_{m+1}, \dots, d_n\}$, com $d_t = \psi^\top(\mu_t \phi) \mu_t^2 - \psi^\top((1 - \mu_t)\phi)(1 - \mu_t)^2 - \psi^\top(\phi)$. Maiores detalhes e derivadas parciais da função log-verossimilhança condicional podem ser vistos em Rocha e Cribari-Neto (2009).

Desta forma, sob certas condições de regularidade, e para tamanhos amostrais grandes, a distribuição conjunta dos EMV's do vetor paramétrico γ é aproximadamente normal com dimensão $k = p + q + 2$, de modo que

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\phi} \\ \hat{\varphi} \\ \hat{\theta} \end{pmatrix} \sim \mathcal{N}_k \left(\begin{pmatrix} \alpha \\ \phi \\ \varphi \\ \theta \end{pmatrix}, K^{-1} \right),$$

sendo $\hat{\alpha}$, $\hat{\phi}$, $\hat{\varphi}$ e $\hat{\theta}$ os EMV's de seus respectivos parâmetros.

Para análise de diagnóstico do modelo ajustado é importante a consideração de um bom resíduo. Os resíduos padronizados são dados por:

$$r_t^p = \frac{y_t - \mu_t}{\sqrt{\text{var}(y_t)}} = \frac{y_t - \mu_t}{\sqrt{\mu_t(1 - \mu_t)/1 + \phi}}. \quad (2.8)$$

Quando o modelo está corretamente ajustado, os resíduos devem estar próximos da distribuição normal padrão, ou seja, espera-se que os resíduos estejam no intervalo $(-3,3)$, com média zero e variância unitária constante.

Para previsões passos a frente, as estimativas $\hat{\gamma}$ de γ são utilizadas para estimar μ_t , para $t > n$. Sabendo que, para $N > n$, as previsões dos erros r_N são iguais a zero, para prever o valor da média do processo, em que $T > n$, deve-se utilizar $\hat{\gamma}$, $\hat{\mu}_t$ e \hat{r}_t , substituindo r_t por zero se $t > n$ (isto é suficiente para obter $\hat{\mu}_{n+1}$ e, analogamente, $\hat{\mu}_{n+2}$, $\hat{\mu}_{n+3}$, e assim sucessivamente) e substituindo y_t por $\hat{\mu}_t$ se $n < t < T$. Por exemplo, a estimativa da média para $n + 1$ é dada por:

$$\hat{\mu}_{n+1} = g^{-1} \left(\hat{\alpha} + \sum_{i=1}^p \hat{\varphi}_i g(y_{n+1-i}) + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{n+1-j} \right).$$

2.3 Critérios de seleção

A escolha do modelo mais adequado para representar uma base de dados é de fundamental relevância, tanto no âmbito da análise de regressão em dados de corte transversal quanto de séries temporais, porém não é uma tarefa elementar. É necessário que o pesquisador se embase em princípios científicos além do conhecimento prévio acerca de um fenômeno (BURNHAN; ANDERSON, 2004). É neste contexto que se inserem os critérios de seleção, os quais são utilizados para identificar quais covariáveis devem ser inclusas em um determinado modelo de regressão.

Como o comportamento dos critérios de seleção pode diferir conforme o modelo a ser utilizado (EMILIANO, 2009), surge a necessidade não apenas do desenvolvimento de novos critérios para a seleção em diferentes classes de modelos (SHI; TSAI, 2002; FAN; TANG, 2013; BAYER; CRIBARI-NETO, 2015a; LV; LIU, 2014), mas também da realização de comparações de desempenho entre os critérios já existentes (STONE, 1979; HERWARTZ, 2007; SINIKSARAN, 2008). Em geral, quando o objeto de estudo é analisar qual critério mais se adapta a uma determinada classe de modelos, são consideradas as propriedades de consistência e eficiência assintótica dos critérios de seleção.

Assumindo que o verdadeiro modelo possui dimensão infinita e, por conseguinte, não está incluso na classe de modelos candidatos, um critério de seleção é dito assintoticamente eficiente se seleciona aquele modelo que possui o menor erro quadrático médio à medida que o tamanho da amostra aumenta (SHIBATA, 1981). Já o conceito de consistência considera que o verdadeiro modelo possui dimensão finita e faz parte do conjunto de modelos candidatos. Um critério de seleção é consistente se identifica o modelo correto com probabilidade assintoticamente igual a um (MCQUARRIE; TSAI, 1998).

A comparação do desempenho entre dois ou mais critérios de seleção em amostras de tamanho finito pode ser realizada através de simulações de Monte Carlo que consideram a frequência em que cada critério identifica o verdadeiro modelo. Este método é utilizado nos trabalhos de Marhuenda, Morales e Pardo (2014), que buscaram comparar o desempenho de alguns critérios de informação para o modelo Fay–Herriot, e Emiliano, Vivanco e Menezes (2014), que avaliaram o comportamento dos critérios de informação em modelos com distribuição normal e ARIMA. Baillie, Kapetanios e Papailias (2014), que propuseram o *Modified Information Criteria* (MIC) para séries temporais com memória longa, também utilizaram esta metodologia ao comparar a performance do MIC com os critérios de informação usuais.

Além disso, os trabalhos de Çetin e Erar (2002), Davies, Neath e Cavanaugh (2005) e Hu e Shao (2008) também buscaram avaliar a proporção em que cada critério é capaz de selecionar o modelo correto. O mesmo ocorre com os critérios de seleção introduzidos por Ishiguro, Sakamoto e Kitagawa (1997), Cavanaugh e Shumway (1997), Shibata (1997), Bengtsson e Cavanaugh (2006), Shang e Cavanaugh (2008) e Seghouane (2010). Estes são destacáveis por serem baseados na log-verossimilhança *bootstrap* e por buscarem melhores resultados no âmbito de pequenas amostras.

A maior parte dos trabalhos acima mencionados utiliza critérios baseados na distância de Kullback-Leibler (K-L) entre o modelo candidato e o verdadeiro. Esta classe de critérios de seleção, também chamados de critérios de informação (IC), é destacável por sua ampla aplicabilidade prática em termos de volume de utilização. Dentre estes, o AIC (AKAIKE, 1974) foi o primeiro a ser desenvolvido, sendo também o mais utilizado. A definição do AIC ocorre conforme segue:

$$\text{AIC} = -2\ell(y; \hat{\gamma}) + 2k,$$

em que o primeiro termo consiste numa estimativa da distância K-L por meio de $\ell(y; \hat{\gamma})$, que representa a log-verossimilhança do modelo estimado, e $2k$ é uma estimativa para o viés, em que

k representa o número de parâmetros do modelo.

O AIC é um estimador assintoticamente não viesado para a distância K-L e também é assintoticamente eficiente, entretanto seu desempenho pode ser pobre em pequenas amostras. Desta forma, Sugiura (1978) e Hurvich e Tsay (1989) buscaram melhorar o resultado do AIC em pequenas amostras e derivaram o AICc, onde

$$\text{AICc} = -2\ell(y; \hat{\gamma}) + \frac{2nk}{(n - k - 1)},$$

o qual representa uma versão corrigida do critério de Akaike em amostras menores. Tanto para o critério AIC quanto para os demais IC, o melhor modelo será aquele que obtiver o menor valor para o critério.

Outro critério bastante referenciado na literatura científica é o critério bayesiano de Schwarz (BIC), proposto por Akaike (1978) e Schwarz (1978) é definido por:

$$\text{BIC} = -2\ell(y; \hat{\gamma}) + k \log(n).$$

O BIC assume que o modelo original possui dimensão infinita e, portanto, não pertence ao grupo de modelos candidatos.

Assim como o AIC, o critério BIC é consistente e assintoticamente eficiente e não necessariamente estende suas propriedades para amostras menores. Com isso, McQuarrie (1999) utilizou a relação entre o AIC e o AICc para definir uma versão corrigida do critério de informação bayesiano, o BICc que é definido como:

$$\text{BICc} = -2\ell(y; \hat{\gamma}) + \frac{nk \log(n)}{n - k - 1}.$$

Hannan e Quinn (1979) também realizaram um estudo acerca dos critérios de seleção, porém com foco em modelos autorregressivos de séries temporais e propuseram outro critério assintoticamente eficiente para a seleção de modelos desse tipo, o HQ:

$$\text{HQ} = -2\ell(y; \hat{\gamma}) + 2k \log(n).$$

O critério HQ também pode ser expandido para outras classes de modelos, porém não é tão utilizado quanto os AIC e BIC.

A versão corrigida do HQ, o HQc, foi proposta por McQuarrie e Tsai (1998) e tem a seguinte estrutura:

$$\text{HQc} = -2\ell(y; \hat{\gamma}) + \frac{2k \log(\log(n))}{n - k - 1}.$$

3 AVALIAÇÃO NUMÉRICA DE CRITÉRIOS DE SELEÇÃO

Com implementação computacional desenvolvida em linguagem R (R Core Team, 2013), a avaliação numérica do desempenho dos critérios de seleção no modelo β ARMA foi realizada via simulações de Monte Carlo. Considerou-se $R = 10.000$ réplicas de Monte Carlo e cinco diferentes tamanhos amostrais, sendo $n = 30, 50, 80, 120, 200$. Em cada iteração são geradas n ocorrências da variável aleatória y_t , com densidade beta dada pela Equação (2.6), em que a média μ_t é determinada por uma estrutura semelhante à da Equação (2.7).

Também foram analisados cinco cenários de modelos. Estes diferem essencialmente nos vetores de parâmetros AR e MA do modelo β ARMA verdadeiro e também nas classes de modelos candidatos consideradas. A Tabela 3.1 descreve detalhadamente a composição de cada cenário.

Tabela 3.1 – Descrição dos cenários de acordo com as classes de modelos candidatos e o modelo verdadeiro

Cenários	Modelo Verdadeiro				Classes de modelos candidatos		
	φ_1	φ_2	θ_1	θ_2	β AR(p)*	β MA(q)**	β ARMA(p, q)***
Cenário 1	0,5	-0,4			Sim		
Cenário 2			0,46	-0,53		Sim	
Cenário 3	0,5	-0,4	0,46	-0,53	Sim	Sim	Sim
Cenário 4	0,5	-0,4			Sim	Sim	Sim
Cenário 5			0,46	-0,53	Sim	Sim	Sim

* com p variando de 1 até 6

** com q variando de 1 até 6

*** com p e q variando de 1 até 6

No Cenário 1, cujo modelo verdadeiro é o β AR(2), consideram-se como candidatos apenas a classe de modelos β AR, com ordem de defasagem p indo de um até seis. No Cenário 2, onde o modelo verdadeiro é um β MA(2), a classe de modelos candidatos analisada é o β MA, com ordem de defasagem q indo de um até seis. Já nos demais cenários, tanto o β AR, quanto o β MA e o β ARMA são considerados como modelos candidatos, onde os valores de p e q variam de um até seis. Todos os cenários serão gerados primeiramente com precisão $\phi = 20$ e posteriormente com $\phi = 120$, além de incluírem o termo de intercepto $\alpha = -1$. A Equação (3.1) apresenta o caso particular do β ARMA(2,2), como no Cenário 3, onde:

$$g(\mu_t) = \alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \theta_1 r_{t-1} + \theta_2 r_{t-2}, \quad (3.1)$$

em que ϕ é o parâmetro de precisão, α é o intercepto, φ_1 e φ_2 são os termos AR e θ_1 e θ_2 são os

termos MA. A função de ligação utilizada é a logit, ou seja,

$$\text{logit}(\mu_t) = \log\left(\frac{\mu_t}{1 - \mu_t}\right).$$

Dessa forma, a média μ_t de cada y_t , para $t = 1, \dots, n$, é:

$$\mu_t = \frac{\exp(\alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \theta_1 r_{t-1} + \theta_2 r_{t-2})}{1 + \exp(\alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \theta_1 r_{t-1} + \theta_2 r_{t-2})}.$$

Para obter tais expressões em relação aos Cenários 1 e 4 basta igualar a zero os parâmetros θ_1 e θ_2 . Analogamente, ao considerar que φ_1 e φ_2 possuem valor zero, obtém-se as expressões que representam os Cenários 2 e 5.

Uma vez que o modelo verdadeiro pertence à classe de modelos candidatos, em cada réplica de Monte Carlo são ajustados e calculados os valores dos critérios em todos os possíveis modelos pertencentes a este conjunto. Desta forma, o critério mais adequado a um dado cenário será aquele que selecionar com maior frequência o verdadeiro modelo. Também será verificada a proporção em que cada critério selecionou um modelo com ordem de defasagens k igual (cenários 3, 4 e 5), superior (sobreidentificação) e inferior (subidentificação) à do modelo verdadeiro. Os critérios de seleção avaliados são os descritos na Seção 2.3, nomeadamente: AIC, AICc, BIC, BICc, HQ e HQc.

Esta abordagem para avaliação de critérios de seleção é amplamente aplicada quando o objeto de estudo é identificar o melhor critério para uma determinada classe de modelos. A metodologia também foi utilizada por Hannan e Quinn (1979), Hurvich e Tsai (1989), McQuarrie, Shumway e Tsai (1997), Pan (1999), Shang e Cavanaugh (2008), Acquah (2010) e Bayer e Cribari-Neto (2015a).

Os resultados das porcentagens de identificação do modelo verdadeiro para cada cenário e diferentes valores de precisão encontram-se na forma de resumo gráfico nas Figuras 3.1 e 3.2 e os resultados completos encontram-se nas Tabelas 3.2, 3.3 e no Apêndice A. Conforme esperado devido às propriedades de consistência e eficiência assintótica, em geral, os critérios melhoraram seu desempenho à medida que o tamanho da amostra aumenta. Também é possível perceber que, os resultados foram muito próximos para os diferentes valores de precisão.

No Cenário 1, verifica-se que todos os critérios obtiveram altos graus de identificabilidade quando a classe de modelos candidatos consiste apenas nos β AR. Para $n = 200$, o AIC, que é o critério com pior resultado neste caso, identificou o modelo correto em 63,72% das vezes quando $\phi = 20$ e 63,23% para $\phi = 120$. Quando os termos de médias móveis passam a integrar a classe modelos candidatos (Cenário 4), há uma queda expressiva na capacidade de

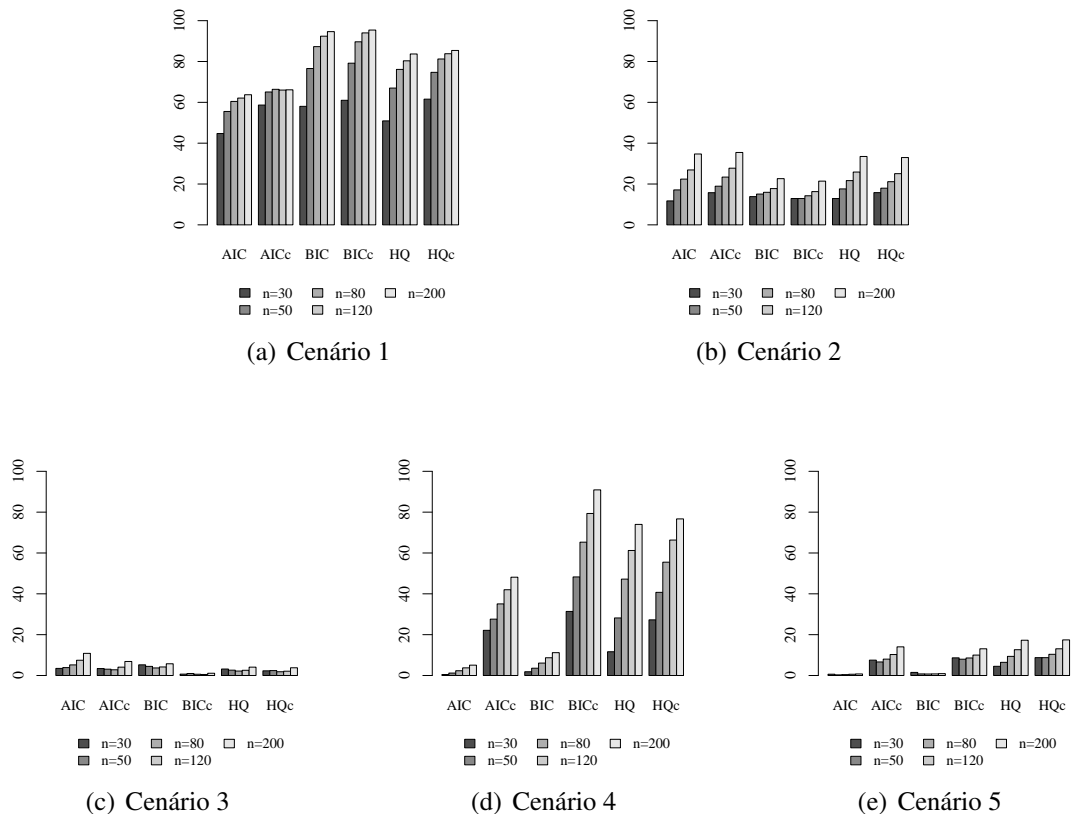


Figura 3.1 – Frequência (em %) de identificação dos modelos com precisão $\phi = 20$

identificação do AIC, AICc e BIC, porém os demais critérios mantiveram a alta identificabilidade mesmo em amostras pequenas.

Para os modelos β MA as porcentagens de identificação foram inferiores em comparação com o β AR. No Cenário 2 o AICc, que obteve o melhor resultado, identificou o modelo verdadeiro 35,45% das vezes para $\phi = 20$ e 34,18% para $\phi = 120$, ambos em $n = 200$. No Cenário 5 o melhor resultado para o maior tamanho amostral foi do HQc (17,44% e 14,37% para $\phi = 20$ e $\phi = 120$, respectivamente). Assim como no β AR, no β MA o AIC e o BIC diminuem consideravelmente seu desempenho quando aumentado o número de modelos candidatos independentemente do grau de precisão.

Com relação ao Cenário 3, foi este o cenário que obteve o desempenho menos favorável em comparação com os demais. As Tabelas 3.2 e 3.3 demonstram que, à medida que o tamanho da amostra aumenta, também aumenta o número de vezes que os critérios identificam o modelo verdadeiro. Entretanto, mesmo com tamanhos amostrais maiores, o β ARMA subidentificou o verdadeiro modelo mais vezes do que identificou para todos os critérios exceto o AIC. .

As tabelas também evidenciam que nem sempre o critério que mais identifica o número

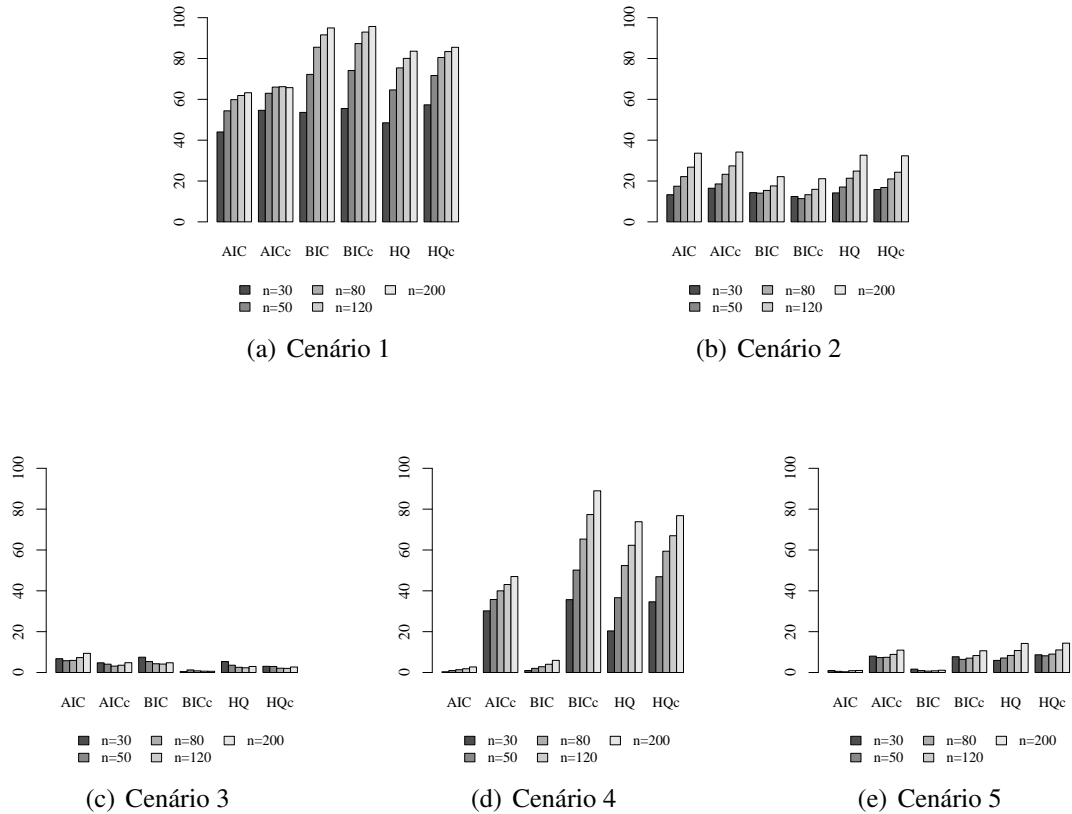


Figura 3.2 – Frequência (em %) de identificação dos modelos com precisão $\phi = 120$

correto de parâmetros k é aquele que mais seleciona o modelo verdadeiro. O caso mais expressivo é o do BIC. Este critério seleciona um modelo com $k = 4$ em 19,01% dos casos quando $n = 30$ e $\phi = 20$. Porém, o grau de identificação do modelo correto, ou seja, o β ARMA(2, 2) é de apenas 5,22% nas mesmas condições. No Apêndice A encontram-se discriminadas para os demais cenários as frequências de sobreidentificação, subidentificação e de identificação da quantidade de defasagens k iguais à do modelo correto.

Em geral, há um decaimento no desempenho de todos os critérios à medida que aumenta a complexidade do modelo verdadeiro ou da classe de modelos candidatos. Além disso, é observável que os termos de médias móveis são menos passíveis de identificação que os autorregressivos. Este resultado converge com o apontado por Emiliano, Vivanco e Menezes (2014) ao realizar simulações para avaliar o grau de identificabilidade do AIC, AICc e BIC em modelos AR(1), AR(2), ARMA(2, 1), ARMA(2, 2), MA(1) e MA(2). Ademais, conforme McQuarrie e Tsai (1998), vale ressaltar que a magnitude dos parâmetros do modelo também pode influenciar na sua identificação.

De acordo com Torres (2005), nem sempre é possível eleger o melhor critério, pois

seu desempenho depende do modelo verdadeiro. Esta afirmação converge com o resultado obtido para os critérios AIC e BIC, que apresentaram altas divergências de resultados em termos de capacidade de identificação do modelo correto entre os diferentes cenários. Já o BICc, HQ e HQc alcançaram os melhores níveis de identificação em comparação com os demais critérios, sendo os mais indicados para a seleção de modelos da classe β ARMA. Quando o modelo verdadeiro é o β AR, o BICc seleciona o modelo correto uma quantidade de vezes ligeiramente superior ao HQ e HQc. Em contrapartida, estes últimos apresentaram melhores porcentagens de identificação quando os modelos verdadeiros são β MA ou β ARMA. Desta forma, ao realizar aplicações do β ARMA, sugere-se a utilização dos critérios BICc, HQ e HQc concomitantemente, uma vez que estes mantiveram um comportamento semelhante entre si.

Devido à semelhança nos resultados, indica-se que, em auxílio à utilização destes três critérios para a seleção de modelos da classe β ARMA, seja também considerada para a escolha do modelo a análise de diagnóstico dos resíduos por meio dos resíduos padronizados, da estatística Q , de Ljung e Box (1978), e a capacidade de previsão através do cálculo da raiz do erro médio quadrático (RMSE) e o erro absoluto médio percentual (MAPE) entre os valores reais observados e os preditos pelo modelo. Já na Seção 4 pode-se verificar o comportamento dos critérios BICc, HQ e HQc no âmbito de aplicações. Nesta, os critérios são utilizados como ferramenta de auxílio na determinação de qual modelo β ARMA é mais adequado ao comportamento do índice de inadimplência de crédito e à relação entre o crédito consignado e o crédito total pessoa física.

Tabela 3.2 – Frequência (em %) de identificação no Cenário 3 com precisão $\phi = 20$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 4$	73,74	22,37	34,36	1,89	49,49	11,01
$k = 4$	12,96	18,52	19,01	5,76	15,65	14,24
$k < 4$	13,30	59,11	46,63	92,35	34,86	74,75
modelo verdadeiro	3,47	3,40	5,22	0,67	3,17	2,28
$n = 50$						
$k > 4$	68,48	33,34	20,01	2,46	34,21	13,43
$k = 4$	11,63	14,29	15,64	5,79	12,40	12,04
$k < 4$	19,89	52,37	64,35	91,75	53,39	74,53
modelo verdadeiro	3,87	3,09	4,46	0,98	2,61	2,41
$n = 80$						
$k > 4$	56,19	32,31	9,47	1,92	20,01	11,07
$k = 4$	15,18	13,41	13,25	4,30	11,32	10,12
$k < 4$	28,63	54,28	77,28	93,78	68,67	78,81
modelo verdadeiro	5,19	2,80	3,70	0,62	2,18	1,84
$n = 120$						
$k > 4$	47,67	29,81	4,85	1,15	12,67	8,22
$k = 4$	17,69	15,02	12,25	3,42	11,26	9,99
$k < 4$	34,64	55,17	82,90	95,43	76,07	81,79
modelo verdadeiro	7,47	4,09	4,18	0,50	2,55	2,10
$n = 200$						
$k > 4$	41,67	27,86	2,71	0,53	8,12	6,21
$k = 4$	21,02	19,29	11,44	3,82	13,30	12,44
$k < 4$	37,31	52,85	85,85	95,65	78,58	81,35
modelo verdadeiro	10,84	6,87	5,74	1,10	4,09	3,78

Tabela 3.3 – Frequência (em %) de identificação no Cenário 3 com precisão $\phi = 120$

Ordens do modelo	AIC	AICc	BIC	BICc	HQ	HQc
<i>n</i> = 30						
<i>k</i> > 4	63,71	14,20	24,27	1,36	36,38	6,56
<i>k</i> = 4	13,17	13,28	15,52	2,98	13,13	8,96
<i>k</i> < 4	23,12	72,52	60,21	95,66	50,49	84,48
modelo verdadeiro	6,75	4,74	7,49	0,43	5,37	3,03
<i>n</i> = 50						
<i>k</i> > 4	60,75	26,03	14,58	1,43	25,96	9,53
<i>k</i> = 4	11,66	11,79	12,1	3,59	10,31	8,76
<i>k</i> < 4	27,59	62,18	73,32	94,98	63,73	81,71
modelo verdadeiro	5,78	4,08	5,35	1,25	3,57	2,95
<i>n</i> = 80						
<i>k</i> > 4	54,90	30,16	7,29	1,14	16,86	8,62
<i>k</i> = 4	12,44	11,22	10,08	2,58	9,12	7,47
<i>k</i> < 4	32,66	58,62	82,63	96,28	74,02	83,91
modelo verdadeiro	5,92	3,14	4,28	0,81	2,55	2,11
<i>n</i> = 120						
<i>k</i> > 4	50,48	30,86	4,81	0,87	12,46	8,03
<i>k</i> = 4	14,94	13,02	9,65	2,70	9,77	8,46
<i>k</i> < 4	34,58	56,12	85,54	96,43	77,77	83,51
modelo verdadeiro	7,30	3,56	4,13	0,64	2,31	2,02
<i>n</i> = 200						
<i>k</i> > 4	43,42	30,46	2,63	0,42	8,75	6,70
<i>k</i> = 4	18,25	15,95	9,53	3,11	11,28	10,40
<i>k</i> < 4	38,33	53,59	87,84	96,47	79,97	82,90
modelo verdadeiro	9,38	4,80	4,76	0,59	2,96	2,69

4 APLICAÇÕES A DADOS REAIS

A aplicação dos critérios de seleção e do β ARMA em dados reais, ocorrerá no contexto do mercado de crédito do SFN. A definição do tema se justifica pela relação complementar entre o setor e o desenvolvimento da atividade econômica do país (ALMEIDA, 2013). Por representar uma possibilidade de financiamento às empresas (SANTOS, 2008), o mercado de crédito provém recursos que permitem investimentos em diversos setores da economia (SOUZA, 2007) e estimulam o crescimento econômico (LEVINE, 1997).

O comportamento do mercado de crédito afeta a tomada de decisões de diferentes segmentos sociais, econômicos e políticos, tendo implicações tanto em momentos de expansão quanto de retração econômica (SAITO, 2012). No Brasil, o comportamento do volume de crédito disponibilizado no SFN apresenta forte crescimento como proporção do Produto Interno Bruto (PIB). Esta relação chegou ao patamar de 55,66% no mês de dezembro de 2014 (Figura 4.1).

Saito, Savoia e Lazier (2014) verificaram que os países integrantes da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) também registraram crescimento consistente de suas relações de Crédito/PIB, apresentando diferenças apenas no ritmo da evolução. Além disso, algumas mudanças institucionais que buscaram atenuar o risco inerente às operações no mercado de crédito também contribuem para tal cenário (MENDONÇA; SACHSIDA, 2013).

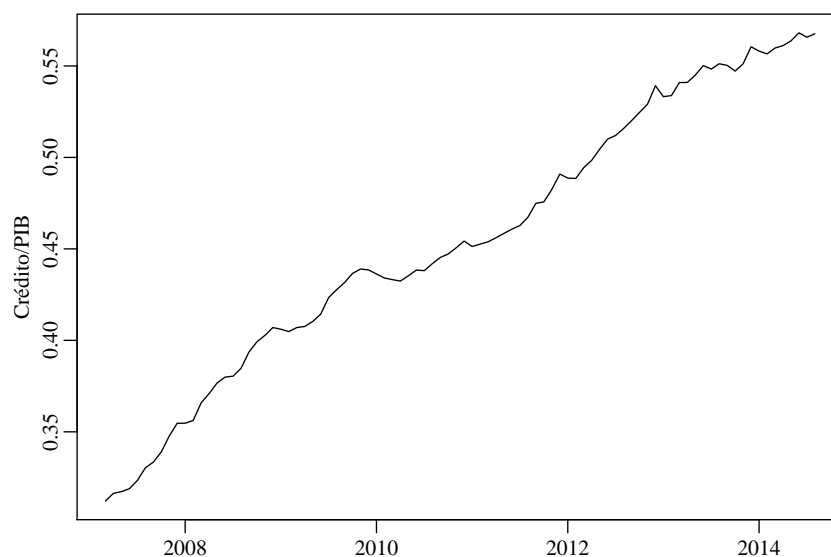


Figura 4.1 – Evolução da participação crédito/PIB no Brasil

Em face a esta realidade, as variáveis aqui utilizadas consistem no índice de inadimplência de crédito e a relação entre o crédito consignado em folha de pagamento e o total da carteira de crédito pessoa física (CC/Crédito PF). Ambas são variáveis percentuais do SFN e, portanto, restritas ao intervalo $(0, 1)$, o que justifica a utilização do β ARMA. Esta escolha é também motivada pela relevância das mesmas na dinâmica de crédito brasileira. Em outros âmbitos da engenharia de produção, o uso do β ARMA também pode ser estendido a séries temporais restritas a intervalos duplamente limitados.

Além destas aplicações, as áreas de economia, ciências sociais aplicadas, física, meteorologia, epidemiologia e os demais campos da engenharia também são exemplos de áreas com amplas possibilidades de aplicações para o β ARMA. Assim, ao avaliar se há superioridade deste modelo na previsão do índice de inadimplência e na relação CC/Crédito PF, bem como apresentar uma aplicação dos critérios de seleção aqui analisados, o presente trabalho pretende fornecer um importante instrumento na construção de aplicações também em outras variáveis. Além da modelagem e previsão das séries temporais através do β ARMA, também é considerado comparativamente o desempenho dos modelos ARIMA.

4.1 Modelando o índice de inadimplência

Por estar diretamente interligado à mensuração do risco de crédito (CIFTER; YILMAZER; CIFTER, 2009; CHAIBI; FTITI, 2015), o nível de inadimplência é destacado na gestão das instituições financeiras e das agências reguladoras (MARINS; NEVES, 2013). Também verifica-se a ligação entre esta variável e o cenário macroeconômico (LINARDI, 2008; CASTRO, 2013), podendo-se destacar sua relação negativa com a oferta de crédito (MENDONÇA; SACHSIDA, 2013), a estabilidade financeira dos bancos (TABAK; CRAVEIRO; CAJUEIRO, 2010) e o *spread* bancário (MANHIÇA; JORGE, 2012).

Um dos principais indicadores de inadimplência é o atraso com mais de 90 dias por parte do tomador, conceito utilizado no Novo Acordo de Capital de Basiléia (Basiléia II) publicado em 2004 (BCBS, 2004). Por ser um indicador de qualidade dos ativos (MEEKER; GRAY, 1987), este índice é acompanhado com bastante atenção pelo mercado financeiro brasileiro (ANNIBAL, 2009). Trabalhos como os de Barseghyan (2010) e Abid, Ouertani e Zouari-Ghorbel (2014) buscam analisar a inadimplência de crédito através da identificação das variáveis determinantes de suas flutuações. Já Crook e Banasik (2012), que estimam modelos ARIMA para comparação com modelos de vetores autorregressivos, e Zaniboni (2013), que

utilizam modelos ARIMA e ARMA com variáveis explicativas (ARMAX), atentam para a necessidade de construir projeções eficazes para a inadimplência também em termos das próprias informações passadas, como é o caso deste estudo.

Para tanto, a base de dados utilizada é proveniente do Sistema Gerenciador de Séries Temporais (SGS)¹ do Banco Central do Brasil. O SGS disponibiliza os valores ao longo do tempo para uma série de variáveis relacionadas ao crédito do SFN e tem como principal fonte o Sistema de Informações de Crédito (SCR), da mesma instituição. Foi utilizada a série temporal mensal da inadimplência agregada no período de janeiro de 2002 a maio de 2014, sendo os últimos seis meses reservados para a análise da capacidade preditiva dos modelos. Este índice consiste no percentual da carteira de crédito com pelo menos uma parcela de atraso superior a 90 dias e engloba operações contratadas no segmento de crédito livre e crédito direcionado, não fazendo distinção por modalidade de crédito e nem entre clientes pessoa física e jurídica.

Ao analisar o comportamento da série ao longo do período, Figura 4.2, observa-se um movimento de declínio. Segundo Mendonça e Sachsida (2013) tais diminuições inicialmente podem ser explicadas pela expansão do volume das operações de crédito ocorrida no Brasil desde os anos 2000. Partindo de patamares por volta dos 0,049 no início de 2002, a partir de 2004 a inadimplência passou a se manter em níveis próximos aos seus valores médios e medianos de 0,038 e 0,036, respectivamente (Tabela 4.1).

Tabela 4.1 – Medidas descritivas para a inadimplência de crédito

Medidas	Inadimplência
Média	0,038
Mediana	0,036
Desvio Padrão	0,006
Mínimo	0,028
Máximo	0,056
Assimetria	1,15201
Curtose	3,84084
Jarque-Bera	p -valor < 0,001

Além do aumento no volume de crédito, o período também coincide com a adoção de algumas reformas institucionais que deram mais segurança na concessão de crédito no Brasil. São estes os casos da criação do SCR em 2002 e das regulamentações do crédito consignado em 2003, da alienação fiduciária no mercado de capitais, de automóveis e imóveis em 2004. Ademais, é de 2004 a publicação de Basiléia II (BCBS, 2004), que impõe medidas para uma

¹ Endereço web: <https://www3.bcb.gov.br/sgspub/>

gestão mais parcimoniosa dos riscos de crédito, de mercado e do risco operacional (YANAKA; HOLLAND, 2010).

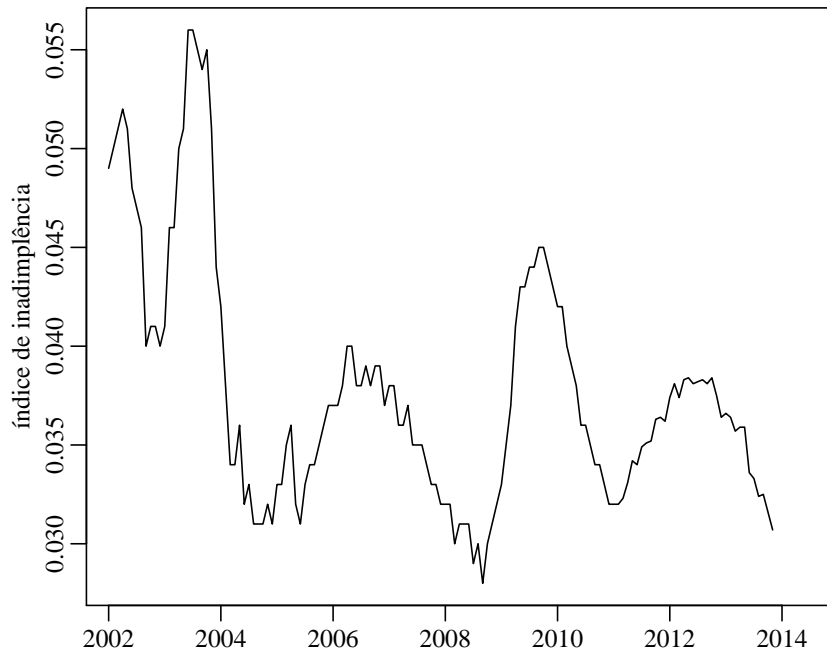


Figura 4.2 – Índice de inadimplência do SFN no período de março de 2011 a dezembro de 2013

O final de 2008 e ano subsequente, época que ocorreu a crise financeira global, também é destacável pela alteração no comportamento da série, onde houve aumento na taxa de inadimplência em comparação com o período anterior à crise. Conforme Vazquez, Tabak e Souto (2012) a crise financeira global impactou o Brasil no terceiro trimestre de 2008 e primeiro trimestre de 2009. Segundo os autores, a rápida deterioração na qualidade do crédito no Brasil que houve neste período foi um reflexo do comportamento das variáveis financeiras e macroeconômicas. Entretanto, Vazquez, Tabak e Souto (2012) também verificaram que o mercado de crédito doméstico alcançou um certo isolamento dos efeitos da crise, o que é provavelmente atribuído à forte expansão do volume de crédito como medida compensatória.

Quanto às características sazonais da variável, fica evidente que a redução da inadimplência do SFN ocorreu de forma gradativa ao longo do tempo, havendo apenas alguns picos que correspondem justamente aos anos de 2004 e 2008 e se estendem por alguns meses. O comportamento da série é indiferente à época do ano, tendo desempenho semelhante ao longo dos meses.

Já os coeficientes de assimetria e curtose, apresentados na Tabela 4.1, permitem verificar que os resultados divergem consideravelmente daqueles esperados no caso em que a variável

possui distribuição normal. Fato que é corroborado pelo teste de Jarque-Bera (JB) (JARQUE; BERA, 1987) que refuta a hipótese de que a variável é normalmente distribuída. O histograma da variável (Figura 4.3) também fornece subsídios para a rejeição da hipótese de normalidade. As observações se concentram essencialmente na primeira metade da figura, sendo que a maior parte dos valores se encontra entre 0,03 e 0,04 e o restante se distribui irregularmente na cauda direita. Cribari-Neto e Souza (2013) também verificaram esse tipo de comportamento em sua análise. Eles salientam que variáveis do tipo taxa e proporção, em geral, não são normalmente distribuídas, sendo tipicamente assimétricas. Neste caso, a utilização do modelo ARIMA, que pressupõe normalidade dos dados, ficaria comprometida.

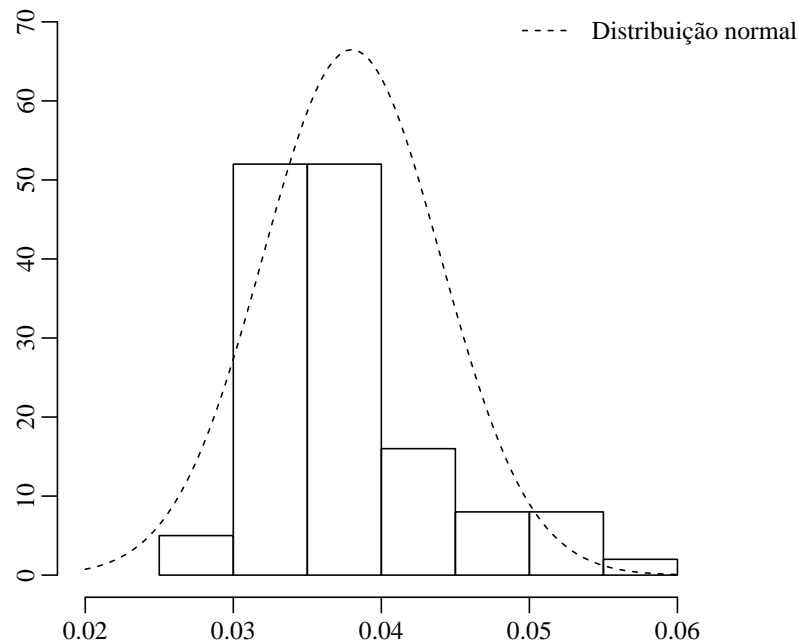


Figura 4.3 – Histograma e densidades estimadas do índice de inadimplência do SFN

Percebe-se que o índice de inadimplência de crédito possui vasta amplitude, curtose e ausência de simetria (Tabela 4.1). Variáveis com grandes oscilações tendem a não serem estacionárias, pois não mantêm sua média e variância constantes ao longo do tempo. Fato este que é agravado pelo comportamento descendente do correlograma da variável, que parte de seus valores máximos e decai lentamente conforme a defasagem aumenta (Figura 4.4). Este é um comportamento típico de séries temporais não estacionárias. Desta forma, cabe a realização de algum teste para assegurar a estacionaridade da série. Neste estudo considera-se o teste Kwiatkowski, Phillips, Schmidt e Shin (KPSS) (KWIATKOWSKI et al., 1992) para estacionaridade.

O resultado do teste KPSS rejeita a hipótese nula de estacionaridade da série, com p -

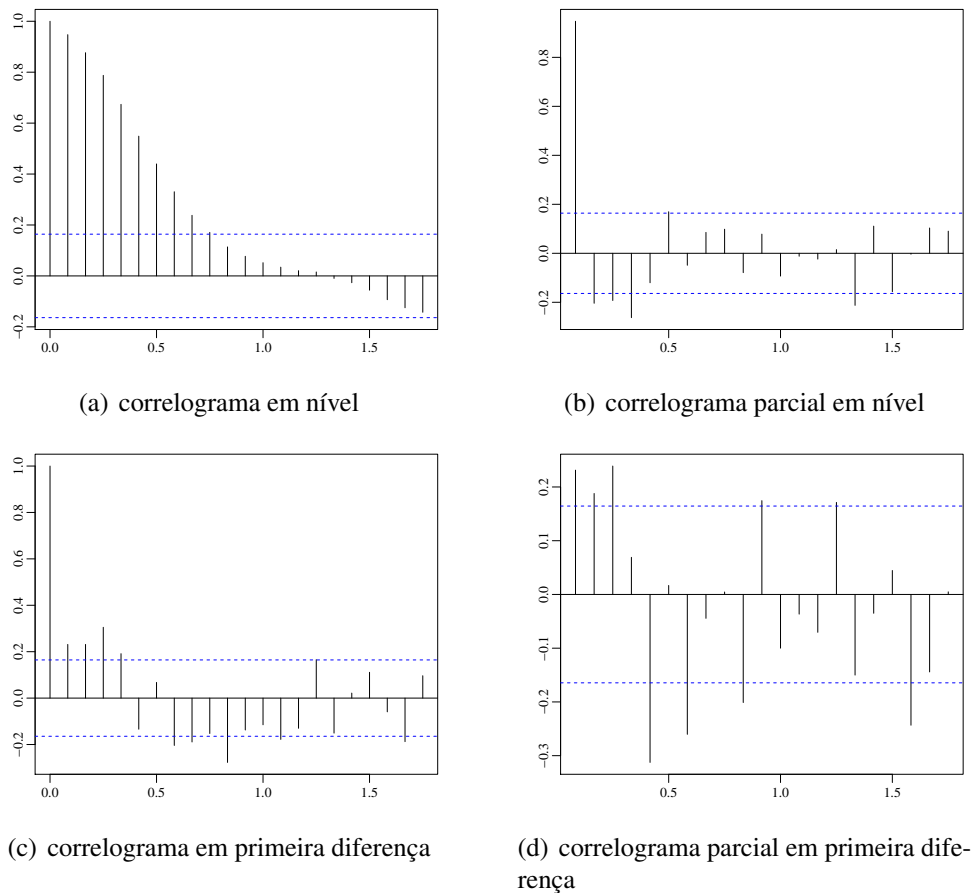


Figura 4.4 – Correlogramas do índice de inadimplência em nível e em primeira diferença

valor igual a 0,06689. Este resultado é convergente com o observado dos correlogramas em nível (Figura 4.4). Assim, conforme já mencionado na Seção 2.1, a alternativa mais indicada quando há ausência de estacionaridade em uma série temporal consiste em tomar a sua diferença. Ao realizar a diferenciação da série temporal observa-se que a mesma deixa de apresentar o padrão decrescente característico de processos não estacionários. Uma vez que o teste KPSS corroborou a hipótese de que a série é estacionária em primeira diferença a um nível de significância de 1%, assumimos que o processo gerador da série é integrado de ordem um $I(1)$ e utilizamos a transformação para aplicação do β ARMA em intervalos limitados. Na Figura 4.5 é apresentado o comportamento da série em primeira diferença.

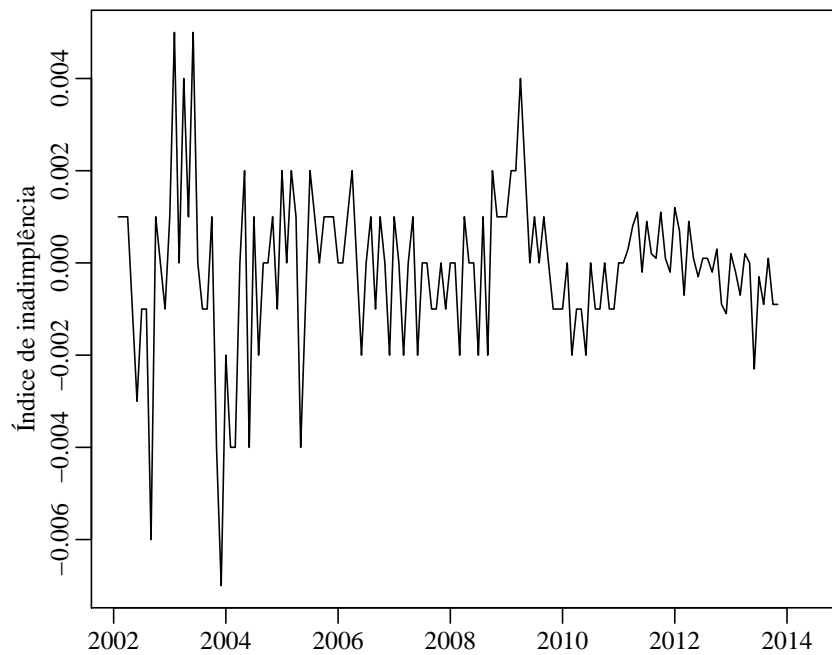


Figura 4.5 – Série de inadimplência de crédito em primeira diferença

4.1.1 Modelagem e discussão

Tendo-se identificado a ordem de integração da série temporal é necessário determinar a ordem de p e q dos modelos ARIMA e β ARMA. Para a definição das defasagens, foram analisados os critérios AIC e BIC para o modelo ARIMA, por serem os mais usuais para o referido modelo (HOTTA; NETO, 1993; MACDONALD; LERER, 1994; DE LUNA, 2000; MAIA, 2008; BISWAS; BHATTACHARYYA, 2013; MONDAL; SHIT; GOSWAM, 2014; KIM et al., 2015). Para o β ARMA foram utilizados os critérios BICc, HQ e HQc, conforme indicado no estudo de simulação (Seção 3). Foram geradas diferentes combinações de p e q , com seus valores variando de zero até dez e selecionados os ARIMA com menores resultados nos critérios AIC e BIC. Para o β ARMA foram selecionados os modelos com menores resultados para o BICc, HQ e HQc.

A partir destes, levou-se em consideração a significância dos parâmetros estimados e a análise de diagnóstico dos resíduos. Esta técnica é denominada superajustamento, ou seja, são estimados modelos com parâmetros extras para verificar se sua inclusão diminui significativamente a variância residual caso estes sejam diferentes de zero (MORETTIN; TOLOI, 2006).

No caso do ARIMA o modelo que apresentou o melhor ajuste foi o ARIMA(1,1,5) com termos de médias móveis que incluem as defasagens 1,3,5,6 e 7 (Tabela 4.2). Dentre os dez modelos com menor valor para os critérios de seleção, este possui resultados ligeiramente inferiores aos demais para o MAPE e RMSE. Além disso, o ARIMA(1,1,5) com defasagens 1,3,5,6 e 7 para as médias móveis possui menor valor de AIC dentre os modelos analisados. De acordo com o teste Ljung e Box, este modelo também não rejeita a hipótese nula de que os resíduos são ruído branco.

Tabela 4.2 – Critérios para identificação do modelo ARIMA para a inadimplência de crédito

Modelo	AIC	BIC	$Q[p\text{-valor}]$	RMSE	MAPE
ARIMA(1,1,1) ¹	-1420,76	-1411,89	0,6300 [0,43]	0,001586	2,97802
ARIMA(1,1,2) ²	-1424,31	-1412,49	0,6879 [0,41]	0,001554	2,87825
ARIMA(1,1,3) ³	-1429,78	-1415,00	0,4166 [0,52]	0,001512	2,86007
ARIMA(1,1,4) ³	-1444,83	-1427,10	0,0590 [0,81]	0,001419	2,67165
ARIMA(1,1,5) ³	-1443,21	-1422,52	0,0266 [0,87]	0,001417	2,68207
ARIMA(1,1,5)²⁴	-1451,25	-1421,69	0,0100 [0,96]	0,001380	2,52979
ARIMA(1,1,6)	-1444,24	-1420,59	0,0012 [0,97]	0,001402	2,64886
ARIMA(1,1,7)	-1447,61	-1421,01	0,0590 [0,81]	0,001373	2,55105
ARIMA(1,1,8)	-1445,62	-1416,07	0,0488 [0,83]	0,001373	2,55537
ARIMA(1,1,9)	-1444,54	-1416,98	0,0002 [0,99]	0,001368	2,56176

¹ Todas as estimativas significativas a pelo menos 1%

² Todas as estimativas significativas a pelo menos 5%

³ Todas as estimativas significativas a pelo menos 10%

⁴ termo de médias móveis inclui as defasagens MA= (1, 3, 5, 6, 7)

Ao analisar os créditos inadimplentes mediante a ótica dos modelos β ARMA, conforme Tabela 4.3, o β ARMA(1,1) é a escolha mais adequada. Além de possuir todas as estimativas dos parâmetros significativas, exceto o intercepto, foi este o modelo que obteve os melhores resultados em termos dos critérios BICc, HQ e HQc. Dentre os modelos com estimativas significativas o β ARMA(1,1) obteve um valor intermediário em termos de MAPE e o menor valor de RMSE, juntamente com β ARMA(2,1). A respeito do teste de Ljung e Box, nenhum dos modelos apresentados refuta a hipótese de que os resíduos desse são ruído branco.

As Tabelas 4.4 e 4.5 apresentam respectivamente as estimativas dos parâmetros dos modelos ARIMA e β ARMA selecionados. No ARIMA todos os parâmetros são significativos a 1%. Já no β ARMA a significância das estimativas ocorre a 5%, exceto no intercepto. Além disso, em ambos os casos foram encontrados resultados razoáveis em termos de MAPE, RMSE. Assim, tendo selecionado os melhores modelos, passa-se à análise das projeções obtidas nas duas modelagens.

Tabela 4.3 – Critérios para identificação do modelo β ARMA para a inadimplência de crédito

Modelo	BICc	HQ	HQc	$Q[p\text{-valor}]$	RMSE	MAPE
β AR(1) ¹	-1589,70	-1595,39	-1595,11	0,3124 [0,58]	0,00081	1,49775
β AR(2) ²	-1577,47	-1585,21	-1584,74	0,3270 [0,57]	0,00080	1,48597
β AR(3)	-1568,27	-1578,14	-1577,43	0,0276 [0,87]	0,00077	1,42646
βARMA(1,1)²	-1591,22	-1598,96	-1598,49	0,6820 [0,41]	0,00079	1,47890
β ARMA(1,2)	-1576,38	-1586,25	-1585,54	0,0831 [0,77]	0,00078	1,45699
β ARMA(2,1) ²	-1575,74	-1585,61	-1584,90	0,0141 [0,91]	0,00079	1,46170
β MA(1,3)	-1568,29	-1580,36	-1579,36	0,0001 [0,99]	0,00075	1,41443
β MA(1)	-1587,40	-1593,09	-1592,82	0,0060 [0,93]	0,00082	1,49718
β MA(2)	-1572,15	-1579,90	-1579,43	0,3115 [0,58]	0,00081	1,48929
β ARMA(2,2) ¹	-1569,37	-1581,44	-1580,45	0,0017 [0,97]	0,00079	1,46598

¹ Todas as estimativas significativas a pelo menos 1%

² Todas as estimativas significativas a pelo menos 5%

Tabela 4.4 – Estimativas do modelo ARIMA selecionado para a inadimplência de crédito

	Estimativa	Desvio Padrão	$z[p\text{-valor}]$
$\hat{\varphi}_1$	0,8406	0,0854	9,8459[<0,001]
$\hat{\theta}_1$	-0,6889	0,1131	-6,0887[<0,001]
$\hat{\theta}_3$	0,1911	0,0686	2,7857[0,0053]
$\hat{\theta}_5$	-0,4352	0,0780	-5,5791[<0,001]
$\hat{\theta}_6$	0,3236	0,0964	3,3574[<0,001]
$\hat{\theta}_7$	-0,3154	0,0688	-4,5870[0,0045]

Tabela 4.5 – Estimativas do modelo β ARMA selecionado para a inadimplência de crédito

	Estimativa	Desvio Padrão	$z[p\text{-valor}]$
$\hat{\alpha}$	-0,0001	0,0003	0,2891[0,7725]
$\hat{\varphi}_1$	0,7347	0,2736	2,6850[0,0073]
$\hat{\theta}_1$	-2,0500	1,1550	1,7748[0,0759]

Uma das formas para verificação de qual modelo explica melhor o comportamento do índice de inadimplência é através da comparação dos resultados das previsões. Em primeira instância a análise gráfica é relevante por apresentar indícios acerca da acurácia preditiva dos modelos. Desta forma, na Figura 4.6 são apresentados os valores preditos dos respectivos modelos ARIMA e β ARMA selecionados em contraponto com a série original.

Verifica-se que ambos os modelos foram capazes de captar adequadamente os movimentos da série temporal, tendo apresentado resultados semelhantes aos originais, porém com superioridade do β ARMA. Além disso, através da Figura 4.7, é possível visualizar que os resíduos padronizados do modelo β ARMA ajustado possuem comportamento aleatório ao redor

do zero e todas as ocorrências no intervalo (-3,3). Por outro lado, os resíduos padronizados do modelo ARIMA apresenta amplitude variável ao longo das observações, indicando possível presença de heteroscedasticidade, e também alguns valores fora do intervalo (-3,3). Já no caso da Tabela 4.6, enquanto o ARIMA forneceu previsões com RMSE de 0,00194 e MAPE de 6,36341, o β ARMA tem um valor de apenas 0,00052 para a raiz do erro médio quadrático e 1,40817 para o erro absoluto médio percentual.

Tabela 4.6 – Previsões para o índice de inadimplência de crédito para meses fora da amostra

Período	Valor observado	Previsão ARIMA	Previsão β ARMA
dez/2013	0,02960	0,03088	0,03026
jan/2014	0,02970	0,03165	0,02990
fev/2014	0,02970	0,03176	0,02995
mar/2014	0,03000	0,03216	0,02990
abr/2014	0,03040	0,03256	0,03002
mai/2014	0,03110	0,03298	0,03004
MAPE	–	6,36341	1,40817
RMSE	–	0,00194	0,00055

Na Figura 4.8 encontram-se os correlogramas dos resíduos padronizados dos modelos ARIMA e β ARMA ajustados. Verifica-se que os resíduos padronizados obtidos através do modelo β ARMA são mais favoráveis, em que todas as autocorrelações e autocorrelações parciais residuais não são significativamente diferentes de zero. Assim, como ainda não há outras aplicações do modelo na área de engenharia de produção e econometria, mesmo no âmbito internacional, este resultado é um primeiro indício empírico de que a utilização do modelo β ARMA pode efetivamente causar um incremento positivo nos resultados das previsões de séries temporais de razões ou proporções, tanto em termos de redução dos erros de previsão, quanto na melhora no ajuste.

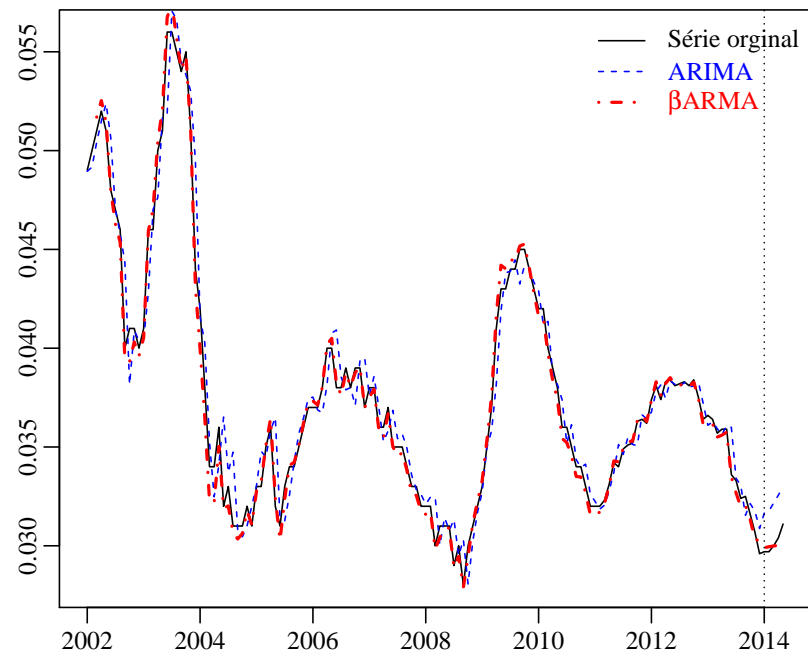


Figura 4.6 – Valores originais e previstos dentro e fora da amostra para a série de inadimplência de crédito do SFN

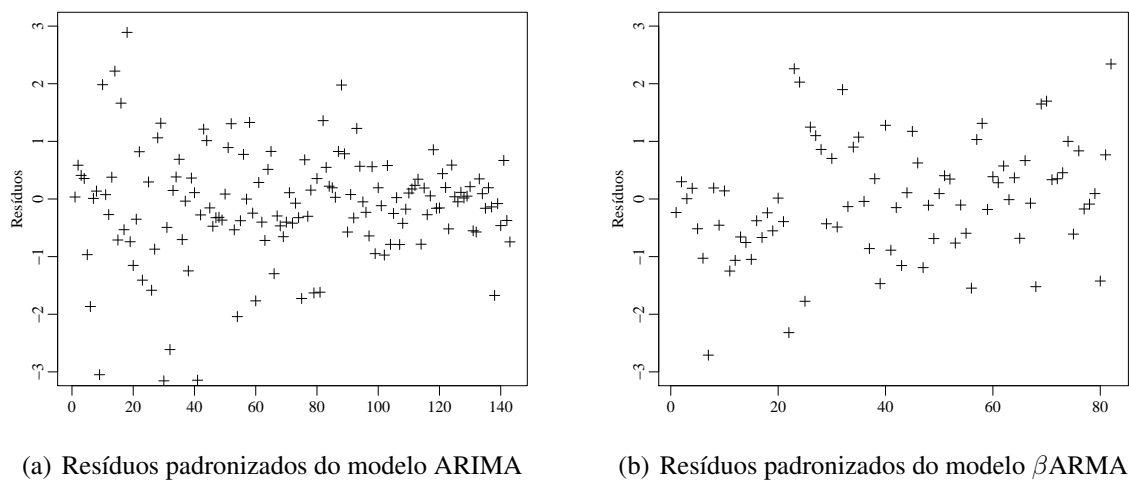
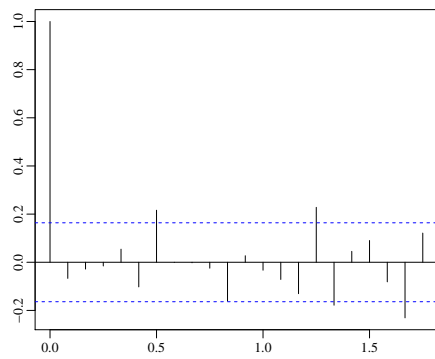
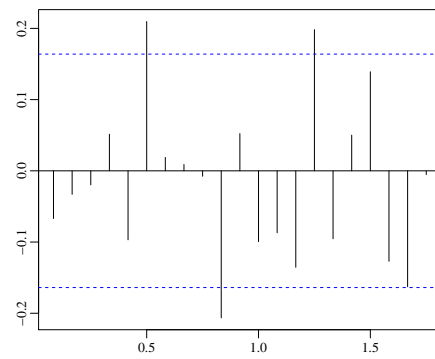


Figura 4.7 – Resíduos padronizados dos modelos para a inadimplência de crédito



(a) correlograma dos resíduos do modelo ARIMA



(b) correlograma parcial dos resíduos do modelo ARIMA

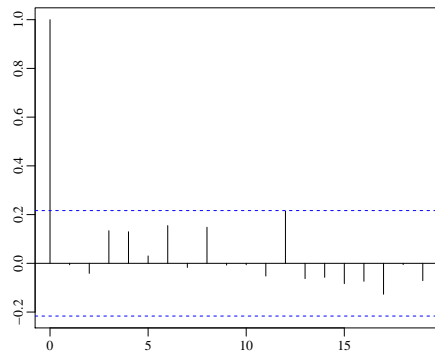
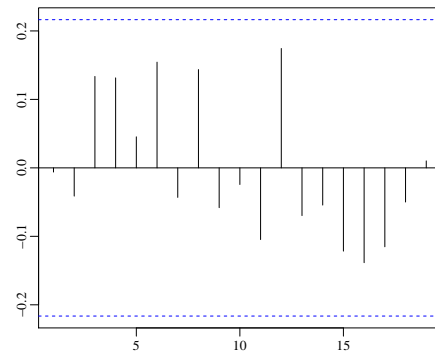
(c) correlograma dos resíduos do modelo β ARMA(d) correlograma parcial dos resíduos do modelo β ARMA

Figura 4.8 – Correlogramas dos resíduos padronizados dos modelos ARIMA e β ARMA da inadimplência de crédito

4.2 Modelando o crédito consignado

A disseminação dos empréstimos consignados em folha de pagamento (CC) ocorrida na última década causou uma reestruturação no mercado de crédito brasileiro. Além de reduzir significativamente a taxa de juros (RODRIGUES et al., 2006), o CC influenciou a redução do risco de inadimplência (SOUZA, 2007; SILVA, 2009) e expansão do volume de crédito no segmento pessoa física (JUNIOR, 2009). A promulgação da Lei 10.820, de 17 de dezembro de 2003, foi a principal responsável por estas transformações. Originada da Medida Provisória 130, de 17 de setembro de 2003, esta dispõe acerca do crédito com consignação em folha de pagamento e regulamenta sua oferta aos trabalhadores da iniciativa privada e aposentados ou pensionistas do Instituto Nacional do Seguro Social – INSS.

A despeito de ter menor rentabilidade em comparação com outras modalidades de crédito pessoal, o baixo custo operacional e menor risco de inadimplência tem aumentado a participação do CC na oferta total de crédito pessoa física. De acordo com Medeiros (2013), esta nova realidade também alterou a estrutura organizacional de muitas instituições financeiras, que passaram a desenvolver atividades especializadas no mercado de crédito consignado.

O índice utilizado neste estudo consiste na participação do crédito consignado sobre o crédito total no segmento pessoa física das instituições financeiras no Brasil - CC/Crédito PF. A série temporal foi calculada a partir de dados SGS. O período considerado abrange março de 2007 a agosto de 2014, sendo os últimos seis meses reservados para a análise da capacidade preditiva do modelo.

Segundo Takeda e Bader (2005), logo nos primeiros anos após a promulgação da Lei 10.820 houve um crescimento do CC em relação ao saldo das carteira de crédito. Este comportamento também é observado na Figura 4.9, onde a série possui um movimento de ascendência ao longo do tempo. A variável parte do patamar de 0,217 no início de 2007, tem valor médio em torno de 0,250 (Tabela 4.7) e segue evoluindo até alcançar seu máximo em agosto de 2014. Marisco e Fernandes (2012) apresentam uma possível justificativa para este comportamento. Segundo os autores, devido às baixas taxas de juros oferecidas, é comum que os consumidores utilizem o CC para liquidar dívidas antigas ou ainda para adquirir bens e serviços de forma mais rápida e menos burocrática.

Nota-se também que o único período de decaimento da série coincide com a crise financeira no final de 2008. De acordo com Martins (2011), tanto a queda da taxa de juros

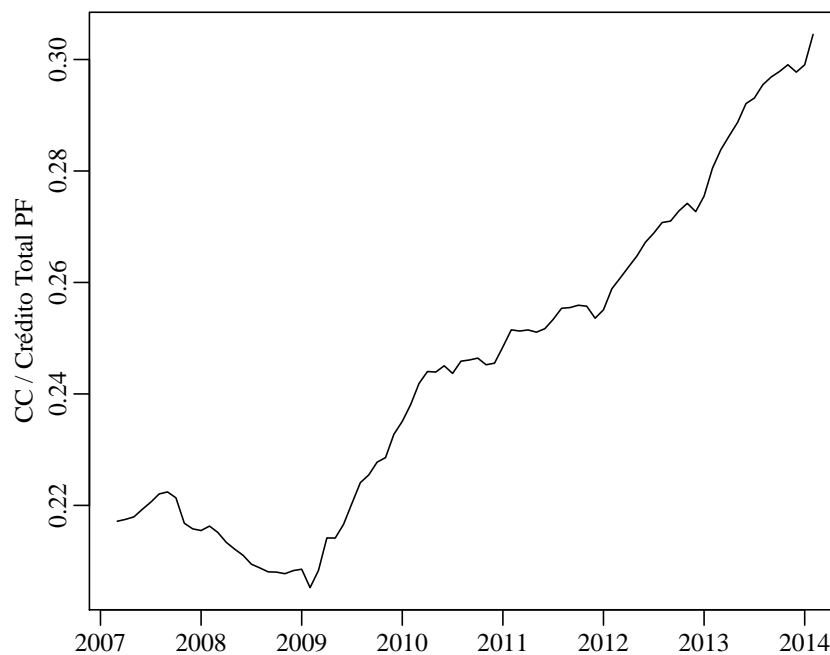


Figura 4.9 – Crédito consignado como proporção do crédito total PF no período de março de 2007 a agosto de 2014

americana quanto a quebra do Banco Lehman Brothers influenciaram a percepção de risco dos bancos brasileiros e provocaram um aumento no *spread* das operações de crédito realizadas. Desta forma, como o crédito consignado possui taxas menores, é provável que neste período os bancos tenham realocado suas carteiras para modalidades mais lucrativas.

De acordo com os valores apresentados na Tabela 4.7, o coeficiente de assimetria encontrado está próximo de zero, porém o resultado de 2,029 para a curtose é mais distante do esperado em variáveis que possuem distribuição normal. Este fato é corroborado pelo teste JB, o qual rejeita a hipótese nula de normalidade a 10% de significância. Neste caso, uma vez que o modelo ARIMA pressupõe normalidade, a utilização do mesmo fica comprometida. A Figura 4.10, que contém o histograma da relação CC/Crédito PF também permite verificar que o comportamento da série é bastante divergente do esperado quando a variável é normalmente distribuída.

Tabela 4.7 – Medidas descritivas para o crédito consignado como proporção do crédito total PF

Medidas	CC/Crédito PF
Média	0,250
Mediana	0,246
Desvio Padrão	0,033
Mínimo	0,205
Máximo	0,318
Assimetria	0,437
Curtose	2,073
Jarque-Bera	p -valor < 0,05

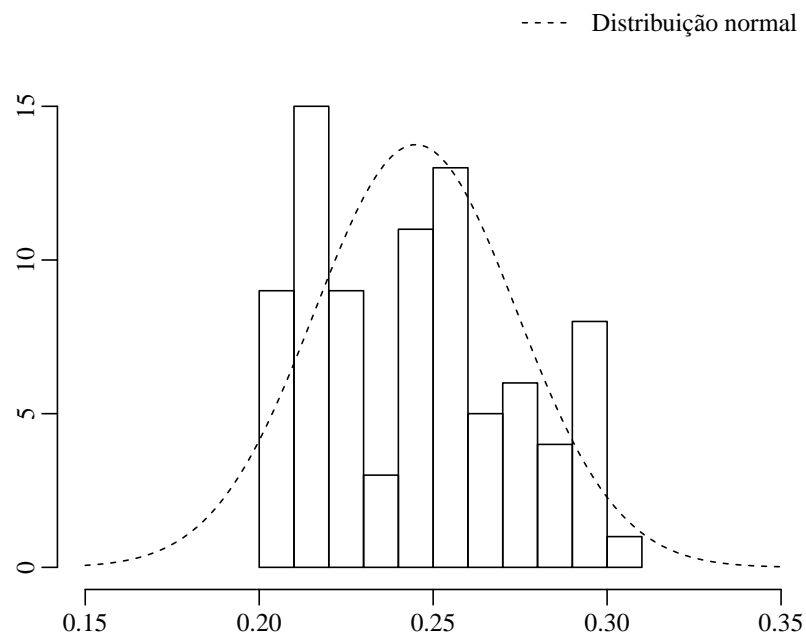


Figura 4.10 – Histograma e densidades estimadas do índice de CC/Crédito PF

Outro fator perceptível se refere ao grau de oscilação da mesma, que mantém-se ao redor de níveis médios em determinados períodos e depois se altera para outros níveis temporários. Conforme Morettin e Tolo (2006) este é um caso típico em séries econômicas, que comumente apresentam não estacionaridade homogênea. A Figura 4.11 é de grande auxílio neste sentido, pois ao verificar que no correlograma em nível há um decaimento da autocorrelação conforme as defasagens aumentam, tendo-se outro indício de ausência de estacionaridade. Além disso, o teste KPSS rejeita a hipótese nula de estacionaridade da série em nível, com estatística calculada de 2,91 para a série em nível e 0,45 para a estacionaridade em tendência. Na Figura 4.12 é apresentado o comportamento da série em primeira diferença.

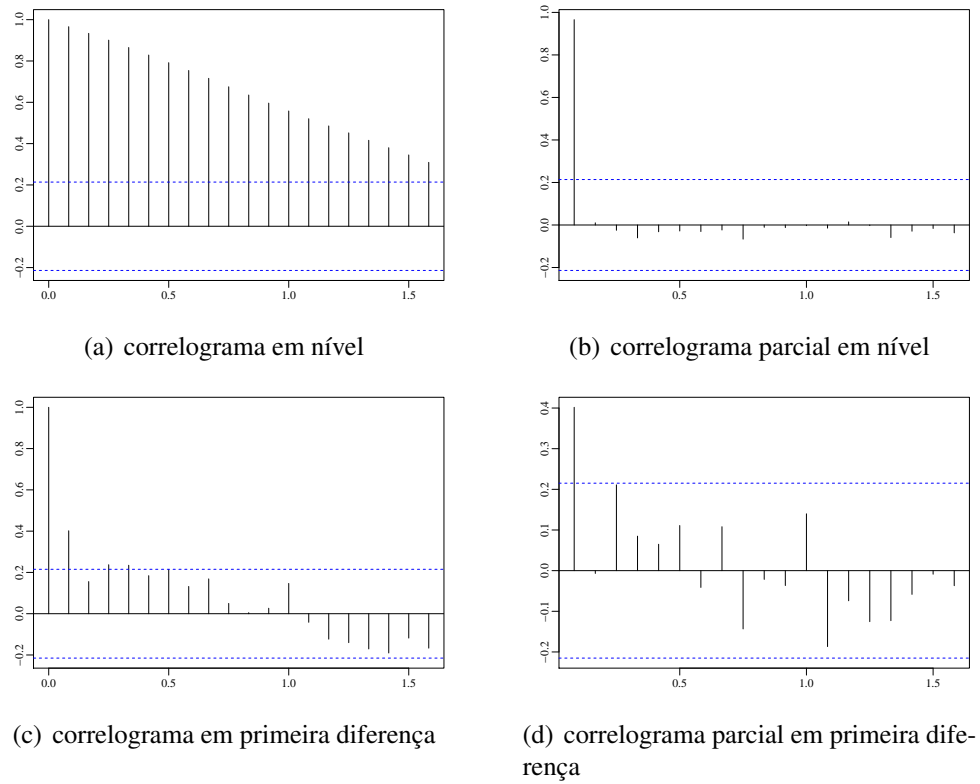


Figura 4.11 – Correlogramas do CC/Crédito PF em nível e em primeira diferença

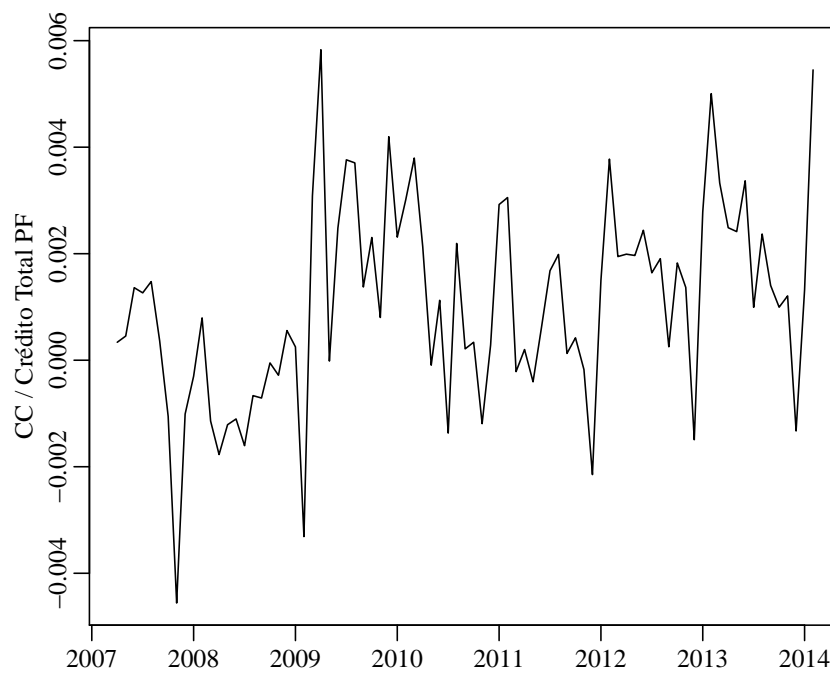


Figura 4.12 – Série de CC/Crédito PF em primeira diferença

Ao realizar a diferenciação da série temporal observa-se que nesta a variável deixa de apresentar o padrão típico de séries temporais não estacionárias (Figura 4.11). Já o teste KPSS corroborou a hipótese de estacionaridade em primeira diferença e com tendência determinística, a um nível de significância de 10%. Este resultado é condizente com o observado na Tabela 4.7, onde verificou-se que a série possui forte tendência crescente. Desta forma, assumimos que o processo gerador da série é integrado de ordem um, $I(1)$, com tendência e utilizamos a transformação para aplicação do β ARMA em intervalos limitados.

4.2.1 Modelagem e discussão

Assim como na Subseção 4.1.1, foram analisados os critérios AIC e BIC para definição das defasagens do modelo ARIMA e o BICc, HQ e HQc para a determinação da ordem de p e q do β ARMA. Também foi utilizada a técnica de superajustamento, levando em consideração a significância dos parâmetros estimados, o teste de Ljung e Box (1978), o RMSE e MAPE.

No caso do ARIMA, conforme evidenciado na Tabela 4.8, o modelo IMA(1,1) foi selecionado pois apresentou o melhor ajuste. Este é o único em que os parâmetros estimados são significativos. Além disso, o IMA(1,1) apresentou o segundo menor RMSE, que é superior apenas ao do IMA(1,10). O último, entretanto, além de utilizar uma quantidade demasiada de parâmetros, possui os piores resultados em termos de AIC e BIC. De acordo com o teste Ljung e Box, o IMA(1,1) não rejeita a hipótese nula de que os resíduos são ruído branco.

Tabela 4.8 – Critérios para identificação do modelo ARIMA para índice de CC/Crédito PF

Modelo	AIC	BIC	$Q[p\text{-valor}]$	RMSE	MAPE
IMA(1,1) ¹	-797,19	-770,59	0,0014[0,97]	0,001621	0,53495
IMA(1,2)	-799,15	-774,96	0,0023[0,96]	0,001622	0,53341
IMA(1,3)	-799,13	-777,36	0,0013[0,97]	0,001643	0,54735
IMA(1,4)	-800,99	-781,64	0,0002[0,99]	0,001646	0,54092
IMA(1,5)	-800,42	-783,48	0,0037[0,95]	0,001674	0,54061
IMA(1,6)	-802,40	-787,89	0,0032[0,96]	0,001674	0,54113
IMA(1,7)	-801,83	-789,73	0,0238[0,88]	0,001702	0,55208
IMA(1,8)	-803,40	-793,73	0,0071[0,93]	0,001706	0,55443
IMA(1,9)	-805,00	-797,74	0,1315[0,72]	0,001710	0,56031
IMA(1,10)	-795,82	-766,79	0,0002[0,99]	0,001605	0,52876

¹ Todas as estimativas significativas a pelo menos 1%

Para a modelagem β ARMA, os resultados expressos na Tabela 4.9 permitem observar que todos os modelos selecionados apresentaram valores bastante próximos em termos de RMSE e MAPE. Além disso, o teste Q de Ljung e Box (1978) não rejeitou a hipótese de que

os resíduos são ruído branco em nenhum dos casos. Tanto para o BICc, quanto para o HQ e HQc o modelo β MA(1) apresentou os melhores resultados, seguindo pelo β AR(1). Além disso, estes modelos foram os únicos que apresentaram significância em todas as estimativas dos parâmetros.

Tabela 4.9 – Critérios para identificação do modelo β ARMA para índice de CC/Crédito PF

Modelo	BICc	HQ	Hqc	$Q[p\text{-valor}]$	RMSE	MAPE
β AR(1) ¹	-910,99	-917,92	-917,16	0,0563[0,81]	0,000829	0,26171
β ARMA(1,1)	-906,96	-915,92	-914,76	0,0044[0,95]	0,000824	0,26594
β ARMA(1,2)	-890,20	-901,33	-899,68	0,0747[0,78]	0,000820	0,26055
β AR(3)	-878,97	-890,09	-888,45	0,0106[0,92]	0,000815	0,25741
β ARMA(2,1)	-889,38	-900,51	-898,86	0,0056[0,94]	0,000825	0,26425
β AR(2)	-894,15	-903,11	-901,95	0,0064[0,94]	0,000827	0,26357
β ARMA(2,2)	-885,04	-898,46	-896,25	0,0019[0,97]	0,000820	0,26078
β MA(3)	-876,55	-887,67	-886,03	0,0001[0,99]	0,000823	0,26233
β MA(2)	-893,64	-902,59	-901,44	0,0003[0,99]	0,000825	0,26487
βMA(1)¹	-911,04	-917,96	-917,19	0,0001[0,99]	0,000824	0,26695

¹ Todas as estimativas significativas a pelo menos 1%

Neste contexto, o modelo utilizado para delinear o comportamento da série é o β MA(1). As Tabelas 4.10 e 4.11 apresentam as estimativas dos modelos ARIMA e β ARMA selecionado, respectivamente. Observa-se a significância das estimativas da primeira defasagem do termo de média móvel. Já o termo \hat{t} representa o elemento de tendência linear identificado no teste KPSS.

Tabela 4.10 – Estimativas do modelo ARIMA selecionado para índice de CC/Crédito PF

	Estimativa	Desvio Padrão	$z[p\text{-valor}]$
$\hat{\theta}_1$	0,4236	0,0931	4,5495 [$<0,0001$]
\hat{t}	0,0011	0,0003	3,6732 [$<0,0001$]

Tabela 4.11 – Estimativas do modelo β ARMA selecionado para índice de CC/Crédito PF

	Estimativa	Desvio Padrão	$z[p\text{-valor}]$
$\hat{\alpha}$	-0,0003	0,0009	0,3622 [0,7172]
$\hat{\theta}_1$	1,5738	0,5212	3,0193 [0,0025]
\hat{t}	0,0001	0,0000	3,2464 [0,0012]

O procedimento utilizado para estimar a variável de tendência é descrito por Morettin e Toloi (2006) como tendência polinomial. Neste caso, a variável de tendência foi ajustada como um polinômio de primeiro grau em função do tempo, que assumirá os valores de 1, 2, 3, ..., n .

Tanto no modelo IMA(1) quanto no β MA(1) foram encontrados valores significativos e positivos para o coeficiente de tendência, o que vai de encontro com o observado no teste KPSS e na inspeção gráfica da série temporal em nível.

No que se refere a eficácia preditiva do modelo selecionado, a Figura 4.13 permite verificar que para ambos há um grau razoável de proximidade entre os dados originais e previstos tanto dentro (antes da linha vertical) quanto fora (após a linha vertical) do horizonte temporal amostrado.

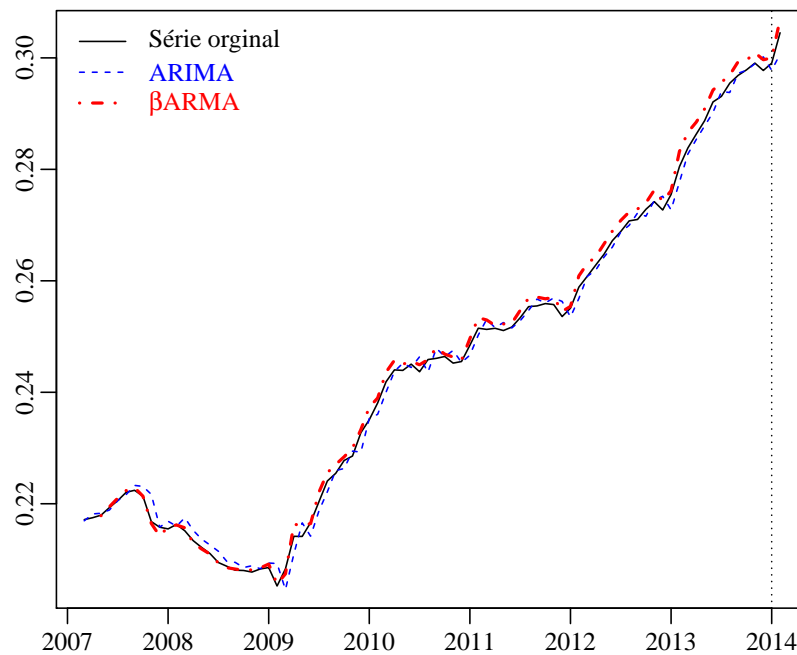


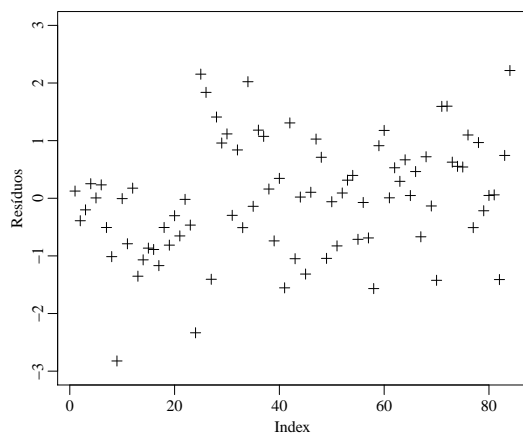
Figura 4.13 – Valores originais e previstos dentro e fora da amostra para a relação de CC/Crédito PF

Os resíduos padronizados dos modelos ajustados estão descritos na Figura 4.14. Nesta é possível observar que em ambos os casos os resíduos padronizados possuem comportamento aleatório próximo de zero e com todas ocorrências pertencentes ao intervalo $(-3, 3)$. Além disso, os correlogramas totais e parciais dos resíduos padronizados (Figura 4.15) corroboram a ideia de que ambos os modelos estão adequados em termos de análise de diagnóstico. Entretanto, ao comparar a capacidade preditiva dos modelos, fica evidente a superioridade do β ARMA selecionado. Além de apresentar resultados visivelmente mais próximos aos da série original (Figuras 4.13 e 4.14), o modelo β MA(1) possui menor MAPE e RMSE para as previsões fora da amostra (Tabela 4.12).

Desta forma, considerando todos os critérios mencionados, pode-se dizer que o ARIMA e o β ARMA, são capazes de representar adequadamente os movimentos do CC como proporção

Tabela 4.12 – Previsões para a relação CC/Crédito PF para meses fora da amostra

Período	Valor observado	Previsão ARIMA	Previsão β ARMA
mar/2014	0,3061	0,3083	0,3080
abr/2014	0,3087	0,3093	0,3103
mai/2014	0,3107	0,3104	0,3126
jun/2014	0,3137	0,3115	0,3150
jul/2014	0,3168	0,3126	0,3174
ago/2014	0,3182	0,3136	0,3198
MAPE	–	0,9089	0,4747
RMSE	–	0,0035	0,0010



(a) Resíduos padronizados do modelo ARIMA

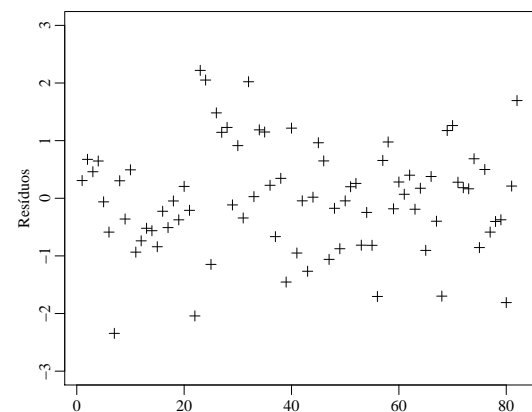
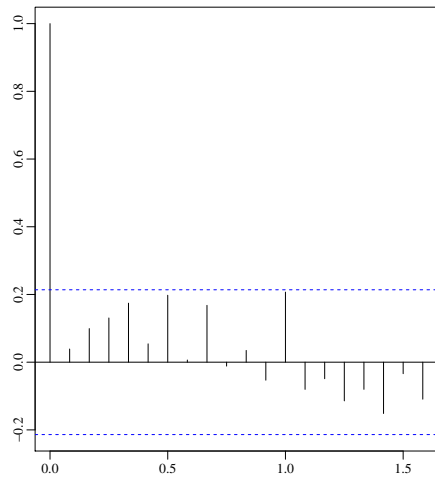
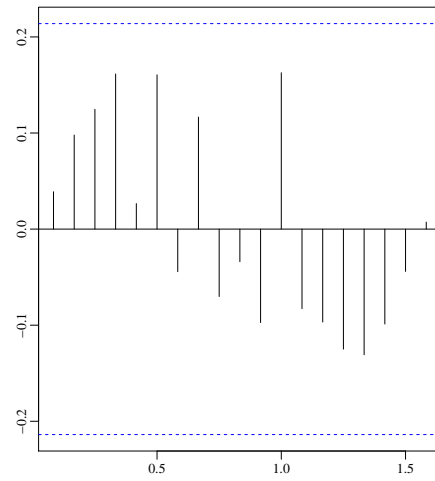
(b) Resíduos padronizados do modelo β ARMA

Figura 4.14 – Resíduos padronizados dos modelos para o CC/Crédito PF

do crédito total. Entretanto, é destacada a superioridade do modelo β ARMA selecionado. No que se refere aos critérios de seleção, vale ressaltar que o BICc, o HQ e o HQc apresentaram comportamento semelhante no que tange à identificação do melhor modelo candidato. Esse resultado vai de encontro com os obtidos nas simulações de Monte Carlo, onde o desempenho dos critérios foi bastante similar dentro de cada um dos cenários.



(a) correlograma dos resíduos do modelo ARIMA



(b) correlograma parcial dos resíduos do modelo ARIMA

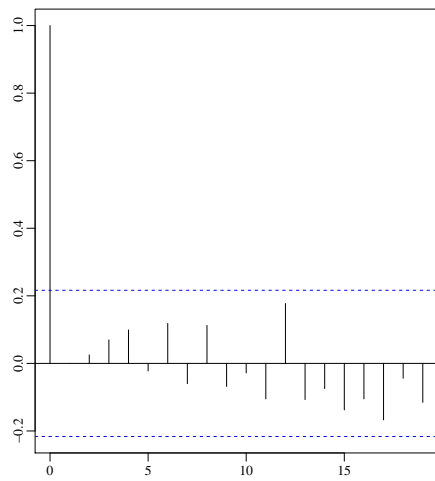
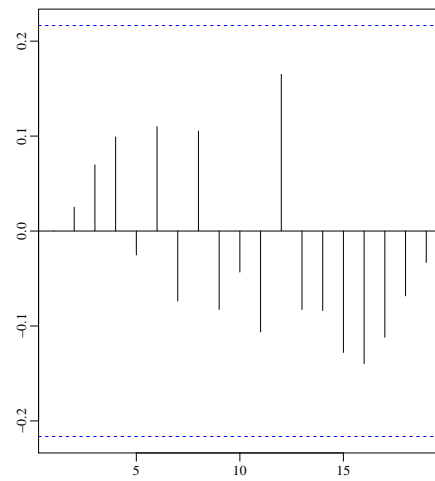
(c) correlograma dos resíduos do modelo β ARMA(d) correlograma parcial dos resíduos do modelo β ARMA

Figura 4.15 – Correlogramas dos resíduos padronizados dos modelos ARIMA e β ARMA do CC/Crédito PF

5 CONCLUSÃO

Os principais propósitos desta dissertação foram avaliar o desempenho de critérios de seleção para o β ARMA e apresentar aplicações deste em variáveis relacionadas ao crédito no SFN. No âmbito específico da modelagem e previsão de séries temporais restritas a intervalos limitados, a estimação de modelos ARIMA pode tornar-se inadequada pela suposição de normalidade. Neste contexto, o β ARMA seria uma alternativa mais apropriada, uma vez que assume que a variável de interesse tem distribuição beta.

Para avaliar o desempenho dos critérios de seleção na identificação do modelo verdadeiro foram realizadas simulações de Monte Carlo. Considerou-se cinco diferentes tamanhos amostrais e cinco tipos de cenários, os quais diferem no vetor de parâmetros AR e MA do modelo verdadeiro e na classe de modelos candidatos, variando também os valores do parâmetro de precisão. Os critérios de seleção analisados melhoraram seu desempenho à medida que o tamanho da amostra aumenta, o que era esperado em virtude das propriedades de consistência e eficiência assintótica. Constatou-se também que não houve alteração no grau de identificação correta dos critérios em função da modificação no valor de ϕ . Além disso, o desempenho de todos os critérios decaiu conforme aumentava a complexidade do modelo verdadeiro e da classe de modelos candidatos. Os modelos β MA e β ARMA obtiveram menor percentual de identificação que o β AR.

Entre os critérios analisados, o AIC, AICc e o BIC foram os que apresentaram as maiores divergências de resultados ao comparar os diferentes cenários. Os demais critérios apresentaram resultados semelhantes entre si, onde o BICc apresentou desempenho ligeiramente superior ao HQ e HQc nos Cenários 1 e 4. Por outro lado, nos demais cenários o HQ e o HQc obtiveram um percentual de acertos levemente superior ao BICc. Devido à semelhança nos bons resultados entre estes, indica-se a utilização dos três critérios para a seleção de modelos β ARMA. Esses três critérios, nomeadamente o BICc, o HQ e o HQc, foram os critérios utilizados neste estudo para a seleção dos modelos β ARMA nas aplicações realizadas. As variáveis selecionadas para aplicação do modelo β ARMA e dos critérios de seleção foram a inadimplência de crédito e a relação entre o CC e o crédito PF, ambas relacionadas ao mercado de crédito do SFN.

No intuito de realizar comparações, foram ajustados e realizadas previsões através da estimação tanto de modelos ARIMA quanto de modelos β ARMA. Para ambas as variáveis o primeiro indício da inadequação da modelagem ARIMA ocorreu já na análise descritiva dos

dados. Já no processo de identificação do modelo para a inadimplência de crédito e a relação CC/Crédito PF, apesar de o ARIMA e o β ARMA captarem adequadamente os movimentos da série, o segundo apresentou previsões mais próximas aos valores originais tanto para os valores ajustados dentro da amostra, quanto para os resultados obtidos fora do horizonte temporal amostrado. Este resultado condiz com o esperado, uma vez que o β ARMA é um modelo específico para séries temporais contínuas restritas ao intervalo unitário padrão.

A análise gráfica dos resíduos dos modelos selecionados também evidencia a superioridade do β ARMA em descrever o comportamento da inadimplência de crédito e do CC/Crédito PF. Com isso, devido ao fato de não haverem outras aplicações do β ARMA na áreas de engenharia de produção e econometria, os resultados obtidos são um primeiro indício empírico de que o β ARMA pode efetivamente causar um incremento positivo tanto em termos de qualidade do ajuste quanto acerca dos resultados para previsões de séries temporais de proporções ou taxas restritas a intervalos limitados.

Além disso, ao tratar dos critérios de seleção no âmbito de aplicações do β ARMA, ressalta-se que os três critérios de melhor desempenho nas simulações de Monte Carlo também foram capazes de identificar modelos de boa qualidade em termos de significância das estimativas e análise de resíduos. Assim como nas simulações de Monte Carlo, o BICc, o HQ e o HQc também mantiveram um comportamento semelhante entre si no contexto da seleção de modelos β ARMA para a inadimplência de crédito e para o CC/Crédito PF.

REFERÊNCIAS

- ABID, L.; OUERTANI, M. N.; ZOUARI-GHORBEL, S. Macroeconomic and Bank-specific Determinants of Household's Non-performing Loans in Tunisia: a dynamic panel data. **Procedia Economics and Finance**, v.13, n.1, p.58 – 68, 2014.
- ACQUAH, H. d.-G. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. **Journal of Development and Agricultural Economics**, v.2, n.1, p.01–06, 2010.
- AKAIKE, H. A. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v.19, n.6, p.716–723, 1974.
- AKAIKE, H. A. Bayesian analysis of the minimum AIC procedure. **Annals of the Institute of Statistical Mathematics A**, v.30, n.1, p.9–14, 1978.
- ALMEIDA, F. D. **Determinantes do spread bancário ex-post no Brasil: uma análise de fatores micro e macroeconômicos**. 2013. Dissertação (Mestrado em Ciência da Computação) — Universidade Católica de Brasília.
- ALMEIDA, P. M.; SOUZA, T. C. d. Estimativas de votos de Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do bolsa família. **Ciência e Natura**, v.37, n.1, p.12–22, 2015.
- AMISANO, G.; CASARIN, R. Particle filters for Markov-Switching Stochastic Correlation Models. In: ITALIAN STATISTICAL SOCIETY CONFERENCE. INTERMEDIATE CONFERENCE, RISK AND PREDICTION. **Anais...** 2007.
- ANNIBAL, C. A. **Inadimplência do setor bancário brasileiro: uma avaliação de suas medidas**. Brasília: Departamento de estudos e pesquisas Banco Central do Brasil, 2009. Trabalho para discussão. (192).
- BABAI, M. et al. Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: theory and empirical analysis. **International Journal of Production Economics**, v.143, n.2, p.463 – 471, 2013.
- BAILLIE, R. T.; KAPETANIOS, G.; PAPAILIAS, F. Modified information criteria and selection of long memory time series models. **Computational Statistics and Data Analysis**, v.76, n.1, p.116–131, 2014.

- BARSEGHYAN, L. Non-performing loans, prospective bailouts, and Japan's slowdown. **Journal of Monetary Economics**, v.57, n.7, p.873 – 890, 2010.
- BAYER, F. M. **Modelagem e inferência em regressão beta**. 2011. Tese (Doutorado em Estatística) — Universidade Federal de Pernambuco.
- BAYER, F. M.; CRIBARI-NETO, F. Bartlett corrections in beta regression models. **Journal of Statistical Planning and Inference (Print)**, v.143, p.531–547, 2013.
- BAYER, F. M.; CRIBARI-NETO, F. Model selection criteria in beta regression with varying dispersion. **Communications in Statistics - Simulation and Computation**, 2015.
- BAYER, F. M.; CRIBARI-NETO, F. Bootstrap-based model selection criteria for beta regressions. <http://arxiv.org/abs/1405.4525>, 2015.
- BCBS. **International Convergence of Capital Measurement and Capital Standards: a revised framework**. [S.l.]: Bank for International Settlements, 2004.
- BENGTSSON, T.; CAVANAUGH, J. An improved Akaike information criterion for state-space model selection. **Computational Statistics and Data Analysis**, v.50, n.10, p.2635–2654, 2006.
- BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. **Journal of the American Statistical Association**, v.98, n.461, p.214–223, 2003.
- BILLIO, M.; CASARIN, R. Beta Autoregressive Transition Markov-switching Models for Business Cycle Analysis. **Studies in Nonlinear Dynamics and Econometrics**, v.15, n.4, p.1–32, 2011.
- BISWAS, R.; BHATTACHARYYA, B. ARIMA modeling to forecast area and production of rice in West Bengal. **Journal of Crop and Weed**, v.9, n.2, p.26–31, 2013.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: forecasting and control**. San Francisco: Holden Day, 1976.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: forecasting and control**. 4.ed. New Jersey: John Wiley & Sons, 2008. 784p.

- BURNHAN, K. P.; ANDERSON, D. R. Multimodel inference: understanding AIC and BIC in model selection. **Sociological Methods and Research**, v.33, n.2, p.261–304, 2004.
- CASARIN, R.; DALLA VALLE, L.; LEINSEN, F. Bayesian model selection for beta autoregressive processes. **Bayesian Analysis**, v.7, p.385–409, 2012.
- CASTRO, V. Macroeconomic determinants of the credit risk in the banking system: the case of the GIPSI. **Economic Modelling**, v.31, n.1, p.672 – 683, 2013.
- CAVANAUGH, J. E.; SHUMWAY, R. H. A Bootstrap Variant of AIC for State-Space Model Selection. **Statistica Sinica**, v.7, n.2, p.473–496, 1997.
- CHAIBI, H.; FTITI, Z. Credit risk determinants: evidence from a cross-country study. **Research in International Business and Finance**, v.33, n.1, p.1 – 16, 2015.
- CIFTER, A.; YILMAZER, S.; CIFTER, E. Analysis of sectoral credit default cycle dependency with wavelet networks: evidence from turkey. **Economic Modelling**, v.26, n.6, p.1382 – 1388, 2009.
- CRIBARI-NETO, F.; SOUZA, T. C. Religious belief and intelligence: worldwide evidence. **Intelligence**, v.41, n.5, p.482–489., 2013.
- CRIBARI-NETO, F.; ZEILEIS, A. Beta Regression in R. **Journal of Statistical Software**, v.34, n.2, p.1–24, 2010.
- CROOK, J.; BANASIK, J. Forecasting and explaining aggregate consumer credit delinquency behaviour. **International Journal of Forecasting**, v.28, n.1, p.145 – 160, 2012.
- DAVIES, S. L.; NEATH, A. A.; CAVANAUGH, J. E. Cross validation model selection criteria for linear regression based on the Kullback–Leibler discrepancy. **Statistical Methodology**, v.2, n.4, p.249 – 266, 2005.
- DE LUNA, X. Prediction Intervals Based on Autoregression Forecasts. **Journal of the Royal Statistical Society: Series D (The Statistician)**, v.49, n.1, p.87–93, 2000.
- EDIGER, V. S.; AKAR, S. ARIMA forecasting of primary energy demand by fuel in Turkey. **Energy Policy**, v.35, n.3, p.1701 – 1708, 2007.

EMILIANO, P. C. **Fundamentos e aplicações dos critérios de informação: Akaike e Schwarz**. 2009. Dissertação (Mestrado em Estatística e Experimentação Agrícola) — Universidade Federal de Lavras.

EMILIANO, P. C. et al. Foundations and comparison of information criteria: Akaike and bayesian. **Revista Brasileira de Biometria**, v.27, n.3, p.394–411, 2009.

EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S. d. Information criteria: how do they behave in different models? **Computational Statistics and Data Analysis**, v.69, n.1, p.141–153, 2014.

ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. Influence diagnostics in beta regression. **Computational Statistics and Data Analysis**, v.52, n.9, p.4417–4431, 2008.

ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On beta regression residuals. **Journal of Applied Statistics**, v.35, n.4, p.407–419, 2008.

FAN, Y.; TANG, C. Y. Tuning parameter selection in high dimensional penalized likelihood. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v.75, n.3, p.531–552, 2013.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v.31, n.7, p.799–815, 2004.

GIAMPAOLI, V.; RAPOSO, M. C. F.; TORRES, S. T. F. Critérios de seleção de modelos para o modelo de regressão beta. **Revista brasileira de Estatística**, v.70, n.232, p.7–28, 2009.

GRANGER, C. W. J.; NEWBOLD, P. **Forecasting Economic Time Series**. London: Academic Press, 1986.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5.ed. Porto Alegre: AMGH, 2011. 924p.

GUOLO, A.; VARIN, C. Beta Regression for time series analysis of bounded data, with application to canada google flu trend. **The Annals of Applied Statistics**, v.8, n.1, p.74–88, 2014.

HANNAN, E. J.; QUINN, B. G. The determination of the order of an autoregression. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v.41, n.2, p.190–195, 1979.

- HERWARTZ, H. **A note on model selection in (time series) regression models - general-to-specific or specific-to-general?** [S.l.]: Christian Albrechts Universitat kiel, 2007. Economics Working Paper. (09).
- HOSKING, J. R. M. Fractional differencing. **Biometrika**, v.68, n.1, p.165–176, 1981.
- HOTTA, L. K.; NETO, J. C. The effect of aggregation on prediction in autorregressive integrated moving average models. **Journal of Time Series Analysis**, v.14, n.3, p.261–269, 1993.
- HU, B.; SHAO, J. Generalized linear model selection using. **Journal of Statistical Planning and Inference**, v.138, n.12, p.3705 – 3712, 2008.
- HURVICH, C. M.; TSAI, C.-L. Regression and time series model selection in small samples. **Biometrika**, v.76, n.2, p.297–307, 1989.
- ISHIGURO, M.; SAKAMOTO, Y.; KITAGAWA, G. Bootstrapping log likelihood and EIC, an extension of AIC. **Annals of the Institute of Statistical Mathematics**, v.49, n.3, p.411–434, 1997.
- JARQUE, C.; BERA, A. A test for normality of observations and regression residuals. **International Statistical Review.**, v.2, n.55, p.163–172, 1987.
- JUNIOR, R. B. Visão geral da produção de estatísticas monetárias e de crédito no Banco Central do Brasil. **São Paulo Perspectiva**, v.23, n.2, p.93–99, 2009.
- KIM, J. Y. et al. Forecasting the monthly abundance of anchovies in the South Sea of Korea using a univariate approach. **Fisheries Research**, v.161, n.1-3, p.293 – 302, 2015.
- KLETTING, P.; GLATTING, G. Model selection for time-activity curves: the corrected akaike information criterion and the f-test. **Zeitschrift für Medizinische Physik**, v.19, n.3, p.200 – 206, 2009.
- KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, v.54, p.159–178,, 1992.
- LEVINE, R. Financial Development and Economic Growth: views and agenda. **Journal of Economic Literature**, v.35, n.2, p.688–726, 1997.

- LINARDI, F. M. **Avaliação dos determinantes macroeconômicos da inadimplência bancária no Brasil**. 2008. Dissertação (Mestrado em Economia) — Universidade Federal de Minas Gerais.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v.65, n.2, p.297–303, 1978.
- LOUZIS, D. P.; VOULDIS, A. T.; METAXAS, V. L. Macroeconomic and bank-specific determinants of non-performing loans in Greece: a comparative study of mortgage, business and consumer loan portfolios. **Journal of Banking & Finance**, v.36, n.4, p.1012 – 1027, 2012.
- LV, J.; LIU, J. S. Model selection principles in misspecified models. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v.76, n.1, p.141–167, 2014.
- MACDONALD, I. L.; LERER, L. B. A Time-Series Analysis of Trends in Firearm-Related Homicide and Suicide. **International Journal of Epidemiology**, v.23, n.1, p.66–72, 1994.
- MAIA, J. P. **Aplicação de redes neurais na previsão de demanda de crédito no sistema financeiro nacional**. 2008. Mestrado (Engenharia Elétrica) — Faculdade de tecnologia da Universidade de Brasília.
- MANHIÇA, F. A.; JORGE, C. T. **O nível da taxa básica de juros e o spread bancário no Brasil: uma análise de dados em painel**. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, 2012. Texto para discussão.
- MARHUENDA, Y.; MORALES, D.; PARDO, M. d. C. Information criteria for Fay–Herriot model selection. **Computational Statistics and Data Analysis**, v.70, n.1, p.268–280, 2014.
- MARINS, J. T. M.; NEVES, M. B. E. d. **Inadimplência de crédito e ciclo econômico: um exame da relação no mercado brasileiro de crédito corporativo**. Brasília: Departamento de estudos e pesquisas do Banco Central do Brasil, 2013. Trabalhos para discussão.
- MARISCO, F. M.; FERNANDES, R. A responsabilidade civil na concessão do empréstimo consignado: uma análise do superendividamento do servidor público no estado de Rondônia. **Revista de Direito Público**, v.7, n.2, p.157–180, 2012.
- MARTINS, B. **Composição, evolução e influência da política econômica no spread bancário praticado no Brasil: 2003 – 2010**. 2011. Monografia (Graduação em Ciências Econômicas) — Universidade Federal do Rio Grande do Sul – UFRGS.

MCCULLAGH, P.; NELDER, J. **Generalized linear models**. 2.ed. London: Chapman and Hall, 1989.

MCQUARRIE, A. D. A small-sample correction for the Schwarz SIC model selection criterion. **Statistics & Probability Letters**, v.44, n.1, p.79–86, 1999.

MCQUARRIE, A. D.; SHUMWAY, R.; TSAI, C.-L. The model selection criterion AICu. **Statistics & Probability Letters**, v.34, n.3, p.285–292, 1997.

MCQUARRIE, A. D.; TSAI, C.-L. **Regression and time series model selection**. Singapore: World Scientific, 1998. 455p.

MEDEIROS, D. A. **Financeirização do território e circuitos da economia urbana: agentes de crédito, técnicas e normas bancárias. um exemplo em alagoas**. 2013. Dissertação (Mestrado em Geografia Humana) — Universidade de São Paulo.

MEEKER, L. G.; GRAY, L. A note on non-performing loans as an indicator of asset quality. **Journal of Banking & Finance**, v.11, n.1, p.161 – 168, 1987.

MENDONÇA, M. J.; SACHSIDA, A. **Identificando a demanda e a oferta de crédito bancário no Brasil**. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, 2013. Texto para discussão.

MITTELHAMMER, R. C.; JUDGE, G. G.; MILLER, D. J. **Econometric Foundations**. New York: Cambridge University Press, 2000.

MONDAL, P.; SHIT, L.; GOSWAM, S. Study of effectiveness of time series modeling (ARIMA in forecasting stock prices. **International Journal of Computer Science, Engineering and Applications (IJCSEA)**, v.4, n.2, p.13–29, 2014.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Blucher, 2006. 538p.

NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. New York: Springer, 1999.

OSPINA, R.; CRIBARI-NETO, F.; VASCONCELOS KLAUS, L. P. Improved point and interval estimation for a beta regression model. **Computational Statistics and Data Analysis**, v.51, n.2, p.960–981, 2006.

OSPINA, R.; FERRARI, S. L. P. A general class of zero-or-one inflated beta regression models. **Pre- print**, 2011.

PAN, W. Bootstrapping likelihood for model selection with small samples. **Journal of Computational and Graphical Statistics**, v.8, n.4, p.687–698, 1999.

PAN, W. Akaike's information criterion in generalized estimating equations. **Biometrics**, v.57, n.1, p.120–125, 2001.

PEREIRA, T. L. **Regressão beta inflacionada: inferência e aplicações**. 2010. Tese de Doutorado em estatística — Universidade Federal de Pernambuco.

POKORNY, M. **An introduction to econometrics**. 1.ed. New York: Basil Blackwell, 1987. 343p.

PRESS, W. H. et al. **Numerical Recipes in C: the art of scientific computing**. 3.ed. New York: Cambridge University Press, 2007. 1256p.

R Core Team. **R: a language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2013.

RAMOS, P.; SANTOS, N.; REBELO, R. Performance of state space and ARIMA models for consumer retail sales forecasting. **Robotics and Computer-Integrated Manufacturing**, 2015. Aceito.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **Test**, v.18, n.3, p.529–545, 2009.

RODRIGUES, E. et al. **O efeito da Consignação em folha nas Taxas de Juros dos Empréstimos Pessoais**. Brasília: Departamento de estudos e pesquisas Banco Central do Brasil, 2006. Trabalhos para discussão. (108).

SAITO, A. T. **Fatores determinantes da disponibilidade de crédito nos países: uma análise de 2004 a 2010**. 2012. Tese (Doutorado em Administração) — Universidade de São Paulo.

SAITO, A. T.; SAVOIA, J. R. F.; LAZIER, I. Determinants of private credit in OECD developed, BRICs and LAC Countries. **American International Journal of Social Science**, v.3, n.3, p.21–35, 2014.

- SANTOS, J. O. d. Análise comparativa de métodos para previsão de insolvência em uma carteira de crédito bancário em empresas de médio porte. **Revista de Gestão USP**, v.15, n.3, p.11–24, jul/set 2008.
- SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v.6, n.2, p.461–464, 1978.
- SEGHOUANE, A.-K. Asymptotic bootstrap corrections of AIC for linear regression models. **Signal Processing**, v.90, n.1, p.217–224, 2010.
- SHANG, J.; CAVANAUGH, J. Bootstrap variants of the Akaike information criterion for mixed model selection. **Computational Statistics and Data Analysis**, v.52, n.4, p.2004–2021, 2008.
- SHI, P.; TSAI, C.-L. Regression model selection—a residual likelihood approach. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v.64, n.2, p.237–252, 2002.
- SHIBATA, R. An optimal selection of regression variables. **Biometrika**, v.68, n.1, p.45–54, 1981.
- SHIBATA, R. Bootstrap estimate of Kullback-Leibler information for model selection. **Statistica Sinica**, v.7, n.2, p.375–394, 1997.
- SILVA, C. R.; SOUZA, T. C. d. Modelagem da taxa de analfabetismo no estado da Paraíba via modelo de regressão beta. **Revista Brasileira de Biometria**, v.32, n.3, p.345–359, 2014.
- SILVA, G. J. C. d. **Ensaio em teoria da firma bancária: estratégia de localização, dinâmica do spread bancário e suas implicações em um modelo baseado em agentes**. 2009. Tese (Doutorado em Economia) — Universidade Federal de Minas Gerais.
- SIMAS, A. B.; SOUZA, W. B.; ROCHA, A. V. Improved estimators for a general class of beta regression models. **Computational Statistics & Data Analysis**, v.54, n.2, 2010.
- SINIKSARAN, E. A geometric interpretation of Mallows' C_p statistic and an alternative plot in variable selection. **Computational Statistics and Data Analysis**, v.52, n.7, p.3459 – 3467, 2008.
- SOUZA, G. J. d. G. **A Interação entre a dinâmica macroeconômica e os bancos: uma perspectiva acerca do risco de crédito**. 2007. Dissertação (Mestrado em Ciências Econômicas) — Universidade Federal Fluminense.

SOUZA, T. C. d.; CRIBARI-NETO, F. Uma estimativa do impacto eleitoral do programa Bolsa-Família. **Revista Brasileira de Biometria**, v.31, n.1, p.79–103, 2013.

STONE, M. Comments on Model Selection Criteria of Akaike and Schwarz. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v.41, n.2, p.276–278, 1979.

SUGIURA, N. Further analysts of the data by Akaike's information criterion and the finite corrections - Further analysts of the data by Akaike's. **Communications in Statistics - Theory and Methods**, v.7, n.1, p.13–26, 1978.

TABAK, B. M.; CRAVEIRO, G. L.; CAJUEIRO, D. O. **Eficiência bancária e inadimplência: testes de causalidade**. Brasília: Departamento de estudos e pesquisas Banco Central do Brasil, 2010. Trabalho para discussão. (220).

TAKEDA, T.; BADER, F. L. C. **Consignação em folha de pagamento – fatores da Impulsão do crédito**. Brasília: Departamento de Estudos e Pesquisas Banco Central do Brasil, 2005. Relatório de Economia Bancária e Crédito.

TORRES, S. T. F. **Avaliação de critérios de seleção de modelos para o modelo de regressão beta**. 2005. Dissertação (Mestrado em Estatística) — Universidade Federal de Pernambuco.

VAZQUEZ, F.; TABAK, B. M.; SOUTO, M. A macro stress test model of credit risk for the Brazilian banking sector. **Journal of Financial Stability**, v.8, n.2, p.69 – 83, 2012.

VICENTE, M. R.; LOPEZ-MENENDEZ, A. J.; PEREZ, R. Forecasting unemployment with internet search data: does it help to improve predictions when job destruction is skyrocketing? **Technological Forecasting and Social Change**, v.92, n.1, p.132 – 139, 2015.

WHEELWRIGHT, S. C.; MAKRIDAKIS, S. **Forecasting Methods for Management**. 4.ed. New York: John Wiley & Sons, 1985.

YANAKA, G. M.; HOLLAND, M. Basiléia II e Exigência de Capital para Risco de Crédito dos Bancos no Brasil. **Revista Brasileira de Finanças**, v.8, n.2, p.167–195, 2010.

ZANIBONI, N. C. **A inadimplência do sistema financeiro no Brasil explicada por meio de fatores macroeconômicos**. 2013. Dissertação (Mestrado em Administração) — Universidade de São Paulo.

ÇETIN, M. C.; ERAR, A. Variable Selection with Akaike Information Criteria: a comparative study. **Hacettepe Journal of Mathematics and Statistics**, v.31, n.1, p.89–97, 2002.

APÊNDICES

APÊNDICE A – Tabelas de frequência (em %) de identificação dos modelos

Nesta seção encontram-se discriminadas as proporções em que cada critério selecionou um modelo com ordem de defasagens k igual à do modelo verdadeiro (Cenários 4 e 5), sobreidentificou e subidentificou o modelo verdadeiro.

Tabela A.1 – Frequência (em %) de identificação no Cenário 1 com $\phi = 20$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 2$	46,06	25,77	21,13	8,23	36,56	17,97
$k = 2$	44,73	58,67	58,04	61,00	50,91	61,56
$k < 2$	9,21	15,56	20,83	30,77	12,53	20,47
$n = 50$						
$k > 2$	41,48	30,73	13,79	7,71	27,81	18,19
$k = 2$	55,53	65,07	76,56	79,17	66,99	74,69
$k < 2$	2,99	4,20	9,65	13,12	5,20	7,12
$n = 80$						
$k > 2$	39,05	33,00	9,89	6,86	22,67	17,22
$k = 2$	60,45	66,38	87,29	89,65	76,13	81,23
$k < 2$	0,50	0,62	2,82	3,49	1,20	1,55
$n = 120$						
$k > 2$	37,87	33,95	7,16	5,46	19,48	15,99
$k = 2$	62,09	66,01	92,40	94,01	80,34	83,81
$k < 2$	0,04	0,04	0,44	0,53	0,18	0,20
$n = 200$						
$k > 2$	36,28	33,87	5,37	4,60	16,31	14,59
$k = 2$	63,72	66,13	94,63	95,40	83,69	85,41
$k < 2$	0,00	0,00	0,00	0,00	0,00	0,00

Tabela A.2 – Frequência (em %) de identificação no Cenário 2 com $\phi = 20$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 2$	61,40	37,77	30,49	10,47	50,69	25,89
$k = 2$	11,72	15,75	13,83	12,90	12,89	15,78
$k < 2$	26,88	46,48	55,68	76,63	36,42	58,33
$n = 50$						
$k > 2$	44,29	32,77	13,02	5,98	29,08	18,30
$k = 2$	17,11	18,93	15,07	12,89	17,60	17,96
$k < 2$	38,60	48,30	71,91	81,13	53,32	63,74
$n = 80$						
$k > 2$	35,56	28,64	6,09	3,84	18,11	13,08
$k = 2$	22,40	23,39	15,94	14,24	21,67	21,14
$k < 2$	42,04	47,97	77,97	81,92	60,22	65,78
$n = 120$						
$k > 2$	33,53	29,38	4,14	3,13	14,69	11,80
$k = 2$	26,89	27,75	17,76	16,28	25,86	25,07
$k < 2$	39,58	42,87	78,10	80,59	59,45	63,13
$n = 200$						
$k > 2$	32,41	29,78	2,36	1,94	12,11	10,55
$k = 2$	34,70	35,45	22,63	21,41	33,49	32,96
$k < 2$	32,89	34,77	75,01	76,65	54,40	56,49

Tabela A.3 – Frequência (em %) de identificação no Cenário 4 com $\phi = 20$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 2$	95,49	55,65	79,91	19,22	75,18	41,02
$k = 2$	3,14	25,96	12,76	35,45	14,10	31,71
$k < 2$	1,37	18,39	7,33	45,33	10,72	27,27
modelo verdadeiro	0,49	22,12	1,83	31,38	11,63	27,25
$n = 50$						
$k > 2$	95,22	60,67	77,36	17,41	58,37	39,04
$k = 2$	3,57	29,66	14,43	50,70	30,11	43,14
$k < 2$	1,21	9,67	8,21	31,89	11,52	17,82
modelo verdadeiro	1,19	27,60	3,57	48,27	28,18	40,72
$n = 80$						
$k > 2$	95,80	59,74	80,35	13,74	43,64	32,98
$k = 2$	3,57	36,02	13,37	66,30	48,23	56,58
$k < 2$	0,63	4,24	6,28	19,96	8,13	10,44
modelo verdadeiro	2,33	35,06	6,08	65,30	47,20	55,50
$n = 120$						
$k > 2$	95,36	56,01	84,21	10,23	34,43	28,78
$k = 2$	4,30	42,26	12,36	79,74	61,53	66,67
$k < 2$	0,34	1,73	3,43	10,03	4,04	4,55
modelo verdadeiro	3,74	42,00	8,69	79,35	61,23	66,33
$n = 200$						
$k > 2$	94,82	51,65	87,72	6,87	25,27	22,47
$k = 2$	5,15	48,16	11,71	91,00	74,03	76,75
$k < 2$	0,03	0,19	0,57	2,13	0,70	0,78
modelo verdadeiro	5,07	48,14	11,18	90,91	74,00	76,70

Tabela A.4 – Frequência (em %) de identificação no Cenário 5 com $\phi = 20$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
<i>n</i> = 30						
<i>k</i> > 2	91,09	49,93	61,30	13,80	69,99	34,39
<i>k</i> = 2	7,77	17,11	32,03	17,65	10,93	19,26
<i>k</i> < 2	1,14	32,96	6,67	68,55	19,08	46,35
modelo verdadeiro	0,63	7,52	1,51	8,64	4,50	8,74
<i>n</i> = 50						
<i>k</i> > 2	86,31	53,11	44,14	9,75	49,85	29,19
<i>k</i> = 2	12,54	16,98	49,27	17,35	15,87	20,58
<i>k</i> < 2	1,15	29,91	6,59	72,90	34,28	50,23
modelo verdadeiro	0,31	6,63	0,77	7,98	6,44	8,74
<i>n</i> = 80						
<i>k</i> > 2	80,51	51,21	32,59	6,76	33,43	23,00
<i>k</i> = 2	17,95	19,80	60,75	18,24	21,88	23,65
<i>k</i> < 2	1,54	28,99	6,66	75,00	44,69	53,35
modelo verdadeiro	0,43	8,02	0,73	8,58	9,37	10,38
<i>n</i> = 120						
<i>k</i> > 2	76,31	47,83	27,06	4,59	24,61	18,70
<i>k</i> = 2	21,92	24,66	67,09	20,12	27,93	28,74
<i>k</i> < 2	1,77	27,51	5,85	75,29	47,46	52,56
modelo verdadeiro	0,61	10,27	0,78	10,00	12,63	13,08
<i>n</i> = 200						
<i>k</i> > 2	74,56	46,27	26,33	2,70	18,55	15,89
<i>k</i> = 2	24,23	30,53	69,42	25,46	35,83	35,88
<i>k</i> < 2	1,21	23,20	4,25	71,84	45,62	48,23
modelo verdadeiro	0,76	14,03	0,91	13,08	17,27	17,44

Tabela A.5 – Frequência (em %) de identificação no Cenário 1 com $\phi = 120$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 2$	44,90	26,77	21,77	8,57	36,14	18,47
$k = 2$	44,02	54,62	53,60	55,52	48,51	57,30
$k < 2$	11,08	18,61	24,63	35,91	15,35	24,23
$n = 50$						
$k > 2$	41,07	30,71	13,86	7,87	27,76	18,01
$k = 2$	54,37	62,97	72,20	74,11	64,61	71,66
$k < 2$	4,56	6,32	13,94	18,02	7,63	10,33
$n = 80$						
$k > 2$	39,09	32,81	9,55	6,63	22,44	16,89
$k = 2$	59,86	65,99	85,51	87,30	75,39	80,49
$k < 2$	1,05	1,20	4,94	6,07	2,17	2,62
$n = 120$						
$k > 2$	37,96	33,65	7,28	5,65	19,61	16,13
$k = 2$	61,89	66,20	91,57	92,94	80,06	83,40
$k < 2$	0,15	0,15	1,15	1,41	0,33	0,47
$n = 200$						
$k > 2$	36,77	34,27	5,02	4,30	16,42	14,51
$k = 2$	63,23	65,73	94,97	95,68	83,58	85,49
$k < 2$	0,00	0,00	0,01	0,02	0,00	0,00

Tabela A.6 – Frequência (em %) de identificação no Cenário 2 com $\phi = 120$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 2$	55,52	30,70	23,62	7,56	43,75	19,86
$k = 2$	13,29	16,46	14,31	12,41	14,17	15,85
$k < 2$	31,19	52,84	62,07	80,03	42,08	64,29
$n = 50$						
$k > 2$	40,07	28,62	10,24	4,79	24,79	14,78
$k = 2$	17,45	18,56	14,06	11,37	17,06	16,82
$k < 2$	42,48	52,82	75,70	83,84	58,15	68,40
$n = 80$						
$k > 2$	33,47	26,69	4,79	3,01	16,26	11,29
$k = 2$	22,16	23,29	15,39	13,29	21,37	21,02
$k < 2$	44,37	50,02	79,82	83,70	62,37	67,69
$n = 120$						
$k > 2$	31,00	26,67	3,16	2,15	12,59	9,84
$k = 2$	26,79	27,39	17,60	15,93	24,87	24,30
$k < 2$	42,21	45,94	79,24	81,92	62,54	65,86
$n = 200$						
$k > 2$	31,56	29,06	2,34	1,84	11,43	9,91
$k = 2$	33,62	34,18	22,13	21,12	32,65	32,34
$k < 2$	34,82	36,76	75,53	77,04	55,92	57,75

Tabela A.7 – Frequência (em %) de identificação no Cenário 4 com $\phi = 120$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
$n = 30$						
$k > 2$	91,92	44,33	73,49	12,87	62,38	30,39
$k = 2$	7,13	34,83	23,37	39,74	23,91	39,39
$k < 2$	0,95	20,84	3,14	47,39	13,71	30,22
modelo verdadeiro	0,35	30,18	0,97	35,68	20,36	34,62
$n = 50$						
$k > 2$	92,63	51,20	71,51	12,37	47,86	30,25
$k = 2$	6,61	38,12	25,41	52,23	38,86	49,49
$k < 2$	0,76	10,68	3,08	35,40	13,28	20,26
modelo verdadeiro	0,95	35,75	2,03	50,16	36,62	46,91
$n = 80$						
$k > 2$	94,67	54,01	76,31	10,35	37,61	27,78
$k = 2$	4,44	41,46	18,70	66,47	53,85	60,93
$k < 2$	0,89	4,53	4,99	23,18	8,54	11,29
modelo verdadeiro	1,36	39,99	2,83	65,34	52,36	59,41
$n = 120$						
$k > 2$	96,37	54,35	82,80	8,93	32,18	26,52
$k = 2$	3,15	43,68	13,35	77,89	62,97	67,67
$k < 2$	0,48	1,97	3,85	13,18	4,85	5,81
modelo verdadeiro	1,84	43,09	4,03	77,35	62,31	66,98
$n = 200$						
$k > 2$	96,88	52,46	89,74	6,49	24,68	21,55
$k = 2$	3,02	47,12	8,77	89,15	73,99	76,98
$k < 2$	0,10	0,42	1,49	4,36	1,33	1,47
modelo verdadeiro	2,72	47,00	5,94	88,97	73,84	76,81

Tabela A.8 – Frequência (em %) de identificação no Cenário 5 com $\phi = 120$

Ordem do modelo	AIC	AICc	BIC	BICc	HQ	HQc
<i>n</i> = 30						
<i>k</i> > 2	82,34	39,83	51,36	10,03	57,21	26,46
<i>k</i> = 2	15,41	19,90	40,96	17,80	15,22	20,75
<i>k</i> < 2	2,25	40,27	7,68	72,17	27,57	52,79
modelo verdadeiro	0,96	8,02	1,65	7,71	5,93	8,67
<i>n</i> = 50						
<i>k</i> > 2	79,47	44,04	38,41	7,79	39,69	23,60
<i>k</i> = 2	17,08	19,20	49,40	15,57	18,17	20,21
<i>k</i> < 2	3,45	36,76	12,19	76,64	42,14	56,19
modelo verdadeiro	0,57	7,30	0,93	6,42	7,10	8,18
<i>n</i> = 80						
<i>k</i> > 2	76,59	46,72	29,71	5,63	29,77	20,13
<i>k</i> = 2	20,19	21,53	59,25	17,94	22,82	24,17
<i>k</i> < 2	3,22	31,75	11,04	76,43	47,41	55,70
modelo verdadeiro	0,35	7,44	0,68	7,03	8,36	9,02
<i>n</i> = 120						
<i>k</i> > 2	73,11	46,24	24,83	3,93	21,95	16,99
<i>k</i> = 2	24,55	24,40	66,52	19,95	27,50	27,92
<i>k</i> < 2	2,34	29,36	8,65	76,12	50,55	55,09
modelo verdadeiro	0,88	8,87	0,83	8,31	10,76	11,05
<i>n</i> = 200						
<i>k</i> > 2	74,56	45,90	24,94	3,01	18,07	15,44
<i>k</i> = 2	18,56	29,50	49,99	24,62	34,72	34,91
<i>k</i> < 2	6,88	24,60	25,07	72,37	47,21	49,65
modelo verdadeiro	0,99	10,95	1,10	10,64	14,23	14,37