

Identificação de Modelos ARMA pelo Método Bootstrap

Identification of ARMA Models Using Bootstrap

Thais Mariane Biembengut Faria

Universidade Federal do Paraná - UFPR, Curitiba, PR
thaismariane@gmail.com

Anselmo Chaves Neto

Universidade Federal do Paraná - UFPR, Curitiba, PR
anselmo@ufpr.br

Resumo: Nesse artigo, avaliou-se o desempenho de um algoritmo *bootstrap moving blocks* na identificação da ordem de modelos ARMA. O método foi aplicado em séries simuladas a partir de modelos de ordem $p + q > 2$ a fim de obter a distribuição amostral das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Por meio de algumas simulações e exemplos, demonstrou-se que o procedimento *bootstrap* proposto possui melhor desempenho em relação ao método assintótico clássico de identificação, sobretudo em pequenas amostras ou ainda na identificação de modelos com baixos valores para as FAC e FACP.

Palavras-chave: autocorrelações; *bootstrap*; identificação; modelos ARMA.

Abstract: In this article, we evaluated the performance of a moving block bootstrap algorithm in identifying the order of ARMA models. The method was applied to simulated series from models of order $p + q > 2$ in order to obtain the sampling distribution of the autocorrelation functions (FAC) and partial autocorrelation (PACF). By means of some examples and simulations, it was demonstrated that the proposed bootstrap procedure has better performance than the classical identification asymptotic method, especially on small samples or models to identify low values for the FAC and FACP.

Recebido em 22/03/2014 - Aceito em 08/06/2014.

RECEN 16(1) p. 17-33 jan/jun 2014 DOI: 10.5935/RECEN.2014.01.02

Key words: ARMA models; autocorrelations; bootstrap; identification.

1 Introdução

Em problemas de inferência estatística pode ocorrer de não se conhecer a distribuição de probabilidade da estatística de interesse e, ainda, os procedimentos teóricos para se estimar a variabilidade dessa estatística não existem ou são baseados em resultados assintóticos. Uma outra restrição que pode ocorrer é a condição de Gaussianidade para os dados, que muitas vezes é imposta para a distribuição da estatística. Nesses casos, o método computacionalmente intensivo conhecido *bootstrap* pode ser empregado para obter uma aproximação da distribuição da estatística de interesse, medidas da sua variabilidade e construção de testes de hipóteses.

O *bootstrap*, aplicado em séries temporais, surgiu nos trabalhos de Freedman e Peters [1], onde o método foi empregado em modelos econométricos de previsão de energia no mercado americano. Chatterjee [2] e Chaves Neto [3] estudaram a estimação do erro-padrão das estimativas dos parâmetros em modelos ARIMA(p,d,q). Nas últimas décadas surgiram inúmeros trabalhos aplicando o *bootstrap* em séries temporais a fim de avaliar a variabilidade de estatísticas necessárias, ao ajuste de modelos ARMA(p,q) e também na construção de intervalos de previsão [4-9].

Os modelos ARMA(p,q) são uma classe da família ARIMA(p,d,q) que descrevem séries temporais univariadas, estacionárias e não sazonais. O popular método ARIMA foi introduzido por Box & Jenkins [10] na década de 70, e desde então tem sido aplicado em diversas áreas do conhecimento. A técnica consiste num ciclo iterativo de 3 etapas: identificação, ajuste do modelo e testes de adequabilidade. (Para um tratamento detalhado da técnica, ver [11, 12]). A identificação da ordem p, q do modelo, é um procedimento sofisticado que requer muitos dados e razoável experiência do analista. Nesta etapa, compara-se os correlogramas amostrais com os teóricos de diversas estruturas, procurando propriedades desejáveis que permitam identificar um possível modelo para a série temporal. Assim, devem ser estimadas as funções de autocorrelação (FAC) e autocorrelação parcial (FACP), porém as expressões clássicas usadas para medir a variabilidade dos respectivos estimadores, são obtidas com base em resultados fortemente assintóticos. Além disso, dispondo de séries com pou-

cas observações, os intervalos de confiança tradicionais para testar as hipóteses nulas possuem baixo desempenho.

Um outro problema é a dificuldade em reconhecer padrões nas FAC e FACP por meio do método Box & Jenkins, por isso vários métodos alternativos tem sido propostos na literatura nas últimas décadas. Choi [13] avaliou e comparou diferentes procedimentos para a identificação de modelos como: o método corner [14]; os métodos de função de autocorrelação amostral estendida (ESACF) e correlação canônica(SCAN) [15, 16]; dentre outros [17, 18]. A principal característica desses métodos de identificação, é apontar um conjunto de modelos candidatos para uma cuidadosa análise posterior. Um problema importante desses procedimentos alternativos é que raramente a distribuição das estatísticas envolvidas na identificação do modelo é conhecida, e, ainda, em alguns procedimentos, a variância assintótica é estimada pela fórmula de Bartlett, baseada em suposição de Gaussianidade. Pesquisas recentes empregam redes neurais e algoritmos genéticos como alternativas para identificar modelos, livres de suposições sobre a natureza da distribuição das estatísticas envolvidas. Chong [19] e Minerva [20], propõem algoritmos genéticos para identificação de modelos ARMA. Rolf e Urfer [21] utilizam algoritmos evolucionários para identificar e estimar os parâmetros do modelo. Machado [22] compara um algoritmo de *neuro-fuzzy back propagation* com procedimentos automáticos na identificação de modelos Box & Jenkins.

Embora o método *bootstrap* esteja bem difundido, poucas pesquisas aplicam o método na identificação de modelos ARMA(p,q). Paparoditis [23] estudou a identificação de modelos considerando o vetor de autocorrelação, e aplicou o *bootstrap* na avaliação da distribuição amostral das correspondentes estatísticas envolvidas. Chaves Neto [3] identificou regiões do espaço paramétrico de modelos ARMA com baixa ordem, onde o método clássico possui fraco desempenho, e propôs o *bootstrap* como alternativa para identificação destes modelos.

Neste trabalho, um algoritmo *bootstrap moving blocks* não paramétrico foi aplicado, afim de obter informações sobre a distribuição das estatísticas envolvidas no processo de identificação de modelos ARMA de ordem $p + q > 2$. Um estudo de simulação avaliou o desempenho do algoritmo proposto na identificação da estrutura,

comparando com o clássico método Box & Jenkins.

2 Identificação da ordem de modelos ARMA(p,q)

Seja um processo estocástico estacionário no sentido amplo no qual ω_t seja solução da equação:

$$\omega_t = \pi_1\omega_{t-1} + \pi_2\omega_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j\omega_{t-j}.$$

A série associada $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$ converge e é não nula para $|B| \leq 1$. Assume-se que o ruído branco a_t é independente e identicamente distribuído com distribuição Normal tendo $E[a_t] = 0$ e $E[a_{t^2}] = \sigma^2 > 0$. Considera-se o caso em que o processo $(\omega_t; t \in \mathbb{Z})$ pode ser descrito por um modelo ARMA(p,q):

$$\phi(B)\omega_t = \theta(B)a_t,$$

onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ e $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$. O operador de retardo B é definido por $B^m \omega_t = \omega_{t-m}$.

No clássico procedimento proposto por Box & Jenkins [10] para identificação da ordem do modelo ARMA(p,q), são estimadas as funções de autocorrelação e autocorrelação parcial com base na série temporal.

Define-se por $\rho_k = \frac{\gamma_k}{\gamma_0}$, a função de autocorrelação (FAC) do processo de lag k , $k \in \mathbb{Z}$. Onde $\gamma_k = E[\omega_t - E(\omega_t)][\omega_{t+k} - E(\omega_t)]$ é a autocovariância de lag k . O estimador da FAC é:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega})}{\sum_{t=1}^n (\omega_t - \bar{\omega})^2}, \quad (1)$$

sendo $\bar{\omega} = \sum_{t=1}^n \frac{\omega_t}{n}$, a média com base na série. Denota-se por ϕ_{kk} , a função de autocorrelação parcial (FACP) de lag k , que pode ser estimada substituindo ρ_k pelas

estimativas obtidas por (1) nas equações de Yule-Walker [10]:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}, \forall k = 1, 2, 3, \dots \quad (2)$$

No procedimento de identificação, compara-se os correlogramas amostrais da FAC e da FACP com os correlogramas teóricos de diversas estruturas (Tabela 1), procurando propriedades desejáveis que identifique um possível modelo para a série [12].

Tabela 1. Comportamentos das funções de autocorrelação e autocorrelação parcial para um processo ARMA(p,q).

MODELO	FAC	FACP
(1,0)	Decaimento exponencial ou oscilatório	$\phi_{kk} = 0$ para $k > 1$
(2,0)	Decaimento exponencial ou senoidal	$\phi_{kk} = 0$ para $k > 2$
(p,0)	Decaimento exponencial ou senoidal	$\phi_{kk} = 0$ para $k > p$
(0,1)	$\rho_k = 0$ para $k > 1$	Decaimento exponencial
(0,2)	$\rho_k = 0$ para $k > 2$	Decaimento exponencial ou senoidal
(0,q)	$\rho_k = 0$ para $k > q$	Combinação linear decaimento exponencial e ou senoidal
(1,1)	Decaimento exponencial a partir da defasagem 1	Decaimento exponencial a partir da defasagem 1
(p,q)	Decaimento exponencial a partir da defasagem $q - p$	Decaimento exponencial e ou senoidal a partir da defasagem $q - p$

Além da dificuldade em reconhecer padrões nos correlogramas amostrais, um outro problema deste procedimento é verificar, por meio de um teste de hipóteses, se a FAC ou a FACP amostrais são nulas além de um certo *lag* k . As distribuições de probabilidade das estatísticas $\hat{\rho}_k$ e $\hat{\phi}_{kk}$ são aproximadas assintoticamente, e, portanto, os intervalos de confiança empregados no teste de hipóteses possuem baixo desempenho. Sobretudo, na identificação de estruturas ARMA(p,q) que possuem baixos

valores para a FAC e/ou FACP ou ainda quando se dispõe de séries temporais com menos de 50 observações.

Sob a suposição que parâmetro estimado ρ_k é nulo e o tamanho da série n é moderado para grande, a distribuição de $\hat{\rho}_k$ é aproximadamente Normal com esperança nula [24], isto é,

$$\hat{\rho}_k \sim N[0, V(\hat{\rho}_k)].$$

A variância assintótica de $\hat{\rho}_k$ pode ser calculada pela fórmula de Bartlett [25]

$$V(\hat{\rho}_k) \cong \frac{1}{n} \left[1 + 2 \sum_{j=1}^q \rho_j^2 \right] \quad k > q,$$

para o caso de correlações teóricas ρ_j nulas para defasagens k maiores do que uma defasagem fixada $q, j > q$.

Considerando um processo autoregressivo de ordem p , Quenouille [26] mostrou que a variância aproximada de $\hat{\phi}_{kk}$ é:

$$V(\hat{\phi}_{kk}) \cong \frac{1}{n},$$

e se o tamanho da série for grande assume-se que $\hat{\phi}_{kk}$ é normalmente distribuído, ou seja,

$$\hat{\phi}_{kk} \sim N\left[0, \frac{1}{n}\right].$$

Assim, para testar as hipóteses:

$$H_{o_1} : \rho_k = 0 \quad e \quad H_{o_2} : \phi_{kk} = 0, \quad (3)$$

emprega-se os intervalos de confiança de nível $(1 - \alpha)$:

$$I = \left[-z_{(1-\frac{\alpha}{2})} \sqrt{V(\hat{\rho}_k)}; \quad z_{(1-\frac{\alpha}{2})} \sqrt{V(\hat{\rho}_k)} \right], \quad (4)$$

$$II = \left[-z_{(1-\frac{\alpha}{2})} \frac{1}{\sqrt{n}}; \quad z_{(1-\frac{\alpha}{2})} \frac{1}{\sqrt{n}} \right], \quad (5)$$

construídos com base nos clássicos resultados assintóticos, afim de verificar se a FAC e a FACP são nulas a partir de um certo *lag* k .

3 O bootstrap na identificação da ordem do modelo ARMA (p,q)

O método *bootstrap*, introduzido no trabalho de Efron [27], baseia-se na construção de distribuições amostrais por reamostragem de uma única amostra existente. Como é bem conhecida, a técnica consiste em substituir a distribuição desconhecida dos dados F da amostra original por um estimador F' , em geral a função de distribuição empírica \hat{F} . Sob a distribuição estimada escolhida para aproximar a original, podem ser extraídas exaustivas amostras e, portanto características que não poderiam ser avaliadas na estrutura original do problema podem ser estimadas nessa pseudo-estrutura criada pelo processo de reprodução [28].

Suponha uma amostra aleatória $\underline{x} = (x_1, x_2, \dots, x_n)$ ($X_i \sim i.i.d$) de uma população de distribuição F desconhecida. De \hat{F} extraem-se *boot* amostras de mesmo tamanho da amostra original, formando o conjunto $\underline{x}^{*l} = (x_1^{*l}, x_2^{*l}, \dots, x_n^{*l})$ $l = 1, \dots, boot$. Calcula-se para cada uma das *boot* amostras as estatísticas *bootstrap*, $\hat{\theta}^{*l} = t(\underline{x}^{*l}, \hat{F})$. O conjunto $(\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*boot})$ é uma aproximação da verdadeira distribuição amostral da estatística $\hat{\theta}$. Assim, tem-se a estimativa *bootstrap* de θ , $\hat{\theta}^* = \frac{\sum_{l=1}^{boot} \hat{\theta}^{*l}}{boot}$, e seu respectivo desvio padrão

$$\hat{\sigma}_B = \sqrt{\frac{\sum_{l=1}^{boot} (\hat{\theta}^{*l} - \hat{\theta}^*)^2}{boot - 1}}.$$

O número de reamostras *boot* pode ser muito grande, da ordem de 10^6 ou maior, dada a facilidade e velocidade de cálculo computacional atuais.

Para aplicar o *bootstrap* em séries temporais, é necessário um algoritmo que preserve a estrutura de correlação da série, como o *moving blocks* [27]. Por essa técnica as observações da série temporal são agrupadas em blocos de comprimento l . As amostras *bootstrap* são obtidas por reamostragem com reposição destes blocos, formando amostras de mesmo tamanho da série original. O algoritmo descrito a seguir, baseado

em *moving blocks*, foi testado aqui para obter as distribuições amostrais *bootstrap* de ρ_k e ϕ_{kk} , necessárias na identificação da ordem do modelo ARMA(p,q), e avaliar as variabilidades destas estatísticas.

Seja a série histórica de dados $\underline{\omega} = \{\omega_t; t = 1, 2, 3, \dots, n\}$, as reamostras *bootstrap* são obtidas por sorteio com reposição de $n - k$ pares da amostra original de pares $\{(w_t, w_{t+k})\}; t = \{1, 2, \dots, n - k\}$. Assim, tem-se na l -ésima replicação *bootstrap* a amostra de pares (w_t^{*l}, w_{t+k}^{*l}) , na qual é obtida a estimativa de $\rho_k, \hat{\rho}_k^{*l}$ de forma usual (1). Repetindo-se o processo *boot* vezes tem-se o estimador *bootstrap* de ρ_k

$$\hat{\rho}_k^* = \frac{\sum_{l=1}^{boot} \hat{\rho}_k^{*l}}{boot}. \quad (6)$$

As estimativas $\hat{\rho}_k^{*l}$ são elementos da distribuição amostral do estimador $\hat{\rho}_k^*$ que constitui uma aproximação da distribuição amostral de $\hat{\rho}_k$, estimativa clássica de ρ_k , se *boot* for um conjunto muito grande. Assumindo-se que $\hat{\rho}_k^{*l}, l = 1, 2, 3, \dots, boot$, são variáveis aleatórias i.i.d, $\hat{\rho}_k^*$ é um estimador consistente de $E(\hat{\rho}_k)$ [3], isto é,

$$\hat{\rho}_k^* = \frac{\sum_{l=1}^{boot} \hat{\rho}_k^{*l}}{boot} \longrightarrow E(\hat{\rho}_k).$$

A distribuição *bootstrap* de $\hat{\phi}_{kk}$ pode ser obtida a partir da distribuição *bootstrap* de $\hat{\rho}_k, \{\hat{\rho}_k^{*l}; l = 1, 2, 3, \dots, boot\}$, calculando-se em cada replicação o valor de $\hat{\phi}_{kk}^{*l}$ em função da autocorrelação *bootstrap* de defasagem k e de defasagens anteriores por meio usual.

A figura 1 ilustra o algoritmo para obtenção de $\hat{\rho}_k^*$ e $\hat{\phi}_{kk}^*$.

Os erros padrão *bootstrap* de $\hat{\rho}_k$ de $\hat{\phi}_{kk}$ são calculados respectivamente por:

$$s^*(\hat{\rho}_k) = \sqrt{\frac{\sum_{l=1}^{boot} (\hat{\rho}_k^{*l} - \hat{\rho}_k^*)^2}{boot}} \quad (7)$$

e

$$s^*(\hat{\phi}_{kk}) = \sqrt{\frac{\sum_{l=1}^{boot} (\hat{\phi}_{kk}^{*l} - \hat{\phi}_{kk}^*)^2}{boot}}. \quad (8)$$

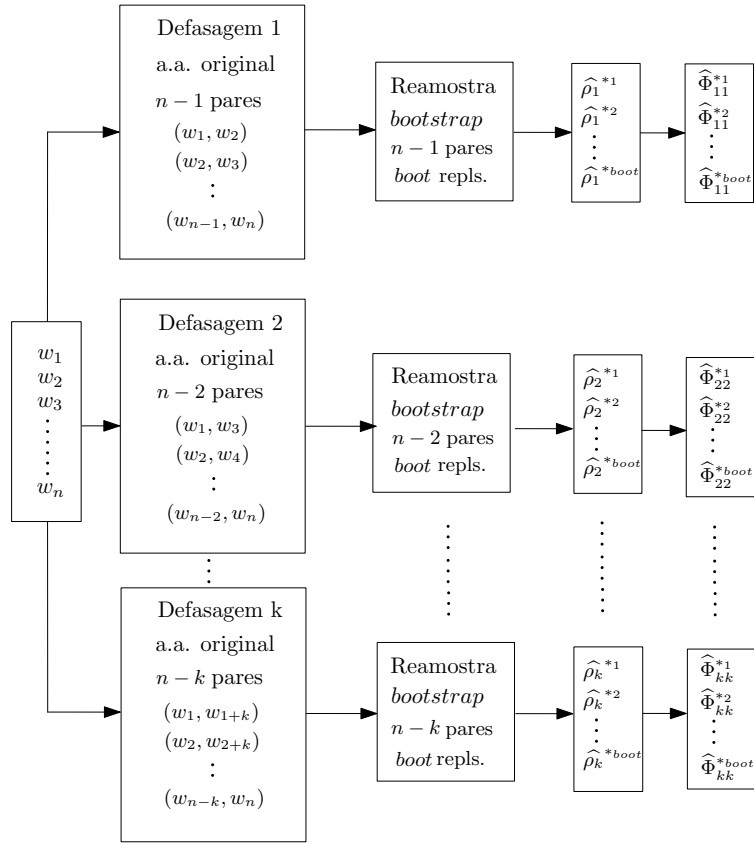


Figura 1. Algoritmo bootstrap - Identificação da Estrutura

Por meio das distribuições de $\hat{\rho}_k^*$ e $\hat{\phi}_{kk}^*$, pode-se obter intervalos de confiança *bootstrap* sem a hipótese de normalidade, por exemplo, os intervalos percentílicos de nível de confiança $1 - \alpha$:

$$[\hat{\rho}_{k_{l\alpha}}^*; \hat{\rho}_{k_{u\alpha}}^*] \quad [\hat{\phi}_{kk_{l\alpha}}^*; \hat{\phi}_{kk_{u\alpha}}^*]. \quad (9)$$

com $l\alpha = 100.\alpha/2\%$ e $u\alpha = 100.(1 - \alpha/2)\%$. Como esses intervalos podem ser assimétricos em relação às estimativas $\hat{\rho}_k^*$ e $\hat{\phi}_{kk}^*$ respectivamente, Efron [29] propôs um intervalo percentílico com vício corrigido (BC)

$$[CDF^{-1}(\phi(2z_o - z_\alpha)); CDF^{-1}(\phi(2z_o + z_\alpha))], \quad (10)$$

para ρ_k ou ϕ_{kk} . Com $z_o = \phi^{-1}(CDF(\hat{\rho}_k^*))$ ou $z_o = \phi^{-1}(CDF(\hat{\phi}_{kk}^*))$, onde ϕ corresponde a função de distribuição da normal padrão.

4 Resultados

A fim de avaliar o desempenho do procedimento *bootstrap*, comparando-o com o método assintótico, simulou-se séries temporais a partir de modelos ARMA de ordem $p + q > 2$. Os resíduos das séries sintéticas são Gaussianos com variância $\sigma_a^2 = 0, 1$, e seu processo gerador é estacionário com expectância nula.

Foram simuladas séries a partir de 15 modelos das estruturas AR(3), MA(3) e ARMA(2,2), algumas com parâmetros escolhidos de forma que $|\rho_k| < c_I$ e $|\phi_{kk}| < c_{II}$, onde c_I e c_{II} são os limites dos intervalos de confiança (4) e (5). Ou seja, escolheu-se modelos com baixos valores da FAC e FACP para avaliar o desempenho do método clássico na identificação deste tipo de estrutura. Como os resultados se repetem em todos os experimentos realizados, por brevidade relata-se apenas uma pequena parte da simulação, que contudo é suficiente para ilustrar os resultados obtidos.

Exemplo 1. Considere os modelos das estruturas AR(3), $\omega_t = 0,7\omega_{t-1} - 0,5\omega_{t-2} + 0,5\omega_{t-3} + a_t$ e $\omega_t = 0,1\omega_{t-1} + 0,1\omega_{t-2} + 0,1\omega_{t-3} + a_t$ com $a_t \sim N(0, 0, 1)$. Estimamos para cada modelo os desvios-padrão das funções de autocorrelação e autocorrelação parcial amostrais. Em 100 repetições Monte Carlo são geradas séries de comprimento $n = 30$, $n = 50$ e $n = 100$ e para cada experimento são obtidos os desvios-padrão exatos, assintóticos e *bootstrap*. Os desvios-padrão exatos foram obtidos por 10000 repetições do modelo. As estimativas assintóticas são calculadas por meio das expressões de Bartlett e Quenoulli. O algoritmo *bootstrap* foi aplicado com $B = 10000$ para cada repetição Monte Carlo. Nas tabelas 2 e 3 estão os valores médios das estimativas dos desvios padrão de ρ_k e ϕ_{kk} para os lags $k = 1, 2, 3$ e 4.

Tabela 2. Estimativas do desvio padrão das FAC e FACP, exato, assintótico e *bootstrap*. Modelo $\omega_t = 0,7\omega_{t-1} - 0,5\omega_{t-2} + 0,5\omega_{t-3} + a_t$

	n = 30			n = 50			n = 100		
	exato	assintótico	<i>bootstrap</i>	exato	assintótico	<i>bootstrap</i>	exato	assintótico	<i>bootstrap</i>
ρ_1	0.164	0.183	0.174	0.137	0.141	0.142	0.103	0.100	0.105
ϕ_{11}	0.164	0.183	0.174	0.137	0.141	0.142	0.097	0.100	0.105
ρ_2	0.173	0.208	0.155	0.148	0.166	0.127	0.107	0.120	0.096
ϕ_{22}	0.164	0.183	0.189	0.132	0.141	0.149	0.089	0.100	0.106
ρ_3	0.183	0.221	0.155	0.159	0.173	0.129	0.123	0.122	0.099
ϕ_{33}	0.158	0.183	0.164	0.125	0.141	0.128	0.095	0.100	0.103
ρ_4	0.174	0.227	0.156	0.146	0.181	0.132	0.109	0.128	0.100
ϕ_{44}	0.154	0.183	0.160	0.129	0.141	0.126	0.096	0.100	0.107

Tabela 3. Estimativas do desvio padrão das FAC e FACP, exato, assintótico e bootstrap. Modelo $\omega_t = 0, 1\omega_{t-1} + 0, 1\omega_{t-2} + 0, 1\omega_{t-3} + a_t$

	n = 30			n = 50			n = 100		
	exato	assintótico	bootstrap	exato	assintótico	bootstrap	exato	assintótico	bootstrap
ρ_1	0.191	0.182	0.158	0.153	0.141	0.130	0.110	0.100	0.096
ϕ_{11}	0.191	0.182	0.158	0.153	0.141	0.130	0.110	0.100	0.095
ρ_2	0.174	0.189	0.159	0.139	0.146	0.124	0.102	0.102	0.097
ϕ_{22}	0.171	0.182	0.166	0.135	0.141	0.129	0.098	0.100	0.099
ρ_3	0.167	0.194	0.153	0.134	0.149	0.125	0.098	0.104	0.096
ϕ_{33}	0.164	0.182	0.165	0.133	0.141	0.133	0.096	0.100	0.099
ρ_4	0.164	0.199	0.155	0.136	0.152	0.125	0.100	0.106	0.096
ϕ_{44}	0.162	0.182	0.173	0.132	0.141	0.136	0.097	0.100	0.101

Em ambos os experimentos, observa-se que as estimativas *bootstrap* têm um bom comportamento em comparação com as estimativas assintóticas, sobretudo nas amostras de tamanhos $n = 30$ e $n = 50$. Nestes casos, na medida que o *lag* k aumenta, as estimativas assintóticas tornam-se mais enviesadas do que as estimativas *bootstrap*.

Também, foram testadas as hipóteses estabelecidas em (3) para cada repetição Monte Carlo, e, assim, pode-se avaliar a probabilidade de cobertura do parâmetro nulo pelos intervalos assintóticos (4) e (5). Os intervalos de confiança *bootstrap* percentílico e percentílico corrigido, foram construídos afim de testar o poder empírico das hipóteses nulas equivalentes:

$$H_{o_1} : [\hat{\rho}_{k_{l_0}}^* ; \hat{\rho}_{k_{u_p}}^*] \supset 0 \quad e \quad H_{o_2} : [\hat{\phi}_{kk_{l_0}}^* ; \hat{\phi}_{kk_{u_p}}^*] \supset 0.$$

Isto é, testa-se a hipótese do zero pertencer aos intervalos. Já nos intervalos clássicos a pergunta é se a estimativa pertence ao intervalo para o parâmetro nulo.

As hipóteses foram testadas para os 4 primeiros *lags* das FAC e FACP de cada uma das estruturas referenciadas. O intervalo percentílico *bootstrap* é do tipo assimétrico com 1% no ramo próximo do zero e 9% na outra extremidade. O nível de confiança de todos os intervalos é de 90%.

Como as séries são simuladas a partir de modelos AR(3), espera-se que $\hat{\phi}_{44}$ seja estatisticamente nulo. Ou seja, $\hat{\phi}_{44}$ pertença aos intervalos clássicos (4) e (5), ou ainda, o zero esteja contido nos intervalos (9) e (10) construídos para este parâmetro.

As tabelas (4) e (5) mostram a probabilidade de cobertura do zero pelos intervalos de confiança assintótico, *bootstrap* percentílico e *bootstrap* BC, para os modelos

considerados.

Tabela 4. Probabilidade de cobertura do zero pelos intervalos assintótico(A), bootstrap percentílico, bootstrap corrigido (BC). Modelo $\omega_t = 0,7\omega_{t-1} + 0,5\omega_{t-2} + 0,5\omega_{t-3} + a_t$

	n = 30			n = 50			n = 100		
	assintótico	percentílico	BC	assintótico	percentílico	BC	assintótico	percentílico	BC
ρ_1	0.35	0.37	0.35	0.16	0.25	0.15	0.00	0.00	0.00
ϕ_{11}	0.35	0.37	0.35	0.16	0.25	0.15	0.00	0.00	0.00
ρ_2	0.77	0.46	0.44	0.63	0.50	0.48	0.76	0.69	0.60
ϕ_{22}	0.53	0.41	0.39	0.35	0.25	0.32	0.19	0.25	0.36
ρ_3	0.81	0.64	0.61	0.76	0.66	0.56	0.30	0.38	0.18
ϕ_{33}	0.45	0.51	0.29	0.14	0.43	0.21	0.00	0.12	0.02
ρ_4	0.81	0.46	0.40	0.48	0.36	0.24	0.05	0.05	0.03
ϕ_{44}	0.96	0.91	0.93	0.91	0.96	1.00	0.89	0.99	0.98

Tabela 5. Probabilidade de cobertura do zero pelos intervalos assintótico(A), bootstrap percentílico, bootstrap corrigido (BC). Modelo $\omega_t = 0,1\omega_{t-1} + 0,1\omega_{t-2} + 0,1\omega_{t-3} + a_t$

	n = 30			n = 50			n = 100		
	assintótico	percentílico	BC	assintótico	percentílico	BC	assintótico	percentílico	BC
ρ_1	0.87	0.78	0.72	0.78	0.76	0.62	0.71	0.80	0.67
ϕ_{11}	0.87	0.78	0.72	0.78	0.76	0.62	0.71	0.80	0.67
ρ_2	0.95	0.76	0.68	0.88	0.82	0.71	0.74	0.85	0.67
ϕ_{22}	0.93	0.79	0.74	0.93	0.85	0.78	0.78	0.88	0.73
ρ_3	0.95	0.82	0.77	0.88	0.85	0.77	0.75	0.78	0.70
ϕ_{33}	0.93	0.80	0.77	0.92	0.88	0.82	0.81	0.85	0.77
ρ_4	0.98	0.82	0.79	0.96	0.80	0.78	0.93	0.89	0.86
ϕ_{44}	0.93	0.82	0.80	0.92	0.79	0.80	0.81	0.84	0.85

Os resultados exibidos na tabela 4 apontam que nas amostras de tamanho $n=30$ e $n=50$ os intervalos *bootstrap*, em especial o BC, possuem maior poder empírico para rejeitar a hipótese nula, ou seja, estimam melhor parâmetros que não são nulos. Já no caso de ϕ_{44} , os intervalos *bootstrap* possuem maior probabilidade de cobertura do zero, ou seja, é mais acurado do que o intervalo assintótico na identificação do parâmetro nulo.

Um problema maior está na identificação da ordem do modelo, a partir das séries simuladas de $\omega_t = 0,1\omega_{t-1} + 0,1\omega_{t-2} + 0,1\omega_{t-3} + a_t$ (tabela 5). Para esta estrutura AR(3), o conjunto de valores dos *lags* da FAC e FACP está contido nos intervalos de confiança assintótico de nível 90% (4) e (5) respectivamente. Observa-se no experimento de simulação que a probabilidade de cobertura do parâmetro nulo é muito alta, inclusive nas amostras de $n = 100$ onde o desempenho do método assintótico é melhor. Assim, a técnica clássica aponta o processo como ruído branco, ao invés de

identificar um modelo com baixos valores para FAC e FACP. Nestes casos o desempenho do *bootstrap* também é superior, em especial nas amostras de tamanho $n = 30$ e $n = 50$, pois a probabilidade de cobertura do zero é menor em ambos intervalos analisados.

Exemplo 2. Considere a estimação dos percentis da distribuição da função de autocorrelação amostral. Na figura 2, a linha sólida representa os percentis 5% e 95% da distribuição exata de $\hat{\rho}_k$ para $k = 1, 2, \dots, 6$ que são obtidos usando 10000 repetições do modelo $\omega_t = 0, 1\omega_{t-1} + 0, 1\omega_{t-2} + 0, 1\omega_{t-3} + a_t$. Os valores médios correspondentes aos percentis análogos da distribuição de $\hat{\rho}_k^*$ ao longo de 200 repetições são representados pela linha pontilhada. Para aplicar o *bootstrap* foi usado $B = 10000$ para cada repetição Monte Carlo. Por fim, os valores médios correspondentes aos percentis da distribuição normal assintótica estão representados pela linha hachurada. As variâncias assintóticas são calculadas em cada uma das 10000 repetições do modelo, por meio da fórmula de Bartlett.

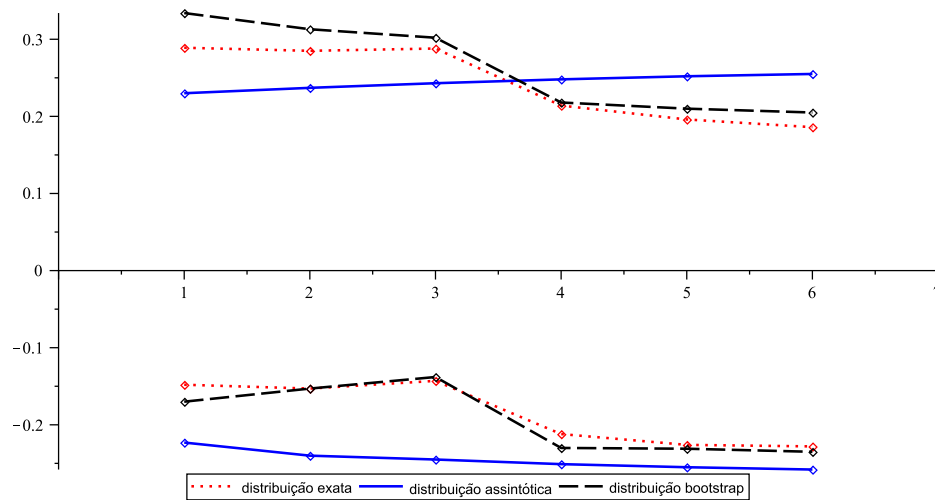


Figura 2. 5% e 95% das distribuições exata, bootstrap e assintótica de ρ_k

Observa-se que as estimativas *bootstrap* refletem mais satisfatoriamente a distribuição amostral da função de autocorrelação parcial do que o método assintótico. Em particular nos casos onde a distribuição não é simétrica o *bootstrap* fornece estimativas mais acuradas.

5 Conclusões

Neste trabalho, foi proposto um procedimento *bootstrap moving blocks* para identificar a ordem de modelos ARMA. O algoritmo foi testado em séries temporais simuladas a partir de modelos das estruturas AR(3), MA(3) e ARMA(2,2). Desta forma, foram determinadas as distribuições amostrais das funções de autocorrelação e autocorrelação parcial, empregadas classicamente na identificação deste tipo de estrutura, livres da suposição de Gaussianidade. Os exemplos mostraram que *bootstrap* tem um bom desempenho em amostras de todos os tamanhos, sendo superior ao método assintótico em amostras pequenas. As estimativas *bootstrap* são mais acuradas, ou seja, possuem uma variabilidade menor do que as estimativas assintóticas.

Na identificação de modelos com baixos valores para a FAC e FACP, o método clássico é ineficaz em amostras de todos os tamanhos, apontado o processo como ruído branco. O *bootstrap* pode ser uma alternativa para este tipo de estrutura, pois os intervalos de confiança possuem uma probabilidade menor de cobertura do parâmetro nulo, ou seja, possuem maior poder para rejeitar as hipóteses nulas.

Estes resultados se repetiram nas séries simuladas a partir de todos os modelos estudados, e pode ser justificado pela distribuição *bootstrap* da FAC e FACP. A comparação com os percentis da distribuição exata, *bootstrap* e assintótica, evidencia que *bootstrap* reproduz de forma mais satisfatória a verdadeira distribuição das funções de autocorrelação e autocorrelação parcial.

Em trabalhos futuros, o algoritmo avaliado aqui será aplicado em séries temporais de ordens maiores. Neste trabalho considerou-se apenas até a ordem 4, pois são modelos que representam a maior parte das séries temporais. Outra possível aplicação para o algoritmo *bootstrap moving blocks* proposto é testá-lo em conjuntos de dados reais, comparando-o com outros métodos de identificação.

Referências

- [1] FREEDMAN, D.; PETERS, S. C. Bootstrapping an econometric model: Some empirical results. *J Bus Econ Stat*, v. 2, n. 2, p. 150-158, 1984.

- [2] CHATTERJEE, S. Bootstrapping arma models: Some simulations. *IEEE Sys, Man Cyb*, v. 2, p. 294–297, 1986.
- [3] CHAVES, Neto A. Bootstrap em Séries Temporais. Tese de Doutorado, PUC, Rio de Janeiro, 1991.
- [4] SAAVEDRA, A.; CAO, R. Rate of convergence of a convolution-type estimator of the marginal density of a ma(1) process. *Stochastic Process*, v. 80, p. 129–155, 1999.
- [5] CAVALIERE, G.; TAYLOR, R. Bootstrap unit root tests for time series with nonstationary volatility. *Econome Theor*, v. 24, p. 43–71, 2008.
- [6] SENSIER, M.; DIJK, Van D. Testing for volatility changes in u.s. macroeconomic time series. *Rev Econ Stat*, v. 86, p. 833–839, 2004.
- [7] HANSEN, B. Regression with nonstationary volatility. *Econometrica*, v. 63, p. 1113–1132, 1995.
- [8] MULLER, U.; SCHICK, A.; WEFELMEYER, W. Weighted residual-based density estimators for nonlinear autoregressive models. *Statist Sinica*, v. 15, p. 177–195, 2005.
- [9] COSKUN, A., CEYHAN, E., INAL, T. C., SERTESER, M. Unsal, I. The comparison of parametric and nonparametric bootstrap methods for reference interval computation in small sample size groups. *Accred Qual Assur*, v. 18, p. 51–60, 2014.
- [10] BOX, G. E. P., JENKINS, G. Time Series Analysis Forecasting and Control. Holden Day, New Jersey, 1976.
- [11] BOX, G. E. P., JENKINS, G., REINSEL, G. C. Time Series Analysis. Prentice Hall, New Jersey, 1994.
- [12] MORETTIN, P.; TOLOI, C. M. C. Análise de Séries Temporais. Blucher, São Paulo, 2006.

- [13] CHOI, B. S. Identification of ARMA Models. Springer, New York, 1992.
- [14] BEGUIN, J. M.; GOURIEROUX, C.; MANFORT, A. Identification of a mixed autoregressive-moving average process : the corner method. Time series : proceedings of the international conference held at Nottingham University, March 1979. p. 423-436, 1980.
- [15] TSAY, R. S.; TIAO, G. C. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *J Am Stat Assoc*, v. 70, p. 84 – 96, 1984.
- [16] TSAY, R. S.; TIAO, G. C. Use of canonical analysis in time series model identification. *Biometrika*, v. 72, n. 2, p. 299–315, 1985.
- [17] KOREISHA, S.; PUKKILA, T. A generalized least-squares approach for estimation of autoregressive moving-average models. *J Time Ser Anal*, v. 2, p. 139–151, 1990.
- [18] HANNAN, E. J.; RISSANEN, J. Recursive estimation of mixed autoregressivemoving average order. *Biometrika*, v. 1, p 81–94, 1982.
- [19] ONG, C. S.; HUANG, J. J.; TZENG, G. H. Model identification of arima family using genetic algorithms. *Appl Math Comput*, v. 164, n. 3, p. 885 – 912, 2005.
- [20] MINERVA, T.; POLI, I. Applications of Evolutionary Computing. Springer, 2001, ch. Building ARMA Models with Genetic Algorithms, pp. 335–342.
- [21] ROLF, S.; SPRAVE, J.; W., U. Model identification and parameter estimation of arma models by means evolutionary algorithms. Proceedings of the *IEEE/IAFE 1 997 Computational Intelligence for Financial Engineering (CIFEr)*, p. 237–243, 1997.
- [22] MACHADO, M. A. S.; SOUZA, R. C.; TRANSCHEIDT, T. Box & Jenkins model identification: A comparison of methodologies. *Independent Journal of Management & Produc*, v.3, n. 2, p. 54–61, 2012.

- [23] PAPANODITIS, E.; STREITBERG, B. Order identification statistics in stationary autoregressive moving-average models: vector autocorrelations and the bootstrap. *J Time Ser Anal*, v. 5, p. 415–434, 1992.
- [24] ANDERSON, R. Distribution of serial correlation coefficient. *Ann Math Stat*, v. 13, p. 1–13, 1942.
- [25] BARTLETT, M. S. On the theoretical specification and sampling properties of autocorrelated time-series. *J R Stat Soc*, v. 27, p. 27–41, 1946.
- [26] QUENOUILLE, M. H. Approximate tests of correlation in time-series. *J Stat Comput Sim*, v. 8, p. 75–80, 1949.
- [27] EFRON, B. Bootstrap methods: another look at jackknife. *Ann Stat*, v. 1, p. 1–26, 1979.
- [28] SILVA, D. O método Bootstrap e aplicações a regressão múltipla. Tese de Doutorado, Unicamp, Campinas, SP, 1995.
- [29] EFRON B. Bootstrap methods for standard errors confidence intervals and other measures of statistics accuracy. *Stat Sci*, v. 1, p. 54–77, 1986.