INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

BOOTSTRAP FOR ORDER IDENTIFICATION IN ARMA(p,q) STRUCTURES

Anselmo Chaves Neto
Universidade Federal do Paraná, Brazil
E-mail: anselmo@ufpr.br

Thais Mariane Biembengut Faria
Universidade Regional de Blumenau, Brazil
E-mail: thaismarianeb@gmail.com

Submission: 14/06/2014
Revision: 25/06/2014
Accept: 13/07/2014

ABSTRACT

The identification of the order p,q, of ARMA models is a critical step in time-series modelling. In the classic Box and Jenkins method of identification, the autocorrelation function (ACF) and the partial autocorrelation (PACF) function should be estimated, but the classical expressions used to measure the variability of the respective estimators are obtained on the basis of asymptotic results. In addition, when having sets of few observations, the traditional confidence intervals to test the null hypotheses display low performance. The bootstrap method may be an alternative for identifying the order of ARMA models, since it allows to obtain an approximation of the distribution of the statistics involved in this step. Therefore it is possible to obtain more accurate confidence intervals than those obtained by the classical method of identification. In this paper we propose a bootstrap procedure to identify the order of ARMA models. The algorithm was tested on simulated time series from models of structures AR(1), AR(2), AR(3), MA(1), MA(2), MA(3), ARMA(1,1) and ARMA (2,2). This way we determined the sampling distributions of ACF and PACF, free from the Gaussian assumption. The examples show that the bootstrap has good performance in samples of all sizes and that it is superior to the asymptotic method for small samples

ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

Keywords: Order Identification; Bootstrap; Correlograms.

1. INTRODUCTION

Let the following be a stationary stochastic process in which ω_t is the solution for the equation

$$\omega_t = \pi_1\omega_{t-1} + \pi_2\omega_{t-2} + \cdots + a_t = \sum_{j=1}^{\infty} \pi_j \omega_{t-j} \quad (1)$$

The associated series $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$ converges and is nonzero for $|B| \leqslant 1$. It is assumed that the white noise a_t, is independent and identically distributed with a normal distribution having $E[a_t] = 0$ and $E[a_{t^2}] = \sigma^2 > 0$. It is considered the case in which the process $(\omega_t; t \in Z)$ can be described by an ARMA(p,q) model, e.g.:

$$\phi(B)\omega_t = \theta(B)a_t, \quad (2)$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ and B is the backshift operator such that $B^m \omega_t = \omega_{t-m}$.

The ARMA(p,q) models are a class of ARIMA(p,d,q) that describe univariate, stationary and unseasonal time series. These models are used in hydrology, econometrics, and other fields. ARMA models can be used to predict behavior of a time series and are widely used for prediction of economic and industrial time series.

The popular ARIMA method was introduced by Box and Jenkins (1976), and the technique consists of an iterative cycle of three steps: identification, model fit and suitability tests.

The identification of the order p, q of the model is a sophisticated procedure that requires a lot of data, and reasonable experience from the analyst. In this step we compare the sample correlograms with the theoretical of various structures, looking for desirable properties which identify a possible model for the time series. This way, the estimated autocorrelation (ACF) and the partial autocorrelation (PACF) functions should be estimated, but the classical expressions used to measure the variability of the respective estimators are obtained on the basis of asymptotic results. In addition, when having sets of few observations, the traditional confidence intervals to test the null hypotheses display low performance.

ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

Another problem is the difficulty in recognizing patterns in the ACF and PACF using the Box and Jenkins method, so several alternative methods have been proposed in the literature over the past decades.

Choi (1992) evaluated and compared different procedures for the identification of models such as the Corner method, the methods of extended sample autocorrelation function (ESACF) and canonical correlation (SCAN). The main feature of these identification methods is to point out a set of candidate models for a posterior careful analysis. A major problem resides in the fact that the distribution of the statistics involved in the identification of the order of the model is rarely known, and in some procedures, the asymptotic variance is estimated by the Bartlett's formula based on the Gaussian assumption.

Recent studies have employed neural networks and genetic algorithms as alternatives to identify models which are free of assumptions about the nature of the distribution of the involved statistics. Minerva (2001) and Ong (2005) have proposed genetic algorithms for the identification of ARMA models. Rolf et al. (1997) have used evolutionary algorithms to identify and estimate the parameters of the model. Machado et al. (2012) have compared an algorithm of neuro-fuzzy back propagation with automatic procedures for identifying Box and Jenkins models.

The bootstrap method may be an alternative for identifying the order of ARMA models, since it allows to obtain an approximation of the distribution of the statistics involved in this step. Therefore it is possible to obtain more accurate confidence intervals than those obtained by the classical method of identification.

In the last decades several studies have applied the bootstrap method in time series with the objective of assessing the variability in the statistics needed, to fit ARMA(p,q) models and also to build prediction intervals (SAAVEDRA; CAO, 1999; CAVALIERE; TAYLOR, 2008; SENSIER; DIJK, 2004; COSKUN; CEYHAN, 2013).

Although the bootstrap method is well known, few studies have applied the method to identify the order of ARMA(p,q) models. Paparoditis (1992) has studied the identification of models by considering the vector of autocorrelation, and by applying the bootstrap in the evaluation of the sampling distributions of the correspondent involved statistics. Chaves Neto (1991) has identified the parameter space of ARMA models with low order, where the classical method has poor performance, and has

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

proposed the bootstrap as an alternative to identify these models.

In this work a moving blocks bootstrap algorithm was applied to obtain information about the distribution of the statistics ACF and PACF involved in identification of ARMA models. Therefore, confidence intervals free of the Gaussian assumption were constructed, classically imposed to obtain the variability of the referenced statistics.

A simulation study evaluated the performance of the proposed algorithm to identify the structure, comparing it with the classical Box and Jenkins method.

2. THE IDENTIFICATION OF ARMA(p,q) MODELS

In the classic procedure for the identification of the order of ARMA(p,q) models proposed by Box and Jenkins (1994), the autocorrelation and partial autocorrelation functions based on the time series are estimated.

The autocorrelation function (ACF) for lag k is defined by $\rho_k = \frac{\gamma_k}{\gamma_o}$ where $\gamma_k = E[\omega_t - E(\omega_t)][\omega_{t+k} - E(\omega_t)]$ is the autocovariance for lag k. The estimator of the ACF is

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega})}{\sum_{t=1}^{n}(\omega_t - \bar{\omega})^2}, \quad (3)$$

where $\bar{\omega} = \sum_{t=1}^{n} \frac{\omega_t}{n}$, is the sample mean of the time series. We denote by ϕ_{kk} the partial autocorrelation function (PACF) of lag k, which can be estimated by substituting estimates from (1) in Yule-Walker equations (BOX; JENKINS, 1994).

In the identification procedure, we compare the sample correlogram of ACF and PACF with theoretical correlograms of various structures, looking for desirable properties that identify a possible model for the series (MORETTIN, 2006).

In addition to the difficulty in recognizing patterns in the sample correlograms, another problem of this procedure is to verify, by means of a hypothesis test, whether the sample ACF or PACF is zero beyond a certain lag k. Probability distributions of the statistics $\hat{\rho}_k$ and $\hat{\varphi}_k$ are approximated asymptotically, and therefore the confidence intervals used in hypothesis testing display low performance, especially in the identification of ARMA(p,q) structures having low values for the ACF and/or PACF or when there are series with less than 50 observations.

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

Under the assumption that the estimated parameter ρ_k is zero and the size series n is moderate to large, the distribution is approximately Normal with zero mean, i.e., $\hat{\rho}_k \sim N[0, V(\hat{\rho}_k)]$ (ANDERSON, 1942).

The asymptotic variance can be calculated by Bartlett's formula

$$V(\hat{\rho}_k) \cong \frac{1}{n}\left[1 + 2\sum_{j=1}^{q} \rho_j^2\right] \; k > q, \tag{4}$$

for the case of zero theoretical correlations ρ_j for lags k greater than a fixed lag q, j>q, (BARTLETT, 1946). Considering an autoregressive process of order p, Quenouille (1949) has showed that the approximate variance of $\hat{\varphi}_{kk}$ is

$$V(\hat{\varphi}_{kk}) \cong \frac{1}{n}\ , \tag{5}$$

and if the size of the series is large, it is assumed that the $\hat{\varphi}_{kk}$ is normally distributed, i.e.,$\hat{\varphi}_{kk} \sim N[0, \frac{1}{n}]$. To test the hypotheses:

$$H_{o_1}: \rho_k = 0 \text{ e } H_{o_2}: \phi_{kk} = 0, \tag{6}$$

we employ confidence intervals at the level $(1-\alpha)$

$$I = [-z_{(1-\frac{\alpha}{2})}\sqrt{V(\hat{\rho}_k)}; z_{(1-\frac{\alpha}{2})}\sqrt{V(\hat{\rho}_k)}] \tag{7}$$

$$II = \left[-z_{\left(1-\frac{\alpha}{2}\right)}\frac{1}{\sqrt{n}}; z_{\left(1-\frac{\alpha}{2}\right)}\frac{1}{\sqrt{n}}\right]. \tag{8}$$

built based on the classical asymptotic results with the objective of verifying whether the ACF and PACF are zero from a certain lag k.

3. THE BOOTSTRAP IN THE IDENTIFICATION OF THE ARMA (p,q) MODEL

The bootstrap method introduced in Efron’s work (1979) is based on the construction of sample distributions by resampling a single existing sample. As it is well known, the technique consists in replacing the unknown distribution of the data F of the original sample data for an F' estimator, in general the empirical distribution function $\hat{F}$. Under the estimated distribution chosen to approximate the original, exhaustive samples can be extracted, and therefore, characteristics that could not be evaluated in the original structure of the problem can now be estimated in this pseudo-structure created by the process of reproduction (SILVA, 1995).

Suppose a random sample $\underline{x} = (x_1, x_2, \ldots, x_n), (X_i \sim i.i.d)$ from a population of

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

unknown distribution F. B samples of the same size of the original sample are extracted from $\hat{F}$, forming the set $\underline{x}^{*l} = (x_1^{*l}, x_2^{*l}, \ldots, x_n^{*l})$, l = 1, ..., B. We calculate the bootstrap statistics $\hat{\theta}^{*l} = t(\underline{x}^{*l}, \hat{F})$ for each of the B samples. The set $(\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B})$ is an approximation of the true sample distribution of the statistics $\hat{\theta}$. This way, we have the bootstrap estimate $\theta, \hat{\theta}^* = \frac{\sum_{l=1}^{B} \hat{\theta}^{*l}}{B}$, and its corresponding standard deviation

$$\hat{\sigma}_B = \sqrt{\frac{\sum_{l=1}^{B} (\hat{\theta}^{*l} - \hat{\theta}^*)^2}{B-1}} \tag{9}$$

To apply the bootstrap in time series, it is necessary to have an algorithm that preserves the correlation structure of the series, such as moving blocks (EFRON, 1979). With this technique the observations of the time series are grouped into blocks of length l. The bootstrap samples are obtained by resampling with replacement of these blocks, forming samples of the same size of the original series. The algorithm described below, based on moving blocks, has been tested here to get the sampling bootstrap distributions ρ_k and ϕ_{kk} which are necessary to evaluate the variability of these statistics in order to identify the order of the ARMA(p,q) model.

In the historical data series $\underline{\omega} = \{\underline{\omega}_t; t = 1,2,3, \ldots, n\}$, the bootstrap samples are obtained by drawing with replacement of n - k pairs of the original sample pairs $\{(\omega_t, \omega_{t+k})\}; t = \{1,2, \ldots, n-k\}$.

This way, we have the l-th bootstrap replication of the sample pairs $({\omega_t}^{*l}, {\omega_{t+k}}^{*l})$, in which the estimate of ρ_k, ${\hat{\rho}_k}^{*l}$ in the usual manner is obtained (1). By repeating the process B times, there is bootstrap estimator of ρ_k,

$${\hat{\rho}_k}^* = \frac{\sum_{l=1}^{B} {\hat{\rho}_k}^{*l}}{B}. \tag{10}$$

The estimates ${\hat{\rho}_k}^{*l}$ are elements of the sampling distribution of the estimator which constitutes an approximation of the sampling distribution of the $\hat{\rho}_k$, classical ρ_k estimate, if B is a very large set.

The $\hat{\varphi}_{kk}$ bootstrap distribution can be obtained from the $\hat{\rho}_k$, $\{{\hat{\rho}_k}^{*l}; l = 1,2,3, \ldots, B\}$ bootstrap distribution, by calculating the ${\hat{\varphi}_{kk}}^{*l}$ value in each replication as a function of the bootstrap autocorrelation lag k and of previous lags reached by usual means.

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

The $\hat{\rho}_k$ and $\hat{\varphi}_{kk}$ bootstrap standard errors are calculated respectively by

$$s(\hat{\rho}_k) = \sqrt{\frac{\sum_{l=1}^{B}\left(\hat{\rho}_k^{*l} - \hat{\rho}_k^{*}\right)^2}{B}}, \quad (11)$$

$$s(\hat{\phi}_{kk}) = \sqrt{\frac{\sum_{l=1}^{B}(\hat{\varphi}_{kk}^{*l} - \hat{\varphi}_{kk}^{*})^2}{B}}. \quad (12)$$

By means of the $\hat{\rho}_k$ and $\hat{\varphi}_{kk}$ distributions, we can obtain bootstrap confidence intervals without the assumption of normality, for instance, the percentile intervals of the confidence level $1 - \alpha$,

$$[\hat{\rho}_{k_{lo}}; \hat{\rho}_{k_{up}}] \quad [\hat{\varphi}_{kk_{lo}}; \hat{\varphi}_{kk_{up}}]. \quad (13)$$

with $\mathrm{lo} = 100.\frac{\alpha}{2}\%$ e $\mathrm{up} = 100.\left(1 - \frac{\alpha}{2}\right)\%$. Since these intervals can be asymmetric in relation to the $\hat{\rho}_k$ and $\hat{\varphi}_{kk}$ estimates, respectively, Efron (1986) has proposed the bias corrected percentile interval (BC)

$$[CDF^{-1}(\phi(2z_o - z_\alpha)); CDF^{-1}(\phi(2z_o + z_\alpha))], \quad (14)$$

for ρ_k or either φ_{kk}. With $z_o = \phi^{-1}(CDF(\hat{\rho}_k))$ou $z_o = \phi^{-1}(CDF(\hat{\varphi}_{kk}))$, where ϕ corresponds to the distribution function of the standard-normal.

4. RESULTS

In order to evaluate the performance of the bootstrap procedure, by comparing it with the asymptotic method, we simulated time series from ARMA models. The residues of synthetic series are Gaussian with variance $\sigma_a^2 = 0{,}1$, and their generating process is stationary with zero mean.

Series were simulated departing from each 15 model structures AR (1), MA (1), AR (2), MA (2), AR (3), MA (3), ARMA(1,1) and ARMA (2,2), some with parameters chosen so that | ρ_k |< c_1 and | ϕ_{kk} |< c_2, where c1 and c2 are the limits of the confidence intervals (7) and (8). That is, models with low values of ACF and PACF were selected to evaluate the performance of the classical method in the identification of this type of structure. As the results are repeated in all of the experiments, we report a small portion of the simulation, which is sufficient to illustrate the results obtained.

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

Consider the model of the structure MA(2), $\omega_t = -0.2a_{t-1} + 0.1a_{t-2} + a_t$ and the model of the structure AR(3), $\omega_t = 0.7\omega_{t-1} - 0.5\omega_{t-2} + 0.5\omega_{t-3} + a_t$ with $a_t \sim N(0,0.1)$. We estimated for each model the standard deviations of the autocorrelation and partial autocorrelation functions of the sample. In 100 Monte Carlo repetitions, length series n = 30, n = 50 and n = 100 are generated, and for each experiment exact standard deviations, asymptotic and bootstrap are obtained. The exact standard deviations were obtained from 10000 replications of the model.

The asymptotic estimates are calculated through the expressions of Bartlett and Quenoulli. The bootstrap algorithm was applied with B = 1000 for each Monte Carlo repetition. Tables 1 and 2 show the average values of the estimated standard deviations for lags k = 1,2,3,4.

Table 1: Estimates of the standard deviation of the ACF and PACF for the model $\omega_t = -0.2a_{t-1} + 0.1a_{t-2} + a_t$

	n = 30			n = 50			n = 100		
	exact	asymptotic	bootstrap	exact	asymptotic	bootstrap	exact	asymptotic	bootstrap
ρ_1	0.129	0.182	0.168	0.094	0.141	0.138	0.072	0.100	0.098
ϕ_{11}	0.129	0.182	0.168	0.094	0.141	0.138	0.072	0.100	0.098
ρ_2	0.139	0.187	0.172	0.122	0.144	0.138	0.072	0.101	0.101
ϕ_{22}	0.139	0.182	0.178	0.114	0.141	0.141	0.069	0.100	0.103
ρ_3	0.193	0.210	0.157	0.144	0.163	0.126	0.118	0.116	0.096
ϕ_{33}	0.149	0.182	0.169	0.115	0.141	0.138	0.091	0.100	0.102
ρ_4	0.186	0.216	0.156	0.158	0.166	0.127	0.107	0.117	0.094
ϕ_{44}	0.159	0.182	0.170	0.129	0.141	0.133	0.093	0.100	0.103

Table 2: Estimates of the standard deviation of the ACF and PACF for the model $\omega_t = 0.7\omega_{t-1} - 0.5\omega_{t-2} + 0.5\omega_{t-3} + a_t$

	n = 30			n = 50			n = 100		
	exact	asymptotic	bootstrap	exact	asymptotic	bootstrap	exact	asymptotic	bootstrap
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	0.096
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

In both experiments we observed that the bootstrap estimates display good behavior in comparison with the asymptotic estimates, especially in samples of size

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

n = 30 and n = 50. In this case, in so far the lag k increases, the asymptotic estimates become more biased than the bootstrap estimates.

Consider the estimation of percentiles of the distribution of the autocorrelation function. In Figure 1, the dotted line represents the percentiles 5% and 95% of the exact distribution of $\hat{\rho}_k$ for k = 1,2, ..,6, which are obtained using 10000 repetitions of the model ARMA(1,1), $\omega_t = 0.4\omega_{t-1} - 0.5a_{t-1} + a_t$. The average values corresponding to the analog percentiles of the $\hat{\rho}_k$ bootstrap distribution over 200 repetitions are represented by the hatched line. To apply the bootstrap, we used B = 1000 for each Monte Carlo repetition. The mean values corresponding to percentiles of the asymptotic normal distribution are represented by the solid line. The asymptotic variances are calculated for each of the 1000 repetitions of the model using the Bartlett's formula.

We observe that the bootstrap estimates reflect more adequately the sampling distribution of the partial autocorrelation function of the asymptotic method. Particularly in cases where the distribution is not symmetric, the bootstrap provides more accurate estimates.

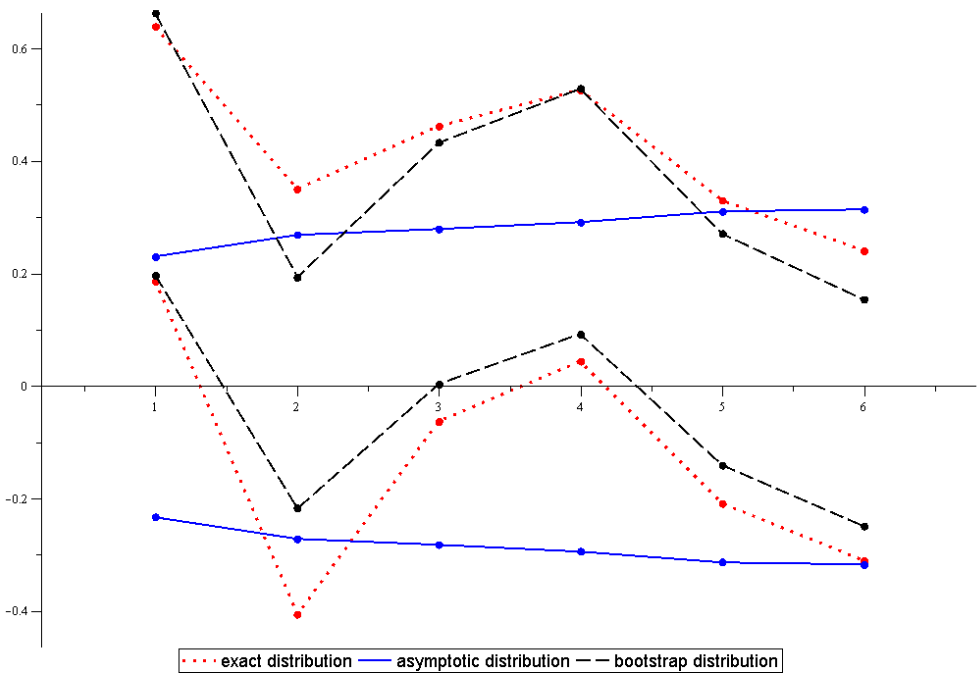

Figure 1: 5% and 95% of the exact, bootstrap and asymptotic distributions of ρ_k

The assumptions set out in (6) for each Monte Carlo repetition were also tested, and this way we could evaluate the coverage probability of the null parameter

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

by the asymptotic intervals (7) and (8). The percentile bootstrap confidence interval (13) and bias corrected percentile interval (14) were constructed to test the equivalent null hypotheses, $H_{o_1}: [\hat{\rho}_{k_{lo}}; \hat{\rho}_{k_{up}}] \supset 0$ and $H_{o_2}: [\hat{\phi}_{kk_{lo}}; \hat{\phi}_{kk_{up}}] \supset 0$. That is, we tested the hypothesis of zero belonging to the intervals. In the classic intervals, the question relies on whether the estimate belongs to the interval for the null parameter.

The hypotheses were tested for the first 4 lags of the ACF and PACF for each of the referred structures. The confidence level of all intervals is 95%. Table (3) displays the probability of coverage of the confidence intervals, for the model AR(3) $\omega_t = 0.7\omega_{t-1} + 0.5\omega_{t-2} + 0.5\omega_{t-3} + a_t$.

Table 3: Probability coverage of zero by asymptotic (A) percentile (B) and bias corrected bootstrap (BC) intervals for the model $\omega_t = 0.7\omega_{t-1} + 0.5\omega_{t-2} + 0.5\omega_{t-3} + a_t$

	n = 30			n = 50			n = 100		
	C	B	BC	C	B	BC	C	B	BC
ρ_1	0.37	0.38	0.35	0.17	0.27	0.16	0.00	0.00	0.00
ϕ_{11}	0.37	0.38	0.35	0.17	0.27	0.16	0.00	0.00	0.00
ρ_2	0.81	0.48	0.47	0.65	0.50	0.47	0.77	0.64	0.60
ϕ_{22}	0.56	0.42	0.39	0.38	0.27	0.19	0.22	0.25	0.30
ρ_3	0.81	0.63	0.60	0.76	0.68	0.54	0.31	0.37	0.20
ϕ_{33}	0.47	0.50	0.32	0.15	0.40	0.17	0.00	0.14	0.03
ρ_4	0.84	0.46	0.42	0.50	0.39	0.25	0.08	0.06	0.05
ϕ_{44}	0.92	0.93	0.95	0.91	0.98	1.00	0.86	0.99	1.00

The results presented in Table 3 reveal that in samples of size n = 30 and n = 50 the bootstrap intervals, especially BC, have higher empirical power to reject the null hypothesis, i.e., they better estimate parameters that are not null.

When the series are simulated departing from AR(3) models, we expected $\hat{\phi}_{44}$ to be statistically null. That is, we expected $\hat{\phi}_{44}$ to belong to the classical intervals (7) and (8), or the zero to be contained in the intervals (13) and (14) constructed for this parameter. In the case of ϕ_{44}, the bootstrap intervals are more likely to cover zero, i.e., they are more accurate than the asymptotic interval in identifying the null parameter.

A major problem relies on the identification of the order of the model departing from the simulated series with parameters chosen so that $| \rho_k |< c_1$ and $| \phi_{kk} |< c_2$. For these structures the set of values of the lags of the ACF and PACF is contained

ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

in the asymptotic confidence interval of 95% level (7) and (8) respectively. We observed in simulation experiments that the probability coverage of the null parameter is very high, even in the samples of n = 100 where the asymptotic performance of the method is better.

This way, the classical technique considers the process as white noise, instead of identifying a model with low values for ACF and PACF. In these cases the bootstrap performance is also superior, especially in samples of size n = 30 n = 50 because the probability coverage of zero is less in both the analyzed intervals.

5. CONCLUSIONS

In this paper we propose a bootstrap procedure to identify the order of ARMA models. The algorithm was tested on simulated time series from models of structures AR(1), AR(2), AR(3), MA(1), MA(2), MA(3), ARMA(1,1) and ARMA (2,2). This way we determined the sampling distributions of the autocorrelation and partial autocorrelation functions, classically used in the identification of this type of structure, free from the Gaussian assumption. The examples show that the bootstrap has good performance in samples of all sizes and that it is superior to the asymptotic method for small samples. The bootstrap estimates are more accurate, i.e., they display less variability than the asymptotic estimates.

In the identification of models with low values for the ACF and PACF, the classic method is ineffective for samples of any size, by considering the process as white noise. The bootstrap can be an alternative to this type of structure, because the confidence intervals have a lower probability coverage of the null parameter, i.e., they have more power to reject the null hypotheses.

These results were repeated in the simulated series from all the models studied and the repetition of results may be justified by the bootstrap distribution of ACF and PACF. The comparison among the percentiles of the exact, asymptotic and bootstrap distribution shows that the bootstrap reproduces more satisfactorily the true distribution of the autocorrelation and partial autocorrelation functions.

As a suggestion for further research, it would be interesting to apply the technique presented here to real time series data with the objective of identifying ARMA(p, q) models, which after adjusted, could be employed, for instance, for predicting the behavior of economic or industrial series.

http://www.ijmp.jor.br v. 6, n. 1, January - March 2015
ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

REFERENCES

ANDERSON, R. (1942) Distribuition of serial correlation coefficient. **Annals of Mathematics Statistics**, n. 13, p. 1-13

BARTLETT, M. S. (1946) On the theoretical specification and sampling properties of autocorrelated time-series. **Journal of the Royal Statistical Society***,* v. 8, n. 27, p. 27-41.

BOX, G. E. P.; JENKINS, G. (1976) **Time Series Analysis Forecasting and Control**. Holen Day: New Jersey.

BOX, G. E. P.; JENKINS, G.; REINSEL, G. C. (1994) **Time Series Analysis**. Prentice Hall: New Jersey.

CAVALIERE, G.; TAYLOR, R. (2008) Bootstrap unit root tests for time series with nonstationary volatility. **Econometric Theory**, n. 24, p. 43-71.

CHOI, B. S. (1992) **Identification of ARMA Models**. Springer: New York.

COSKUN, A.; CEYHAN, E.; INAL, T. C.; SERTESER, M; UNSAL, I. (2013) The comparison of parametric and nonparametric bootstrap methods for reference interval computation in small sample size groups. **Accred Qual Assur**, n. 18, p. 51-60.

EFRON, B. (1979) Bootstrap methods: another look at jackknife. **Annals of Statistics** v. 7, n. 1, p. 1-26.

EFRON, B. (1986) Bootstrap methods for standard errors confidence intervals and other measures of statistics accuracy. **Statistical Science***,* v. 1, n. 1, p. 54-77.

MACHADO, M. A. S; SOUZA, R. C. (2012) Box & Jenkins model identification: A comparison of methodologies. **Independent Journal of Management & Production***,* v. 3, n. 2, p. 54-61.

MINERVA, T.; POLI, I. (2001) ARMA models with genetic algorithms, in: **Applications of Evolutionary Computing**. Springer: New York, p. 335-342.

MORETTIN, P.; TOLOI, C. M. C.(2006) **Análise de séries temporais**. Blucher: São Paulo.

MULLER, U.; SCHICK, A.; WEFELMEYER, W.(2005) Weighted residual-based density estimators for nonlinear autoregressive models. **Statistc Sinica**, n. 15, p. 177-195.

NETO CHAVES, A.(1991) **Bootstrap em séries temporais**. Thesi (PhD in Eletric Engineering), PUC: Rio de Janeiro.

ONG, C.S.; HUANG, J.J.; TZENG, G.H. (2005) Model identification of arima family using genetic algorithms. **Applied Mathematics and Computation**, v. 164, n. 3, p. 885-912.

PAPARODITIS, E.; STREITBERG, B. (1992) Order identification statistics in stationary autoregressive moving-average models: vector autocorrelations and the bootstrap. **Journal of Time Series Analysis***,* v. 13, n. 5, p. 415-434.

QUENOUILLE, M. H. (1949) Approximate tests of correlation in time-series. **Journal of Statistical Computation and Simulation**, n. 8, p. 75-80.

ISSN: 2236-269X
DOI: 10.14807/ijmp.v6i1.244

ROLF, S.; SPRAVE, J. (1997) Model identification and parameter estimation of arma models by means evolutionary algorithms. **Computational Intelligence for Financial Engineering (CIFEr)**, v. 1, n. 997, p. 237-243.

SAAVEDRA, A.; CAO, R. (1999) Rate of convergence of a convolution-type estimator of the marginal density of a ma(1) process. **Stochastic Process**, n. 80, p. 129-155.

SENSIER, M.; VAN DIJK, D. (2004) Testing for volatility changes in u.s. macroeconomic time series. **Review of Economics and Statistics**, n. 86, p. 833–839.

SILVA, D. (1995) **O método bootstrap e aplicações a regressão múltipla**. Dissertation (Master in Statistics), Unicamp: Campinas.