

BOOTSTRAP PREDICTIVE INFERENCE FOR ARIMA PROCESSES

BY LORENZO PASCUAL, JUAN ROMO AND ESTHER RUIZ

Universidad Carlos III de Madrid



Abstract. In this study, we propose a new bootstrap strategy to obtain prediction intervals for autoregressive integrated moving average processes. Its main advantage over other bootstrap methods previously proposed for autoregressive integrated processes is that variability due to parameter estimation can be incorporated into prediction intervals without requiring the backward representation of the process. Consequently, the procedure is very flexible and can be extended to processes even if their backward representation is not available. Furthermore, its implementation is very simple. The asymptotic properties of the bootstrap prediction densities are obtained. Extensive finite sample Monte Carlo experiments are carried out to compare the performance of the proposed strategy vs. alternative procedures. The behaviour of our proposal equals or outperforms the alternatives in most of the cases. Furthermore, our bootstrap strategy is also applied for the first time to obtain the prediction density of processes with moving average components.

Keywords. Forecasting; non Gaussian distributions; prediction density; resampling methods; simulation.

1. INTRODUCTION

Forecasting is one of the main goals in univariate time-series analysis. The problem is providing information about the distribution of the variable Y_{T+k} conditional on a realization of the past variables $\mathbf{Y}_T = \{Y_1, \dots, Y_T\}$. In particular, the objective is to construct prediction intervals $I(\mathbf{Y}_T) = \{L(\mathbf{Y}_T), U(\mathbf{Y}_T)\}$ designed to capture the future value of Y_{T+k} with a fixed probability, the nominal coverage. We will focus on prediction of future values of time series generated by autoregressive integrated moving-average (ARIMA) processes with possibly non-Gaussian innovations.

The standard prediction approach for ARIMA processes (Box and Jenkins, 1976) assumes Gaussian innovations and known parameters. Consequently, the resulting prediction intervals are centered around the conditional expectation which is a linear function of past observations and do not incorporate the uncertainty due to parameter estimation.

Alternatively, bootstrap-based methods provide prediction intervals without any distributional assumption on the innovations. There are several bootstrap alternatives in the literature to construct prediction intervals for autoregressive models of order p (AR(p)). Findley (1986), Stine (1987), Masarotto (1990) and

Grigoletto (1998) use bootstrap methods to estimate the density of the prediction errors including uncertainty due to parameter estimation. As in the standard method, they centre the forecast intervals at a linear combination of past observations. Alternatively, Thombs and Schucany (1990) and Breidt *et al.* (1995) directly estimate the distribution of Y_{T+k} conditional on \mathbf{Y}_T . In an AR(p) process, conditioning on \mathbf{Y}_T is equivalent to conditioning on the last p observations. Consequently, Thombs and Schucany (1990) and Breidt *et al.* (1995) use the backward representation of AR(p) models to generate bootstrap series that mimic the structure of the original data with fixed last p observations. McCullogh (1994) applies the results in Thombs and Schucany (1990) and Breidt *et al.* (1995) to real data implementing also the bias-correction bootstrap of Efron (1982). García-Jurado *et al.* (1995) extend the bootstrap approach of Thombs and Schucany (1990) to autoregressive integrated (ARI) processes. They use the backward representation of the autoregressive model to construct bootstrap replicates of the differenced variable k periods ahead and then obtain bootstrap samples of the original variable Y_{T+k} by solving a $(k+d) \times (k+d)$ linear system, where d is the number of unit roots. When forecasting far ahead, the huge dimension makes the system difficult to handle. The need of the backward representation to generate bootstrap series makes all these methods computationally expensive and, what is more important, restricts their applicability to models having a backward representation, excluding, for example, generalized autoregressive conditionally heteroscedastic (GARCH) processes. Furthermore, the prediction of moving-average processes cannot be handled by these techniques because the infinite order of their autoregressive representation requires that, at least theoretically, the whole sample should be fixed to generate bootstrap replicates. An additional difficulty with this backward representation approach is that, although in AR(p) processes the distribution of Y_{T+k} conditional on \mathbf{Y}_T coincides with the distribution conditional on the last p observations under known parameters, if the parameters are estimated, these distributions are different for finite sample sizes. Kabaila (1993) questions whether predictive inference should be carried out conditioning on the last p observed values. Finally, Cao *et al.* (1997) present an alternative bootstrap method for constructing prediction intervals for stationary AR(p) models which does not require the backward representation. However, their intervals do not incorporate the variability due to parameter estimation.

In this study, we propose a simple resampling procedure for ARIMA processes to estimate the conditional distribution of Y_{T+k} directly, incorporating the variability due to parameter estimation. Our strategy makes the backward representation unnecessary and, as a consequence, this bootstrap procedure can be easily extended to forecasting with more general models.

The paper is organized as follows. Section 2 presents the resampling procedure to estimate prediction distributions and establishes its asymptotic validity for AR(p) processes. In Section 3, we extend this procedure to ARIMA processes and we establish its asymptotic validity. Section 4 contains an extensive Monte Carlo simulation study which compares the performance of several available bootstrap prediction techniques for different ARIMA models and error distributions.

Finally, the conclusions and some ideas for further research can be found in Section 5.

2. BOOTSTRAP PREDICTION INTERVALS FOR STATIONARY AR(p) PROCESSES

Let $\mathbf{y}_T = \{y_1, \dots, y_T\}$ be a sequence of T observations generated by a stationary AR(p) process given by

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t, \quad t = \dots, -2, -1, 0, 1, 2, \dots, \quad (1)$$

where $\{a_t\}$ is a sequence of zero-mean independent random variables with common distribution function F_a such that $E(a_t^2) = \sigma_a^2 < \infty$, $\phi = (\phi_0, \phi_1, \dots, \phi_p)$ are unknown parameters and all the roots of the autoregressive polynomial $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ lie outside the unit circle.

Conditional on $\mathbf{Y}_T = \{Y_1, \dots, Y_T\}$, the minimum mean square error (MSE) predictor of Y_{T+k} is given by the conditional mean of Y_{T+k} ,

$$\tilde{Y}_{T+k} = \phi_0 + \phi_1 \tilde{Y}_{T+k-1} + \dots + \phi_p \tilde{Y}_{T+k-p}, \quad (2)$$

where $\tilde{Y}_{T+j} = Y_{T+j}$ for $j \leq 0$. The prediction error is a combination of future innovations a_{T+j} , $j = 1, \dots, k$, given by

$$\tilde{e}_{T+k} = Y_{T+k} - \tilde{Y}_{T+k} = \sum_{i=0}^{k-1} \Psi_i a_{T+k-i}, \quad (3)$$

where Ψ_i are the coefficients of the moving-average representation of the AR(p) model obtained from $\Psi(B) = \Phi(B)^{-1}$, where B is the backshift operator. The prediction MSE is

$$\text{MSE}(\tilde{e}_{T+k}) = \sigma_a^2 \sum_{i=0}^{k-1} \Psi_i^2. \quad (4)$$

Usually, since the parameters are unknown, predictions are made with estimated parameters. The actual predictor is then given by

$$\hat{Y}_{T+k} = \hat{\phi}_0 + \hat{\phi}_1 \hat{Y}_{T+k-1} + \dots + \hat{\phi}_p \hat{Y}_{T+k-p}, \quad (5)$$

where $\hat{\phi} = (\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p)$ are parameter estimators and $\hat{Y}_{T+j} = Y_{T+j}$ for $j \leq 0$. The corresponding prediction error can be separated into two parts by writing

$$\hat{e}_{T+k} = Y_{T+k} - \hat{Y}_{T+k} = (Y_{T+k} - \tilde{Y}_{T+k}) + (\tilde{Y}_{T+k} - \hat{Y}_{T+k}). \quad (6)$$

The first term in (6) is the prediction error in (3). The second term appears because parameter estimates are used instead of true values. So, in practice, the uncertainty due to parameter estimation has to be included in the expression of the prediction MSE.

The prediction intervals for Y_{T+k} constructed using the Box and Jenkins (1976) procedure are given by

$$\left\{ \hat{Y}_{T+k} - z_{\alpha/2} \left(\hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\Psi}_j^2 \right)^{1/2}, \hat{Y}_{T+k} + z_{\alpha/2} \left(\hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\Psi}_j^2 \right)^{1/2} \right\}, \quad (7)$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution, $\hat{\sigma}_a^2$ is the usual estimate of the innovations variance and $\hat{\Psi}_j$ are the estimated coefficients of the moving average representation. The prediction intervals in (7) just consider the MSE in (4) and replace the unknown parameters by appropriate estimates. However, they do not incorporate the variability due to parameter estimation. Moreover, these intervals have two additional problems when the distribution of a_t is not normal. First, the value of the standard normal quantile may not be appropriate. To handle this question, Findley (1986), Stine (1987), Masarotto (1990) and Grigoletto (1998) proposed different ways of bootstrapping from the residuals of the estimated model,

$$\hat{a}_t = y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p}, \quad t = p+1, \dots, T, \quad (8)$$

to estimate the distribution function of the prediction error. The second difficulty is that these bootstrap prediction intervals are still centered in (5) and when the innovation distribution is not symmetric, this could be inappropriate.

To solve this problem, Thombs and Schucany (1990) introduced a bootstrap method based on directly estimating the distribution of Y_{T+k} conditional on the available variables \mathbf{Y}_T . To incorporate the uncertainty due to parameter estimation in the prediction intervals, they generated bootstrap replicates $\mathbf{y}_T^* = \{y_1^*, \dots, y_T^*\}$ that mimic the structure of the original series. Since the prediction is conditional on the last p values of the series, all the bootstrap replicates are generated fixing the last p values; this requires the backward representation of stationary AR(p) models, where Y_t is expressed as a linear combination of future values plus an error term. Using the backward representation makes the procedure computationally demanding and it constitutes an obstacle to extend the resampling procedure to models without backward representation. To overcome this problem, Cao *et al.* (1997) proposed a fast procedure, called conditional bootstrap, to generate prediction intervals based on resampling residuals but their method does not incorporate variability due to parameter estimation.

In this section, we introduce a new resampling strategy to build prediction intervals in AR(p) models. Our method is based on fixing the last p observations to obtain bootstrap replicates of future values Y_{T+k} but the estimated parameters are bootstrapped without fixing any observation in the sample. As a consequence, we do not need the backward representation of the model and, therefore, the method can be easily extended to more general models.

Our proposal to obtain bootstrap replicates of the series is as follows. Given a set of estimates of the AR(p) model, obtain the residuals by (8) and centre and

rescale them, as suggested by Stine (1987), by the factor $(T - p/T - 2p)^{\frac{1}{2}}$. From a set of p initial values, say $\mathbf{y}_0^* = \{y_{p+1}^*, \dots, y_T^*\}$, construct a bootstrap series $\{y_1^*, \dots, y_T^*\}$ from

$$Y_t^* = \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1}^* + \dots + \hat{\phi}_p Y_{t-p}^* + \hat{a}_t^*, \quad t = 1, \dots, T, \quad (9)$$

where \hat{a}_t^* are independent observations obtained by resampling from \hat{F}_a , the empirical distribution function of the centered and rescaled residuals. Once the parameters of this bootstrap series are estimated, say $\hat{\phi}^* = (\hat{\phi}_0^*, \hat{\phi}_1^*, \dots, \hat{\phi}_p^*)$, we forecast through the recursion of the autoregressive model with the bootstrap parameters and fixing the last p observations of the original series,

$$Y_{T+k}^* = \hat{\phi}_0^* + \sum_{j=1}^p \hat{\phi}_j^* Y_{T+k-j}^* + \hat{a}_{T+k}^*, \quad (10)$$

with \hat{a}_{T+k}^* being a random draw from \hat{F}_a and $Y_{T+h}^* = y_{T+h}$, $h \leq 0$. Once we obtain a set of B bootstrap replicates $\{y_{T+k}^{*(1)}, \dots, y_{T+k}^{*(B)}\}$, we proceed as in Thombs and Schucany (1990). The prediction limits are defined as the quantiles of the bootstrap distribution function of Y_{T+k}^* . More specifically, if $G^*(h) = \Pr(Y_{T+k}^* \leq h)$ is the distribution function of Y_{T+k}^* and $G_B^*(h) = \#(y_{T+k}^{*(b)} \leq h)/B$ is its Monte Carlo estimate, a $100\alpha\%$ prediction interval for Y_{T+k}^* is given by

$$\{L_B^*(\mathbf{y}), U_B^*(\mathbf{y})\} = \{Q_B^*(\frac{1}{2} - \frac{1}{2}\alpha), Q_B^*(\frac{1}{2} + \frac{1}{2}\alpha)\}, \quad (11)$$

where $Q_B^* = G_B^{*-1}$. The main difference between our bootstrap strategy and Thombs and Schucany's (1990) is that our bootstrap parameter estimates are not conditional on the last p observations and this allows us to overcome the computational burden associated with resampling through the backward representation. Moreover, this procedure can be extended to forecasting with more general and complex models.

Summarizing, the steps for obtaining bootstrap prediction intervals are:

- Step 1. Compute the residuals \hat{a}_t as in (8). Let \hat{F}_a be the empirical distribution function of the centered and rescaled residuals.
- Step 2. Generate a bootstrap series using the recursion in (9) and calculate the estimates $\hat{\phi}^*$.
- Step 3. Obtain a bootstrap future value by expression (10). Note that the last p values of the series are fixed in this step but not in the previous one.
- Step 4. Repeat the last two steps B times and then go to step 5.
- Step 5. The endpoints of the prediction interval are given by quantiles of G_B^* , the bootstrap distribution function of Y_{T+k}^* .

The asymptotic properties of the proposed bootstrap procedure are analysed in the following section where we deal with the more general ARIMA(p, d, q) model.

3. PREDICTION FOR ARIMA MODELS

In this section, we generalize the resampling scheme introduced above to ARIMA(p, d, q) process given by

$$\begin{aligned} \nabla^d Y_t &= \phi_0 + \phi_1 \nabla^d Y_{t-1} + \cdots + \phi_p \nabla^d Y_{t-p} + a_t \\ &+ \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \quad t = \dots, -2, -1, 0, 1, 2, \dots \end{aligned} \quad (12)$$

where $\nabla = (1-B)$ is the first difference operator and the roots of the autoregressive and moving-average polynomials satisfy the usual stationary and invertibility conditions, respectively. The innovations $a_t = \sum_{j=0}^{\infty} \pi_j \nabla^d Y_{t-j}$ can be approximated by

$$a_t(\phi, \theta) = \sum_{j=0}^{t-1} \pi_j \nabla^d Y_{t-j} = \sum_{j=0}^{t-1} \beta_j(\theta) \left(\nabla^d Y_{t-j} - \sum_{i=1}^p \phi_i \nabla^d Y_{t-j-i} \right),$$

where π_j are the parameters of the infinite autoregressive representation of the stationary process $\nabla^d Y_t$ and

$$\sum_{j=0}^{\infty} \beta_j(\theta) z^j = \left(1 + \sum_{i=1}^q \theta_i z^i \right)^{-1}, \quad |z| \leq 1,$$

where $\theta = (\theta_1, \dots, \theta_q)$. If y_{T-n+1}, \dots, y_T are observations of the ARIMA(p, d, q) process and $\hat{\phi}$ and $\hat{\theta}$ are estimates of ϕ and θ , obtained using the stationary transformation $\nabla^d Y_t$, define $\hat{a}_t = a_t(\hat{\phi}, \hat{\theta})$, $t = T-n+1, \dots, T$. We have that, for t large enough, $\hat{a}_t - a_t$ tends to zero in probability as the sample size n goes to infinity since $|\hat{a}_t - a_t| \leq |\hat{a}_t - a_t(\phi, \theta)| + |a_t(\phi, \theta) - a_t|$, the first term goes to zero in probability as $n \rightarrow \infty$ (Kreiss and Franke, 1992; Lemmas 2.1 and 2.2) and the second one tends to zero in probability when the sample size tends to infinity, due to the invertibility condition (Brockwell and Davis, 1991).

The bootstrap prediction strategy proceeds now by adapting the five steps described in Section 2 to model (12). For simplicity of the exposition, we first consider the stationary ARMA model and, then, describe how to deal with the general ARIMA model. In particular, if $d = 0$, in step 2, instead of equation (9), the bootstrap series used to obtain bootstrap estimates of the parameters is generated by the following recursion

$$\begin{aligned} Y_t^* &= \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1}^* + \cdots + \hat{\phi}_p Y_{t-p}^* + \hat{a}_t^* \\ &+ \hat{\theta}_1 \hat{a}_{t-1}^* + \cdots + \hat{\theta}_q \hat{a}_{t-q}^*, \quad t = T-n+1, \dots, T, \end{aligned} \quad (13)$$

and, in step 3, the bootstrap future value is obtained by

$$Y_{T+k}^* = \hat{\phi}_0^* + \sum_{j=1}^p \hat{\phi}_j^* Y_{T+k-j}^* + \hat{a}_{T+k}^* + \sum_{j=1}^q \hat{\theta}_j^* \hat{a}_{T+k-j}^*, \quad (14)$$

where $Y_{T+h}^* = y_{T+h}$ and $\hat{a}_{T+h}^* = \hat{a}_{T+h}$ if $h \leq 0$. Note that, once more, the last p values of the series are fixed in (14) but not in (13).

If $d \neq 0$, equation (9) is replaced by the appropriate recursion. For example, for the ARI(1,1) model,

$$Y_t^* = \hat{\phi}_0 + (1 + \hat{\phi}_1)Y_{t-1}^* - \hat{\phi}_1 Y_{t-2}^* + \hat{a}_t^*, \quad t = 1, \dots, T. \quad (15)$$

Then, equation (10) is replaced by the following recursions,

$$\begin{aligned} Y_{T+1}^* &= \hat{\phi}_0^* + (1 + \hat{\phi}_1^*)Y_T - \hat{\phi}_1^* Y_{T-1} + \hat{a}_{T+1}^*, \\ Y_{T+2}^* &= \hat{\phi}_0^* + (1 + \hat{\phi}_1^*)Y_{T+1}^* - \hat{\phi}_1^* Y_T + \hat{a}_{T+2}^*, \end{aligned} \quad (16)$$

and so on.

To analyse the asymptotic properties of the proposed bootstrap procedure, we need the following definitions. Let Y_{T+k} be an observation of the law P conditional on \mathbf{y}_T and let Y_{T+k}^* be a bootstrap observation with distribution P^* conditional on \mathbf{y}_T . We say that Y_{T+k}^* converges weakly (P^*) in probability (P) to Y_{T+k} if for any distance d metrizing weak convergence, $d(Y_{T+k}^*, Y_{T+k}) \xrightarrow{P} 0$. We also say that Y_{T+k}^* converges in probability (P^*) in probability (P) to Y_{T+k} if for any distance d metrizing convergence in probability, $d(Y_{T+k}^*, Y_{T+k}) \xrightarrow{P} 0$. Finally, Y_{T+k}^* converges weakly (P^*) almost surely (P) to Y_{T+k} if for any distance d metrizing weak convergence, $d(Y_{T+k}^*, Y_{T+k}) \xrightarrow{0}$ for almost all sample sequences. The validity of the proposed method is established in the following theorem.

THEOREM 1. *Let $\mathbf{y}_T = \{y_T, y_{T+1}, \dots, y_T\}$ be a realization of an ARIMA(p, d, q) process $\{Y_t\}$ with $E(a_t) = 0$ and $E(a_t^4) < \infty$ and the roots of the autoregressive and moving-average polynomials satisfying the usual stationary and invertibility conditions respectively. Let $(\hat{\phi}, \hat{\theta})$ be any M-estimate of (ϕ, θ) and let Y_{T+k}^* be obtained following steps 1 to 5. Then, given \mathbf{y}_T , Y_{T+k}^* converges weakly in probability to Y_{T+k} as n tends to infinity.*

PROOF OF THEOREM 1. Assuming, for simplicity, that the integration parameter d is zero, weak convergence in probability of the bootstrap M-estimates, $(\hat{\phi}^*, \hat{\theta}^*)$, to the true parameters (ϕ, θ) , follows from Theorem 4.1 in Kreiss and Franke (1992). We express Y_{T+k}^* as a sum involving the available and fixed values y_T, y_{T+1}, \dots, y_T , the independent random draws \hat{a}_{T+j}^* , estimated innovations \hat{a}_T and continuous functions of the bootstrap parameter estimates $(\hat{\phi}^*, \hat{\theta}^*)$. For a forecast horizon k , we have that

$$\begin{aligned} Y_{T+k}^* &= g_0(\hat{\phi}^*) + g_1(\hat{\phi}^*)y_T + \dots + g_p(\hat{\phi}^*)y_{T-p+1} + h_1(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_{T+1}^* \\ &\quad + \dots + h_{k-1}(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_{T+k-1}^* + \hat{a}_{T+k}^* + l_1(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_T + \dots + l_q(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_{T-q+1}. \end{aligned}$$

The functions g_p , h_j and l_j are different for each prediction horizon, but for simplicity we use the same notation.

Following the arguments in Thombs and Schucany (1990), we have that $g_0(\hat{\phi}^*)$ converges weakly in probability to $g_0(\phi)$ and the products of fixed values y_T, y_{T-j} and functions of the bootstrap parameter estimates $g_j(\hat{\phi}^*)y_{T-j+1}$ converge also weakly

in probability to $g_j(\phi)Y_{T-j+1}$. Now, by Theorem 3.1 in Kreiss and Franke (1992), the terms \hat{a}_{T+j}^* tend weakly to a_{T+j} in probability and the products $h_j(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_{T+j}^*$ converge in distribution in probability to $h_j(\phi, \theta)a_{T+j}$, by the bootstrap version of Slutsky's Theorem. Since the \hat{a}_{T+j}^* 's are independent, the sum containing them and the continuous functions h_j s converge weakly in distribution in probability to the corresponding limit sum. Finally, the remaining term $l_1(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_T + \dots + l_q(\hat{\phi}^*, \hat{\theta}^*)\hat{a}_{T-q+1}$ can be rewritten as

$$\begin{aligned} & l_1(\phi, \theta)a_T + \dots + l_q(\phi, \theta)a_{T-q+1} + \{l_1(\hat{\phi}^*, \hat{\theta}^*) - l_1(\phi, \theta)\}\hat{a}_T \\ & + \dots + \{l_q(\hat{\phi}^*, \hat{\theta}^*) - l_q(\phi, \theta)\}\hat{a}_{T-q+1} + l_1(\phi, \theta)(\hat{a}_T - a_T) \\ & + \dots + l_q(\phi, \theta)(\hat{a}_{T-q+1} - a_{T-q+1}). \end{aligned}$$

The first q terms do not depend on n . The next q terms converge weakly to zero in probability since $\{l_j(\hat{\phi}^*, \hat{\theta}^*) - l_j(\phi, \theta)\}$ converges to zero in probability and the elements \hat{a}_{T-j+1} converge to a_{T-j+1} in distribution. Finally, the last q terms go to zero in probability since $l_j(\phi, \theta)$ is a fixed value and $(\hat{a}_{T-j+1} - a_{T-j+1})$ tends to zero in probability. It follows that for any fixed forecast horizon k , Y_{T+k}^* converges to Y_{T+k} weakly in probability, as n tends to infinity.

In the ARIMA(p, d, q) model with $d \neq 0$, the parameters (ϕ, θ) can be estimated from the stationary transformation, $\nabla^d Y_t$. Taking into account that the ARIMA(p, d, q) model can be seen as an ARMA($p+d, q$) model where the coefficients of the new autoregressive part (of order $p+d$) are continuous functions of the $p+1$ autoregressive parameters ϕ the proof concludes following the same lines as before. QED

Notice that, if the model does not contain a moving-average component, it is possible to obtain ordinary least squares (OLS) estimates of the autoregressive parameters ϕ . In this case, if the innovations satisfy the weaker condition $E(|a_t|^\gamma) < \infty$, for some $\gamma > 2$, Freedman (1985) proved the consistency of the OLS bootstrap estimates in conditional probability for almost all sample sequences. Then, following the same arguments as in the proof of Theorem 1, it can be shown that, if Y_{T+k}^* is obtained following steps 1 to 5, Y_{T+k}^* converges weakly almost surely to Y_{T+k} as n tends to infinity.

Finally, notice that the proposed bootstrap procedure is valid for any estimator that satisfies weak convergence in probability of bootstrap estimates to the true parameters.

4. SIMULATION RESULTS

The coverage of prediction intervals for finite samples is usually different from the asymptotic nominal coverage and depends on the model, the distribution of the innovations and the parameter estimation method. In this section, we present several Monte Carlo experiments carried out to analyse the finite sample

behaviour of the proposed bootstrap estimates of prediction densities for ARIMA(p,d,q) processes. We compare our proposal Pascual, Romo and Ruiz (PRR) with Box and Jenkins (BJ) intervals and with alternative bootstrap intervals. For stationary AR(p) processes, we compare PRR intervals with intervals introduced by Thombs and Schucany (1990) (TS), Breidt *et al.* (1995) (BDD) and Kabaila (1993) (KAB). For integrated autoregressive models, we compare PRR intervals with BJ intervals and with intervals constructed following García-Jurado *et al.* (1995) (GGP). Finally, the behaviour of our technique is analysed in forecasting future values of MA(q) models. As far as we know, the prediction density of MA(q) models has not been previously estimated by bootstrap methods; therefore, we only present PRR and BJ prediction intervals.

To study the different prediction intervals, we consider their coverage and length, and the proportion of observations lying out to the left and to the right. We compare these measures with those corresponding to the empirical prediction distribution obtained for a particular series generated by a specified process, sample size and error distribution F_a , generating $R = 1000$ future values y_{T+k} from that series. Then, for that particular series and for each of the methods considered, we obtain a $100\alpha\%$ prediction interval denoted by (L^*, U^*) (based on $B = 1000$ replicates in the case of bootstrap intervals) and estimate the conditional coverage for each procedure by

$$\hat{\alpha}^* = \frac{\#(L^* \leq y_{T+k}^r \leq U^*)}{R},$$

where y_{T+k}^r ($r = 1, \dots, R$) are the values generated previously. We have carried out 1000 Monte Carlo experiments and report average coverage, average length and average proportion of observations on the left and on the right for each method and for the empirical distribution.

We consider the following models:

- (i) $y_t = 1.75y_{t-1} - 0.76y_{t-2} + a_t$
- (ii) $\nabla^2 y_t = 0.5\nabla^2 y_{t-1} + a_t$
- (iii) $y_t = a_t - 0.3a_{t-1} + 0.7a_{t-2}$
- (iv) $y_t = 0.7y_{t-1} + a_t - 0.3a_{t-1}$,

where the innovations distribution F_a is normal, exponential or a contaminated distribution $0.9F_1 + 0.1F_2$, with $F_1 \sim N(-1,1)$ and $F_2 \sim N(9,1)$. Each distribution has been centered to have zero mean. The sample sizes considered are 25, 50 and 100, the prediction horizons are $k = 1$ and 3, and we construct intervals with nominal coverage α equal to 0.80 and 0.95. Results of some selected experiments appear in Tables I to VI. Results for other cases are available from the authors upon request. In Tables I and II, corresponding to model (i), it can be observed that the behaviour of all bootstrap prediction intervals, except for Kabaila (1993), is rather similar for all horizons, when nominal coverages and distributions are considered. The intervals constructed by Kabaila's method, although asymptotically correct, are, in general, too wide for moderate sample sizes.

TABLE I
MONTE CARLO RESULTS FOR THREE STEP AHEAD PREDICTIONS OF MODEL $Y_t = 1.75Y_{t-1} + 0.76Y_{t-2} + a_t$ WITH GAUSSIAN INNOVATIONS

Sample size	Method	Average coverage (SE)	Coverage (below/above)	Average length (SE)
n	Empirical	80%	10%/10%	7.83
25	BJ	70.01 (0.13)	15.8/14.2	7.31 (1.54)
	TS	70.14 (0.13)	15.8/14.1	7.28 (1.67)
	BDD	72.73 (0.11)	14.4/12.9	7.44 (1.52)
	KAB	63.68 (0.19)	18.8/17.5	10.3 (9.70)
	PRR	73.31 (0.14)	13.9/12.8	8.07 (2.43)
50	BJ	75.67 (0.08)	12.3/12.0	7.60 (1.04)
	TS	75.22 (0.07)	12.6/12.2	7.49 (1.12)
	BDD	76.26 (0.07)	12.0/11.7	7.68 (1.15)
	KAB	66.38 (0.12)	17.5/16.1	7.70 (3.16)
	PRR	76.92 (0.08)	11.7/11.3	7.83 (1.27)
100	BJ	78.03 (0.05)	10.7/11.3	7.74 (0.73)
	TS	77.64 (0.05)	10.8/11.6	7.70 (0.78)
	BDD	77.98 (0.05)	10.7/11.3	7.74 (0.79)
	KAB	67.67 (0.07)	15.9/16.4	6.86 (1.10)
	PRR	78.29 (0.05)	10.6/11.1	7.80 (0.79)

TABLE II
MONTE CARLO RESULTS FOR THREE STEP AHEAD PREDICTIONS OF MODEL $Y_t = 1.75Y_{t-1} + 0.76Y_{t-2} + a_t$ WITH CONTAMINATED INNOVATIONS

Sample size	Method	Average coverage (SE)	Coverage (below/above)	Average length (SE)
n	Empirical	95%	2.5%/2.5%	34.05
25	BJ	87.46 (0.11)	2.3/10.2	34.27 (10.8)
	TS	86.02 (0.14)	7.1/6.8	33.99 (10.6)
	BDD	88.53 (0.13)	5.7/5.8	35.33 (10.5)
	KAB	78.65 (0.20)	12.9/8.4	48.20 (49.2)
	PRR	87.75 (0.13)	6.25/6.0	37.75 (15.3)
50	BJ	91.03 (0.08)	0.83/8.14	36.34 (7.52)
	TS	89.46 (0.11)	6.3/4.2	34.73 (7.07)
	BDD	90.67 (0.10)	5.5/3.8	35.45 (6.92)
	KAB	82.73 (0.14)	10.5/6.7	37.60 (16.9)
	PRR	91.13 (0.09)	4.9/4.0	36.57 (8.30)
100	BJ	92.74 (0.04)	0.1/7.16	37.14 (5.35)
	TS	92.57 (0.07)	4.2/3.2	34.93 (4.85)
	BDD	92.86 (0.07)	4.0/3.1	35.06 (4.71)
	KAB	85.49 (0.10)	8.5/6.0	32.69 (7.24)
	PRR	93.03 (0.06)	3.8/3.2	35.54 (5.16)

When looking at the results for Gaussian innovations, we may see that although the BJ intervals are built assuming the correct error distribution, bootstrap intervals have better properties for a 80% nominal coverage. This may be due to the fact that BJ intervals do not incorporate the variability due to parameter estimation and to the well-known good bootstrap behaviour for small samples. Moreover, we may observe that the BJ intervals have worse coverage properties than PRR intervals when forecasting three periods ahead. Notice that when model parameters are estimated, the distribution of the forecasting errors is not

TABLE III
MONTE CARLO RESULTS FOR PREDICTIONS OF MODEL (1 $B)^2(1 - 0.5B)Y_t$ a_t WITH GAUSSIAN INNOVATIONS

Lead item	Sample size	Method	Average coverage (SE)	Coverage (below/above)	Average length (SE)		
1	n	Empirical	95%	2.5%/2.5%	3.92		
		25	BJ	93.25 (0.04)	3.44/3.31	3.89 (0.59)	
		GGP	91.39 (0.06)	4.43/4.2	3.81 (0.68)		
	50	PRR	91.63 (0.05)	4.3/4.06	3.82 (0.68)		
		BJ	94.22 (0.03)	2.9/2.9	3.92 (0.41)		
		GGP	93.06 (0.04)	3.5/3.5	3.87 (0.54)		
	100	PRR	93.08 (0.04)	3.5/3.4	3.87 (0.54)		
		BJ	94.64 (0.02)	2.6/2.7	3.92 (0.29)		
		GGP	94.01 (0.03)	2.9/3.05	3.90 (0.39)		
	3	n	Empirical	95%	2.5%/2.5%	19.72	
			25	BJ	97.91 (0.04)	1.05/1.03	27.10 (5.52)
			GGP	90.93 (0.07)	4.5/4.5	18.76 (3.69)	
50		PRR	91.29 (0.06)	4.09/4.2	18.77 (3.61)		
		BJ	98.79 (0.02)	0.62/0.58	27.47 (3.84)		
		GGP	93.16 (0.04)	3.4/3.4	19.31 (2.64)		
100		PRR	93.21 (0.04)	3.4/3.4	19.32 (2.62)		
		BJ	99.16 (0.01)	0.41/0.43	27.66 (2.71)		
		GGP	94.04 (0.03)	2.9/3.04	19.50 (1.90)		
PRR		94.05 (0.03)	2.9/3.04	19.50 (1.89)			

TABLE IV
MONTE CARLO RESULTS FOR PREDICTIONS OF MODEL (1 $B)^2(1 - 0.5B)Y_t$ a_t WITH EXPONENTIAL INNOVATIONS

Lead item	Sample size	Method	Average coverage (SE)	Coverage (below/above)	Average length (SE)		
1	n	Empirical	80%	10%/10%	2.19		
		25	BJ	84.60 (0.10)	3.7/11.7	2.46 (0.68)	
		GGP	75.23 (0.15)	13.68/11.09	2.23 (0.62)		
	50	PRR	76.06 (0.14)	12.72/11.21	2.23 (0.61)		
		BJ	87.44 (0.07)	1.6/10.98	2.51 (0.47)		
		GGP	77.53 (0.12)	11.76/10.70	2.20 (0.42)		
	100	PRR	77.96 (0.11)	11.25/10.97	2.20 (0.42)		
		BJ	88.86 (0.03)	0.52/10.62	2.53 (0.35)		
		GGP	78.27 (0.10)	11.40/10.33	2.20 (0.31)		
	3	n	Empirical	80%	10%/10%	11.82	
			25	BJ	90.14 (0.08)	1.97/7.9	17.03 (5.10)
			GGP	74.07 (0.13)	14.02/11.91	11.46 (3.46)	
50		PRR	75.12 (0.13)	12.92/11.94	11.47 (3.16)		
		BJ	92.51 (0.04)	0.69/6.8	17.60 (3.51)		
		GGP	76.89 (0.10)	12.11/11.00	11.61 (2.15)		
100		PRR	77.37 (0.90)	11.55/11.07	11.63 (2.15)		
		BJ	93.54 (0.02)	0.17/6.3	17.86 (2.63)		
		GGP	78.15 (0.07)	11.42/10.42	11.74 (1.58)		
PRR		78.48 (0.07)	10.99/10.5	11.73 (1.58)			

normal even if the innovations are Gaussian. This is due to the fact that the predictors are linear combinations of products of asymptotically normal random variables which, in general, are non-normal. This is the reason why, even for

TABLE V
MONTE CARLO RESULTS FOR PREDICTIONS OF MODEL $Y_t = a_t - 0.3a_{t-1} + 0.7a_{t-2}$ WITH EXPONENTIAL INNOVATIONS

Lead item	Sample size	Method	Average coverage (SE)	Coverage (below/above)	Average length (SE)
1	n	Empirical	80%	10%/10%	2.19
		25	BJ	82.75 (0.13)	5.75/11.51
	50	PRR	76.24 (0.17)	12.7/11.1	2.36 (0.63)
		BJ	86.15 (0.09)	3.1/10.7	2.57 (0.51)
	100	PRR	76.60 (0.14)	12.8/10.6	2.26 (0.44)
		BJ	88.56 (0.05)	0.94/10.5	2.55 (0.36)
3	n	Empirical	80%	10%/10%	2.93
		25	BJ	83.44 (0.09)	5.81/10.75
	50	PRR	78.55 (0.11)	10.37/11.08	2.98 (0.74)
		BJ	84.48 (0.07)	4.97/10.54	3.28 (0.69)
	100	PRR	79.35 (0.09)	10.08/10.56	2.98 (0.58)
		BJ	85.15 (0.05)	4.31/10.53	3.23 (0.47)
		PRR	79.63 (0.07)	10.02/10.34	2.95 (0.40)

TABLE VI
MONTE CARLO RESULTS FOR PREDICTIONS OF MODEL $Y_t = a_t - 0.3a_{t-1} + 0.7a_{t-2}$ WITH CONTAMINATED INNOVATIONS

Lead item	Sample size	Method	Average coverage (SE)	Coverage (below/above)	Average length (SE)
1	n	Empirical	95%	2.5%/2.5%	12.56
		25	BJ	89.84 (0.07)	0.97/9.19
	50	PRR	89.60 (0.11)	5.45/4.95	12.83 (3.10)
		BJ	90.31 (0.03)	0.18/9.50	12.35 (2.28)
	100	PRR	91.72 (0.08)	3.9/4.4	12.65 (1.93)
		BJ	90.20 (0.02)	0.01/9.8	12.37 (1.59)
3	n	Empirical	95%	2.5%/2.5%	14.80
		25	BJ	91.34 (0.07)	0.98/7.68
	50	PRR	91.99 (0.09)	3.67/4.35	15.19 (3.92)
		BJ	92.30 (0.05)	0.37/7.33	15.75 (3.06)
	100	PRR	94.08 (0.05)	2.78/3.14	15.30 (2.41)
		BJ	92.53 (0.03)	0.15/7.32	15.64 (2.09)
		PRR	94.70 (0.03)	2.60/2.70	15.16 (1.56)

Gaussian innovations, constructing bootstrap forecasting intervals could improve the forecast properties. Looking at Table II, which reports results for the contaminated distribution, we observe that the BJ intervals are too wide and still are not able to cope with the shape of the error distribution. This can be seen more clearly in Figure 1 that plots prediction densities of one-step ahead predictions for a particular series of size 100, estimated by our bootstrap procedure and by the BJ methodology together with the empirical density. The bootstrap density is obtained by applying a kernel density estimator of S-Plus with a rectangular box and a smoothing parameter of 1, to the bootstrap replicates of y_{T+1} , i.e. $\{y_{T+1}^{*(1)}, \dots, y_{T+1}^{*(999)}\}$. The empirical density is calculated using the same kernel estimator with the replicates of y_{T+1} generated for this particular

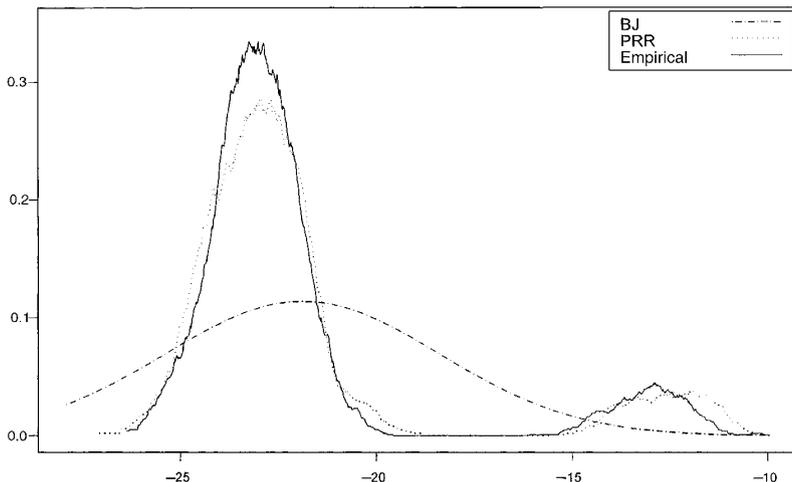


FIGURE 1. Densities of one step ahead predictions of one time series of size 100 generated by model $y_t = 1.75y_{t-1} - 0.76y_{t-2} + a_t$ with Contaminated innovations.

series. Finally, comparing the PRR, TS and BDD intervals, we may observe that for all distributions, sample sizes and coverages considered, the behaviour of the three methods is very similar. Our procedure does not work worse and, in some cases, seems to be slightly better than the others. The potential gains of PRR over TS and BDD could be due to the fact that the variance of the parameter estimates is reduced when the last p observations are not fixed to obtain bootstrap estimates of the parameters. Since our method is much simpler to implement and less computationally demanding than the other bootstrap methods, it seems to be an interesting alternative even for $AR(p)$ models.

Tables III and IV report results for model (ii) with Gaussian and exponential innovations respectively, comparing PRR forecast intervals with standard intervals and intervals built by the method proposed by García-Jurado *et al.* (1995). It is possible to observe that, even for Gaussian errors, standard intervals deteriorate very seriously when predicting three-steps ahead. As expected from results in the stationary case, the behaviour of standard intervals is even worse when the error distribution is not Gaussian. However, PRR and GGP intervals are very similar. Moreover, constructing GGP intervals requires solving a system which could be difficult to handle when forecasting far into the future, complicating the implementation of the method. Furthermore, in the cases considered in this study, PRR intervals slightly outperform the GGP intervals. The simulation results for $ARI(p,d)$ models are illustrated in Figure 2 where we represent one-step-ahead prediction densities estimated using the standard and PRR procedures together with the empirical density for a sample size of 100 and exponential innovations. Finally, as in the stationary case, we may observe that

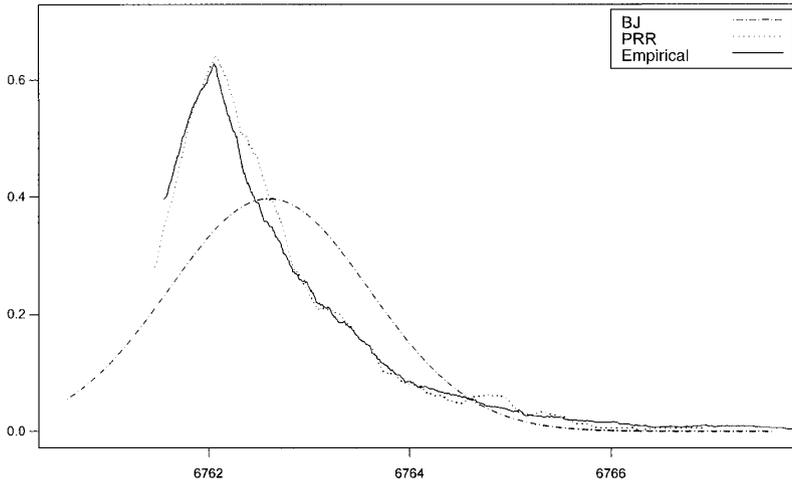


FIGURE 2. Densities of one step ahead predictions of one time series of size 100 generated by model $\Delta^2 y_t = 0.5\Delta^2 y_{t-1} + a_t$ with exponential innovations.

when the sample size increases, the average coverage and average length converge to the empirical values according to the results in Section 3.

Finally, Tables V and VI report the results of the Monte Carlo experiments to check the behaviour of our technique when forecasting processes with moving-average components by considering the MA(2) model in (iii) with exponential and contaminated innovations respectively. In this case, there are no alternative bootstrap methods proposed in the literature and we only compare our strategy with BJ intervals. To predict the future values of a moving-average process, we need estimates of the within-sample innovations. This is an additional source of uncertainty in forecasting MA processes which makes the construction of forecast intervals a more difficult task. However, our bootstrap method is easy to implement even in the presence of moving-average components and, as we will see, it works reasonably well. There are several alternatives to estimate the innovations of moving-average processes and, in this study, we consider the simplest one which consists in conditioning on the value of all innovations previous to the sample period being equal to their expected value zero. The estimation of the parameters is carried out by conditional quasi-maximum likelihood. The results of the Monte Carlo simulations are similar to those corresponding to the previous models. BJ intervals are not able to deal with asymmetric distributions. Figure 3 shows the standard and bootstrap densities together with the empirical density for one-step-ahead predictions built with a sample size of 100 and exponential innovations. It is clear that the BJ density does not mimic the empirical prediction distribution.

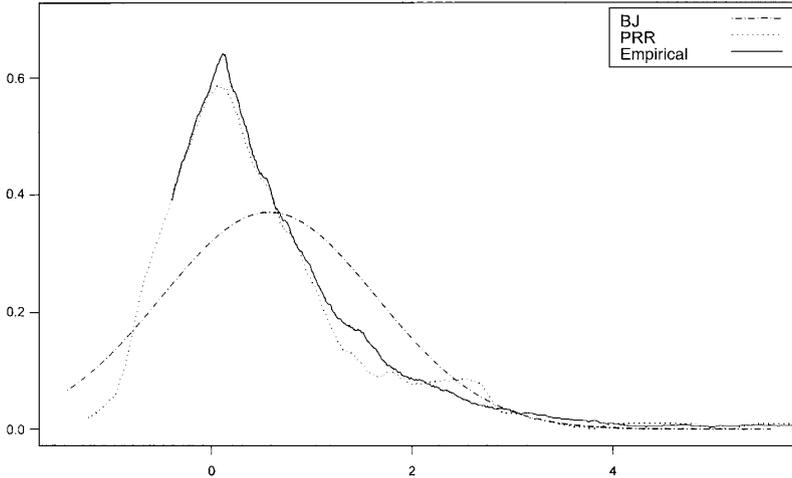


FIGURE 3. Densities of one step ahead predictions of one time series of size 100 generated by model $y_t = a_t - 0.3a_{t-1} + 0.7a_{t-2}$ with exponential innovations.

It is important to note that the results presented in this section have been obtained using OLS or conditional quasi-maximum likelihood estimates of the parameters. It seems clear that these results could be improved by using estimates more appropriate for non-normal innovations. The effects of the estimation method on bootstrap prediction densities is analysed by Pascual *et al.* (2001). Moreover, in moving-average models, the innovations are estimated conditioning on pre-sample values being zero. These estimates can be improved using the unconditional residuals, which can be obtained, for example, via the Kalman filter. Then, the resampling procedure can be applied to the unconditional residuals.

5. CONCLUSIONS

A new bootstrap approach to estimate the prediction density of ARIMA processes has been presented in this paper. The proposed bootstrap density is estimated directly from the bootstrap predictions and incorporates the uncertainty due to parameter estimation. The main advantage of this prediction resampling strategy with respect to previous bootstrap prediction methods with similar properties is that the backward representation of the process is not required to obtain bootstrap replicates of the series. Consequently, our method is flexible and easy to implement, allowing the generalization to models with moving-average components and also to processes without a backward

representation. We have established the asymptotic properties of the bootstrap prediction intervals and carried out Monte Carlo experiments to analyse their behaviour for finite samples. We have compared them with standard intervals as proposed by Box and Jenkins (1976) and with bootstrap intervals based on Thombs and Schucany (1990). The results of these experiments show that for non-normal innovations, Box and Jenkins prediction intervals can be heavily distorted. We have seen that all bootstrap intervals have rather similar properties for $ARI(p,d)$ processes and our intervals are slightly better in some cases. Moreover, our proposal is more flexible allowing for the construction of prediction densities for processes with moving-average components which cannot be handled by previous methods. Monte Carlo simulations show that the proposed bootstrap prediction intervals work well in forecasting future values of processes with moving-average components.

Finally, the flexibility of this method allows to extend the construction of prediction intervals even for models without a backward representation, such as GARCH models. Miguel and Olave (1998) study a bootstrap procedure for GARCH processes based on Cao *et al.* (1995), where the resampling is conditional on the parameter estimates. Currently, we are investigating the application of our bootstrap strategy to prediction densities of GARCH processes with very promising results.

ACKNOWLEDGEMENTS

We are grateful to an Associate Editor and an anonymous referee for helpful comments. Financial support from project BEC2002-03720 from the Spanish Government is gratefully acknowledged.

REFERENCES

- BOX, G. E. P. and JENKINS, G. M. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day.
- BREIDT, F. J., DAVIS, R. A. and DUNSMUIR, W. T. (1995) Improved bootstrap prediction intervals for autoregressions. *J. Time Ser. Anal.* 16, 177–200.
- BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*. New York: Springer Verlag.
- CAO, R., FEBRERO BANDE, M., GONZÁLEZ MANTEIGA, W., PRADA SÁNCHEZ, J. M. and GARCÍA JURADO, I. (1997) Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Commun. Statist. B* 26, 961–78.
- EFRON, B. (1982) The Jackknife, the Bootstrap and Other Resampling Plans. CMBS NSF Monographs 38, Philadelphia, PA: SIAM.
- FINDLEY, D. F. (1986) On bootstrap estimates of forecast mean square errors for autoregressive processes. In *Computer Science and Statistics: The Interface* (ed. D. M. ALLEN). Amsterdam: North Holland, pp. 11–7.
- FREEDMAN, D. A. (1985) On bootstrapping two stage least squares estimates in stationary linear models. *Ann. Statist.* 12, 827–42.

- GARCÍA JURADO, I., GONZÁLEZ MANTEIGA, W., PRADA SÁNCHEZ, J. M., FEBRERO BANDE, M. and CAO, R. (1995) Predicting using Box Jenkins, nonparametric and bootstrap techniques. *Technometrics* 37, 303 10.
- GRIGOLETTO, M. (1998) Bootstrap prediction intervals for autoregressions: some alternatives. *Int. J. Forecasting* 14, 447 56.
- KABAILA, P. (1993) On bootstrap predictive inference for autoregressive processes. *J. Time Ser. Anal.* 14, 473 84.
- KREISS, J. P. and FRANKE, J. (1992) Bootstrapping stationary autoregressive moving average models. *J. Time Ser. Anal.* 13, 297 317.
- MASAROTTO, G. (1990) Bootstrap prediction intervals for autoregressions. *Int. J. Forecasting* 6, 229 39.
- MCCULLOUGH, B. D. (1994) Bootstrapping forecast intervals: an application to AR(p) models. *J. Forecasting* 13, 51 66.
- MIGUEL, J. A. and OLAVE, P. (1998) Bootstrapping forecast intervals in ARCH models. *TEST* 18, 345 64.
- PASCUAL, L., ROMO, J. and RUIZ, E. (2001) Effects of parameter estimation on prediction densities: a bootstrap approach. *Int. J. Forecasting* 17, 83 103.
- STINE, R. A. (1987) Estimating properties of autoregressive forecasts. *J. Am. Statist. Assoc.* 82, 1072 8.
- THOMBS, L. A. and SCHUCANY, W. R. (1990) Bootstrap prediction intervals for autoregression. *J. Am. Statist. Assoc.* 85, 486 92.