

Bootstrap prediction intervals for power-transformed time series

Lorenzo Pascual^{a,1}, Juan Romo^{b,2}, Esther Ruiz^{b,*}



^a*HidroCantábrico, Serrano Galvache 56, Centro Empresarial Parque Norte, 28033 Madrid, Spain*

^b*Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain*

Abstract

In this paper, we propose a bootstrap procedure to construct prediction intervals for future values of a variable after a linear ARIMA model has been fitted to its power transformation. The procedure is easy to implement and provides a useful tool in empirical applications given that it is often the case that, for example, the log transformation is modeled when the variable of interest for prediction is the original one. The advantages over existing methods for computing prediction intervals of power transformed time series are that the proposed bootstrap intervals incorporate the variability due to parameter estimation and do not rely on distributional assumptions neither on the original variable nor on the transformed one. We derive the asymptotic distribution and show the good behavior of the bootstrap approach versus alternative procedures by means of Monte Carlo experiments. Finally, the procedure is illustrated by analyzing three real time series data sets.

Keywords: Forecasting; Non-Gaussian distributions; Box-Cox transformations; Resampling methods

1. Introduction

Forecasting future values of time series data is one of the main objectives of time series analysis. Generally, predictions are given as point forecasts, although it is even more important to provide

interval forecasts as well; see, for example, [Chatfield \(1993\)](#).

In empirical time series analysis, it is common to transform the data using power transformation prior to the estimation of the model used for forecasting. There are several reasons for transforming the data before fitting a suitable model, for example, the necessity of stabilizing the increasing variance of trending time series, to reduce the impact of outliers, to make the normal distribution a better approximation to the data distribution, or because the transformed variable has a convenient economic interpretation; for example, first differenced log-transformed data correspond to growth rates.

* Corresponding author. Tel.: +34 916249851; fax: +34 916249849.

E-mail addresses: lorenzo.pascualcaneiro@h-c.es (L. Pascual), romo@est-econ.uc3m.es (J. Romo), ortega@est-econ.uc3m.es (E. Ruiz).

¹ Tel.: +34 917819338; fax: +34 917819359.

² Tel.: +34 916249805; fax: +34 916249849.

The well-known family of Box–Cox transformations is given by

$$g(X_t) = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(X_t), & \lambda = 0, \end{cases} \quad (1)$$

where $\{X_t\}$ denotes the observed time series with $X_t > 0$, $\ln(\cdot)$ denotes the natural logarithm, and λ is a real constant. The transformation for $\lambda=0$ follows from the fact that $\lim_{\lambda \rightarrow 0} \frac{X_t^\lambda - 1}{\lambda} = \ln(X_t)$; see [Box and Cox \(1964\)](#). Subtracting 1 and dividing it by λ does not influence the stochastic structure of X_t^λ , and hence, without loss of generality, one often considers the following transformation suggested by [Tukey \(1957\)](#)

$$g(X_t) = \begin{cases} X_t^\lambda, & \lambda \neq 0 \\ \ln(X_t), & \lambda = 0, \end{cases} \quad (2)$$

instead of the Box–Cox transformation in Eq. (1).

Once a model has been estimated, point and interval forecasts can be obtained for the transformed series $Y_{T+k} = g(X_{T+k})$ as defined in Eq. (2) for $k=1, 2, \dots$. We will focus on ARIMA models fitted to the series y_t , $t=1, \dots, T$. The specification of the model and the parameter λ will be assumed to be known. Notice that this is an interesting case given that, in many empirical applications, the logarithmic transformation is assumed. If the objective is to predict future values of X_t , the retransformed point forecasts induce bias in the forecasts, as shown for linear models in [Granger and Newbold \(1976\)](#). When Y_{T+k} is normally distributed and the point forecast of X_{T+k} is just the inverse transformation of the forecast obtained for the transformed variable, this naive point prediction is not the minimum mean squared error (MMSE) forecast but the minimum mean absolute error (MMAE), which is the median of the conditional probability density function (pdf) of X_{T+k} . Therefore, if the error loss function is quadratic, this naive prediction of X_{T+k} is not optimal; see [Guerrero \(1993\)](#).

Assuming Gaussianity of Y_t , [Granger and Newbold \(1976\)](#) propose a debiasing factor to reduce the transformation bias in the point forecast. Unfortunately, inasmuch as they solve the problem using Hermite polynomial expansions, their procedure becomes very complicated for many fractional power transformations, making this approach not useful in practice. Later, [Taylor \(1986\)](#) proposes a simpler expression for the debiasing factor, but it does not provide an adequate

solution for $\lambda=0$. Notice that, as pointed out above, the logarithmic transformation is one of the most popular in practice. Another alternative proposed by [Pankratz and Dudley \(1987\)](#) is complicated and, additionally, only admits a closed form expression when λ is a fractional positive integer. Finally, the method proposed by [Guerrero \(1993\)](#) avoids all the drawbacks found in previous approaches. His proposal is both simple and general. In a comparative study, [Guerrero \(1993\)](#) shows that his method has a performance similar to or better than the other procedures.

Although it is relatively well studied how to obtain a good estimate for the conditional mean in the original metric, there is no generally accepted method of constructing prediction intervals for the untransformed variable. One solution is based on a normal assumption on X_{T+k} , providing a symmetric interval. This seems not to be a good choice unless the distribution of X_{T+k} is close to be Gaussian; see [Chatfield \(1993\)](#). Another alternative is to construct prediction intervals for X_{T+k} by retransforming the upper and lower values of the corresponding prediction interval for Y_{T+k} . Finally, [Guerrero \(1993\)](#) suggests to correct for bias the endpoints of the latter prediction intervals, using a procedure similar to the one he proposes for the point forecast.

In this paper, we propose a bootstrap resampling scheme to obtain an estimate of the pdf of X_{T+k} , conditional on the available data when an ARIMA model has been fitted to y_t , $t=1, \dots, T$. Given this density, the required prediction intervals for X_{T+k} can be constructed. There are several advantages over the methods previously described. First of all, the bootstrap procedure does not rely on distributional assumptions neither on the transformed data nor on the original scale. The second advantage is that the bootstrap intervals incorporate the variability due to parameter estimation, which is not allowed by any of the alternative procedures. Finally, the method is very easy to implement.

The finite sample behavior of the bootstrap intervals is compared with the alternative intervals by means of an extensive simulation study. It is shown that the proposed procedure performs as well as the best alternatives when Y_t is Gaussian and tends to outperform its competitors when leaving this assumption.

The paper is organized as follows. Section 2 presents a description of the existing methods for

obtaining prediction intervals for a variable in its original scale. In Section 3, we introduce the bootstrap approach. A Monte Carlo study comparing the finite sample performance of all existing methods is presented in Section 4. In Section 5, we illustrate the procedure analyzing empirically three real data sets. Finally, we conclude with some remarks and suggestions for future research in Section 6.

2. Prediction intervals for transformed time series

There are two main alternatives proposed in the literature to obtain prediction intervals for X_{T+k} given the observed series $x_t, t=1, \dots, T$ after an ARIMA model has been fitted to the power-transformed variable $y_t, t=1, \dots, T$. In this section, these two procedures are described.

Consider that $\{x_1, \dots, x_T\}$ is an available sequence of T observations such that, for any of the reasons previously mentioned, it needs to be transformed adequately by a function $g(\cdot)$ defined in Eq. (2) to obtain a new sequence y_1, \dots, y_T . Lets also assume that the transformed sequence is well fitted by an ARIMA(p, d, q) process given by

$$\phi(L)\nabla^d y_t = \phi_0 + \theta(L)a_t, \quad (3)$$

where a_t is a white noise process, $\phi(L)$ and $\theta(L)$ are autoregressive and moving average polynomials defined as $\phi(L)=1-\phi_1L-\dots-\phi_pL^p$ and $\theta(L)=1-\theta_1L-\dots-\theta_qL^q$, respectively, where L is the lag operator such that $Lx^h=x_{t-h}$, ∇ is the difference operator such that $\nabla=(1-L)$ and d is the number of differences needed for the series y_t to be stationary. The polynomials $\phi(L)$ and $\theta(L)$ satisfy the usual conditions to guarantee stationarity, invertibility and noncommon roots. From the transformed series $\{y_1, y_2, \dots, y_T\}$, the parameters of model (3) can be estimated by a consistent estimator, for example, conditional quasi-maximum likelihood (QML). Given the estimates $(\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$, the residuals are calculated by the following recursion

$$\begin{aligned} \hat{a}_t = & \nabla^d y_t - \hat{\phi}_0 - \hat{\phi}_1 \nabla^d y_{t-1} - \dots - \hat{\phi}_p \nabla^d y_{t-p} \\ & - \hat{\theta}_1 \hat{a}_{t-1} - \dots - \hat{\theta}_q \hat{a}_{t-q}, t = p+d+1, \dots, T, \end{aligned} \quad (4)$$

where the residuals corresponding to periods of time $t=p+d, p+d-1, \dots, 1, 0, -1, -2, \dots$ are set equal to 0; see, for example, [Harvey \(1993\)](#).

Once the ARIMA model has been estimated, the optimal linear predictor of $Y_{T+k}, k=1, 2, \dots$, denoted by $\hat{Y}_T(k)$, can be obtained in the usual way. If, for example, $d=0$, then the optimal predictor is given by

$$\begin{aligned} \hat{Y}_T(k) = & \hat{\phi}_0 + \hat{\phi}_1 \hat{Y}_T(k-1) + \hat{\phi}_2 \hat{Y}_T(k-2) + \dots \\ & + \hat{\phi}_p \hat{Y}_T(k-p) + \hat{\theta}_1 \hat{a}_{T+k-1} + \dots + \hat{\theta}_q \hat{a}_{T+k-q} \end{aligned} \quad (5)$$

where $\hat{Y}_T(j)=Y_{T+j}$ for $j \leq 0$ and $\hat{a}_{T+j}=0$ for $j \geq 0$. Alternatively, if for example $d=1$, then $\hat{Y}_T(k)$ is given by

$$\begin{aligned} \hat{Y}_T(k) = & \hat{\phi}_0 + (1 - \hat{\phi}_1) \hat{Y}_T(k-1) \\ & + (\hat{\phi}_2 - \hat{\phi}_1) \hat{Y}_T(k-2) + \dots \\ & + (\hat{\phi}_p - \hat{\phi}_{p-1}) \hat{Y}_T(k-p) \\ & + \hat{\theta}_1 \hat{a}_{T+k-1} + \dots + \hat{\theta}_q \hat{a}_{T+k-q}, \end{aligned} \quad (6)$$

Expressions of the optimal predictor for other values of d can be obtained similarly. The usual [Box and Jenkins \(1976\)](#) prediction intervals for Y_{T+k} are given by

$$\begin{aligned} \left[\hat{Y}_T(k) - z_{\alpha/2} \left(\hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j^2 \right)^{1/2}, \hat{Y}_T(k) \right. \\ \left. + z_{\alpha/2} \left(\hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j^2 \right)^{1/2} \right], \end{aligned} \quad (7)$$

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 $y_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}$, where the parameters ψ_i are the coefficients of the polynomial given by $\psi(L)=\theta(L)/\phi(L)\nabla^d$.

2.1. Symmetric prediction intervals

Multistep symmetric prediction intervals have been widely used in linear time series models. These intervals are constructed under the assumption of normality for the variable of interest. Therefore, they

provide a reasonable performance in terms of coverage and length if the density of the forecast error is well approximated by the normal distribution.

To obtain a symmetric prediction interval k periods into the future, a point forecast $\hat{X}_T(k)$ for X_{T+k} is needed first, usually corrected by a bias using one of the methods previously mentioned to compute the debiasing factor, and secondly, an explicit expression for the k -step ahead conditional mean squared error (MSE), say $V_c(k)$. Then, assuming normality, it follows that the conditional distribution of X_{T+k} given the available data is normal with mean estimated by $\hat{X}_T(k)$ and MSE estimated by $\hat{V}_c(k)$. In such a case, the k -step ahead prediction interval is given by

$$\left[\hat{X}_T(k) - z_{\frac{\alpha}{2}} \hat{V}_c(k)^{\frac{1}{2}}, \hat{X}_T(k) + z_{\frac{\alpha}{2}} \hat{V}_c(k)^{\frac{1}{2}} \right]. \quad (8)$$

The expression of $V_c(k)$ given by Granger and Newbold (1976) heavily depends on the Gaussian assumption for the series Y_t . Furthermore, this expression is derived by using Hermite polynomials, and it is not easy to obtain for a general transformation $g(\cdot)$. In fact, Granger and Newbold (1976) only give the final expression of $V_c(k)$ for the logarithmic and square root transformations. For example, if the logarithmic transformation is considered, $\hat{V}_c(k)$ is given by $\exp \{2\hat{Y}_T(k) + \hat{\sigma}^2(k)\} [\exp \{\hat{\sigma}^2(k)\} - 1]$, where $\hat{\sigma}^2(k) = \hat{\sigma}_a^2 \sum_{j=0}^{k-1} \hat{\psi}_j^2$.

Furthermore, the prediction intervals in Eq. (8) ignore the skewness and all higher moments in the distribution of the forecast error by assuming that it is approximately normal and therefore will only be accurate if the corresponding forecast error is approximately normally distributed.

Notice that usually, Y_t is assumed to be normally distributed, and consequently, the untransformed variable X_t will be not normally distributed unless the parameter λ in the Box–Cox transformation is equal to 1; that is, X_t is not transformed.

2.2. Naive prediction intervals

Alternatively, prediction intervals for the variable in the original scale can be constructed by retransforming the upper and lower values of the corresponding prediction intervals for the transformed variable Y_t given by Eq. (7). If the prediction interval for Y_t has a prescribed probability, say $(1-\alpha)$, then the

retransformed prediction interval for X_t should have the same prescribed probability; see Harvey (1989). The corresponding prediction interval with nominal coverage of $1-\alpha$ is given by

$$\left[g^{-1} \left\{ \hat{Y}_T(k) - z_{\frac{\alpha}{2}} \hat{\sigma}(k) \right\}, g^{-1} \left\{ \hat{Y}_T(k) + z_{\frac{\alpha}{2}} \hat{\sigma}(k) \right\} \right]. \quad (9)$$

Additionally, as proposed by Guerrero (1993), it is possible to correct for bias in the previous confidence interval by multiplying the end points of Eq. (9) by the following debiasing factor

$$C_{\lambda}(k) \begin{cases} \left\{ 0.5 + 0.5 [1 + 2(\lambda^{-1} - 1)\sigma^2(k)/\hat{Y}_T^2(k)]^{1/2} \right\}^{1/\lambda}, & \lambda \neq 0 \\ \exp(\sigma^2(k)/2), & \lambda = 0. \end{cases} \quad (10)$$

Notice that the prediction intervals in Eq. (9) are able to cope with the potential asymmetry of the distribution of X_t , although they still rely on the Gaussianity assumption for the transformed variable, Y_t , and do not incorporate the uncertainty due to parameter estimation.

3. The bootstrap approach

In this section, we describe a bootstrap procedure to obtain prediction densities and prediction intervals of future values of the series of interest, X_t . The resampling scheme is based on the proposal by Pascual, Romo, and Ruiz (2004) to estimate prediction densities and intervals of series generated by ARIMA(p, d, q) processes.

Denoted by \hat{F}_a , the empirical distribution function of the centered residuals of the ARIMA model for y_t given in Eq. (4). Given the $p+d$ initial values of the variable y_t , say $\{y_1, \dots, y_{p+d}\}$, a bootstrap replicate of the transformed series $\{y_1^*, \dots, y_T^*\}$ is constructed by the following equation

$$\nabla^d y_t^* = \hat{\phi}_0 + \sum_{j=1}^p \hat{\phi}_j \nabla^d y_{t-j}^* + \sum_{j=1}^q \hat{\theta}_j \hat{a}_{t-j}^* + \hat{a}_t^*, \quad t = p + d + 1, \dots, T, \quad (11)$$

where $y_t^* = y_t$, $t=1, \dots, p+d$ and $\hat{a}_{1+p+d-q}^*, \dots, \hat{a}_T^*$ are random draws from \hat{F}_a . Once the parameters of this bootstrap series are estimated, say $(\hat{\phi}_0^*, \hat{\phi}_1^*, \dots, \hat{\phi}_p^*,$

$\hat{\theta}_1^*, \dots, \hat{\theta}_q^*$), the bootstrap forecast k steps ahead is obtained as follows,

$$\begin{aligned} \nabla^d y_{T+k}^* &= \hat{\phi}_0^* + \sum_{j=1}^p \hat{\phi}_j^* \nabla^d y_{T+k-j}^* + \sum_{j=1}^q \hat{\theta}_j^* \hat{a}_{T+k-j}^* \\ &+ \hat{a}_{T+k}^*, k = 1, 2, \dots \end{aligned} \quad (12)$$

where $y_{T+k-j}^* = y_{T+k-j}$, $j \geq k$ and $\hat{a}_{T+k-j}^* = \hat{a}_{T+k-j}$, $j \geq k$; that is, the last $p+d$ observations of the series and the last q residuals are fixed in order to obtain the prediction density conditional on the observed data. Finally, in expression (12), \hat{a}_{T+k-j}^* , $j < k$ are random draws from \hat{F}_a .

Once B bootstrap replicates of Y_{T+k}^* are obtained, it is possible to construct a bootstrap estimate of the distribution of Y_{T+k} conditional on $\{y_1, \dots, y_T\}$ and the corresponding prediction intervals. Pascual et al. (2004) prove that, for the transformed series $\{y_t\}$, $Y_{T+k}^* \rightarrow Y_{T+k}$ in conditional probability, in probability, as the sample size T goes to infinity. They also show that the finite sample properties of the bootstrap procedure just described outperforms other alternative bootstrap mechanisms proposed to compute prediction intervals in stationary AR(p) models.

However, in this paper, the objective is to estimate the distribution of X_{T+k} conditional on $\{x_1, \dots, x_T\}$. In this case, a new step has to be introduced in the described procedure. The bootstrap forecast k steps ahead for the variable in the original metric is then obtained as

$$x_{T+k}^* = g^{-1}(y_{T+k}^*), k = 1, 2, \dots \quad (13)$$

This procedure is repeated B times to obtain a set of B bootstrap replicates for X_{T+k} , say $(x_{T+k}^{*(1)}, \dots, x_{T+k}^{*(B)})$. Then, the prediction limits are defined as the quantiles of the bootstrap distribution function of X_{T+k}^* ; that is, if $G^*(h) = \Pr(X_{T+k}^* \leq h)$ is the distribution function of X_{T+k}^* and its Monte Carlo estimate is $G_B^*(h) = \#(x_{T+k}^{*(b)} \leq h)/B$, a $100(1-\alpha)\%$ prediction interval for X_{T+k}^* is given by

$$[L_B^*, U_B^*] = \left[Q_B^* \left(\frac{\alpha}{2} \right), Q_B^* \left(1 - \frac{\alpha}{2} \right) \right], \quad (14)$$

where $Q_B^* = G_B^{*^{-1}}$.

Before summarizing the steps for obtaining bootstrap prediction densities and intervals for X_{T+k} , we illustrate the method with a simple example. Suppose that after taking an adequate power transformation, the sequence $\{y_1, \dots, y_T\}$ follows an ARIMA(0, 1, 2) model without constant term, i.e.

$$\nabla y_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}. \quad (15)$$

Once the parameters of model (15) have been estimated and the bootstrap draws $\hat{a}_0^*, \hat{a}_1^*, \dots, \hat{a}_T^*$ are available, a bootstrap replicate of the transformed series is constructed by

$$y_t^* = y_{t-1}^* + \hat{a}_t^* + \hat{\theta}_1 \hat{a}_{t-1}^* + \hat{\theta}_2 \hat{a}_{t-2}^*, t = 2, \dots, T, \quad (16)$$

where $y_1^* = y_1$. Then, bootstrap estimates $\hat{\theta}_1^*$ and $\hat{\theta}_2^*$ are obtained for the bootstrap series, and bootstrap replicates of future values of the transformed series are generated by

$$y_{T+1}^* = y_T + \hat{a}_{T+1}^* + \hat{\theta}_1^* \hat{a}_T^* + \hat{\theta}_2^* \hat{a}_{T-1}^*$$

$$y_{T+2}^* = y_{T+1}^* + \hat{a}_{T+2}^* + \hat{\theta}_1^* \hat{a}_{T+1}^* + \hat{\theta}_2^* \hat{a}_T^*$$

$$y_{T+3}^* = y_{T+2}^* + \hat{a}_{T+3}^* + \hat{\theta}_1^* \hat{a}_{T+2}^* + \hat{\theta}_2^* \hat{a}_{T+1}^*.$$

It is important to note that inasmuch as the predictions are conditional on the sample information available at time T , in the recursions above, \hat{a}_T and \hat{a}_{T-1} are kept fixed in the different bootstrap replicates of y_{T+1}^* and y_{T+2}^* while \hat{a}_{T+1}^* , \hat{a}_{T+2}^* and \hat{a}_{T+3}^* change from one replicate to another. Finally, bootstrap replicates of future values of the series in the original scale are generated by expression (13).

Now, we summarize all the steps needed for obtaining bootstrap prediction intervals for X_{T+k} .

Step 1. Compute the residuals \hat{a}_t as in Eq. (4) for the transformed series. Let \hat{F}_a be the empirical distribution function of the centered residuals.

Step 2. Generate a bootstrap series using the recursion in Eq. (11) and calculate the estimates $(\hat{\phi}_0^*, \hat{\phi}_1^*, \dots, \hat{\phi}_p^*, \hat{\theta}_1^*, \dots, \hat{\theta}_q^*)$.

Step 3. Obtain a bootstrap future value for the transformed series by expression (12). Note that the last $p+d$ values of the transformed series and the final

q residuals are fixed in this step but not in the previous one.

Step 4. Obtain a bootstrap future value for the series in the original scale by expression (13).

Step 5. Repeat the last four steps B times and then go to Step 6.

Step 6. The endpoints of the prediction interval are given by quantiles of G_B^* , the bootstrap distribution function of X_{T+k}^* , given by expression (14).

Alternatively, the bootstrap procedure just described could be also applied to construct prediction intervals conditional on the parameter estimates; hereafter CB (conditional bootstrap). This procedure has been previously proposed by [Cao, Febrero-Bande, González-Manteiga, Prada-Sánchez and García-Jurado \(1997\)](#) for series $\{x_1, \dots, x_T\}$ following an AR(p) processes, and has been generalized by [Pascual, Romo, and Ruiz \(2001\)](#) for the general ARIMA(p, d, q) processes. With this method, the parameters are estimated once, and these estimates are used in the calculation of all bootstrap forecasts x_{T+k}^* . The steps to obtain bootstrap forecasts are similar to those presented above, except that Step 2 is avoided inasmuch as now it is not necessary to generate bootstrap replicates of the transformed series. Then, the expression to obtain bootstrap future values for the transformed series in Step 3 is replaced by

$$\nabla^d y_{T+k}^* = \hat{\phi}_0 + \sum_{j=1}^p \hat{\phi}_j \nabla^d y_{T+k-j}^* + \sum_{j=1}^q \hat{\theta}_j \hat{a}_{T+k-j}^* + \hat{a}_{T+k}^*, \quad k = 1, 2, \dots,$$

where y_{T+k-j}^* and \hat{a}_{T+k-j}^* are defined as in Eq. (12). Inasmuch as the parameter estimates are kept fixed in all bootstrap replicates of future values, the CB prediction intervals do not incorporate the uncertainty due to parameter estimation.

Notice that the estimated bootstrap density of X_{T+k} can also be used to obtain a bootstrap estimate of the expected value and/or the median of X_{T+k} conditional on the available series. These estimates can then be taken as point forecasts of X_{T+k} .

Finally, using the asymptotic results in [Pascual et al. \(2004\)](#) and inasmuch as $g(\cdot)$ is a known

continuous invertible function, it is straightforward to prove, using the bootstrap version of the Continuity Theorem, that $g^{-1}(Y_{T+k}^*) \rightarrow g^{-1}(Y_{T+k})$, i.e., $X_{T+k}^* \rightarrow X_{T+k}$, in conditional probability, in probability, as $T \rightarrow \infty$.

4. Small sample properties

4.1. Monte Carlo design

We now describe the results of several Monte Carlo experiments carried out to study the finite sample performance of the prediction intervals built by the alternative procedures considered in this paper. Prediction intervals built by the proposed bootstrap procedure (PRR) are compared with CB intervals and with the nonbootstrap methods described in Section 2. As previously mentioned, PRR is the only approach that does not condition on parameter estimates and, consequently, introduces the variability due to parameter estimation in the intervals. Comparing PRR with CB intervals, we are studying the effect of parameter estimation variability on the shape of estimated prediction densities. The basic symmetric prediction intervals in Eq. (8) will be denoted hereafter by STD1, the intervals based on retransforming the ends in Eq. (9) will be denoted by STD2, and finally, the corrected by bias prediction intervals using Eq. (10) will be denoted by STD3.

The focus of the simulation experiments is on prediction of future values of a series X_t , such that a linear ARIMA(p, d, q) model is fitted to a power transformation of its original values, say y_t . We consider the following three ARIMA processes for the transformed variable

$$y_t = 0.95y_{t-1} + a_t, \quad \sigma_a^2 = 0.1 \quad (17)$$

$$y_t = 0.7y_{t-1} + a_t - 0.3a_{t-1}, \quad \sigma_a^2 = 0.5 \quad (18)$$

$$\Delta y_t = 0.3\Delta y_{t-1} + a_t, \quad \sigma_a^2 = 0.05. \quad (19)$$

The AR(1) model was chosen because the autoregressive root is 0.95 and, therefore, the model is close to the nonstationarity region. The ARMA(1, 1) model was chosen to analyze the finite sample properties of the proposed procedure in models with

moving average components. Finally, the ARI(1,1) model was chosen to illustrate the performance of the PRR bootstrap procedure in the context of integrated series. The variances of the noises have been chosen in each case to generate original series with properties similar to the ones obtained in real-time series.

For each model considered, we generate artificial series with three alternative error distributions, in particular, Gaussian, Student- t with five degrees of freedom and (minus) exponential. In all cases, we have centered the errors to have 0 mean. The Gaussian distribution has been chosen as a benchmark for comparative purposes given that the STD1, STD2 and STD3 intervals have been derived under this assumption. On the other hand, the Student- t distribution allows us to analyze the effects on the prediction intervals for the original series of having leptokurtic although symmetric innovations after transformation. Finally, the exponential distribution illustrates the effects of having nonsymmetric innovations. Although the exponential distribution may seem rather unrealistic for real data, it may be of interest when analyzing nonlinear time series; see, for example, Granger and Sin (2000) who assume that absolute financial returns, computed as increments of logarithmic prices, are exponential. Furthermore, Harvey and Newbold (2003) present strong evidence of skewness for the errors of some macroeconomic series even after transformation.

Although the Box–Jenkins prediction intervals have been derived assuming Gaussian errors, they are extensively implemented in practice even if the errors are non-Gaussian. Therefore, it seems interesting to compare the bootstrap with STD1, STD2 and STD3 intervals when the errors are non-Gaussian even if the latter is not optimal in these circumstances.

We only report the results obtained for the logarithmic transformation, i.e., $y_t = \log(X_t)$. The conclusions with other power transformations and models considered are the same and therefore are not reported to save space.

All the models for the log-transformed series are estimated by conditional QML, with the parameters restricted to satisfy the stationarity and invertibility conditions. We consider two sample sizes: $T=50$ and 100. The prediction horizons under study are $k=1$ and 3, and the corresponding intervals are constructed with a nominal coverage $1-\alpha$ equal to 0.80, 0.95 and

0.99. For each particular series generated by any of the models considered, with a particular sample size and error distribution F_a , we generated $R=1000$ future values of X_{T+k} from that series and obtain $100(1-\alpha)\%$ prediction intervals, denoted by (L_X, U_X) by each of the five procedures considered. PRR and CB prediction intervals are constructed based on $B=999$ bootstrap replicates. The conditional coverage of each procedure is computed by

$$\hat{\alpha}_X = \#\{L_X \leq x_{T+k}^r \leq U_X\} / R,$$

where x_{T+k}^r , $r=1, \dots, R$ are future values of the variable generated previously. For each artificial series, we also estimate the empirical density of the future values x_{T+k}^r . All the estimated densities in this paper have been obtained using a kernel estimator of S-Plus with a rectangular box and a smoothing parameter of 1. The Monte Carlo results are based on 1000 replicates.

There is no obvious way to decide whether a prediction interval is good. Chatfield (1993) and his discussants conclude that the interval should be such that its coverage should be as close as possible to the nominal coverage with the smallest length possible, and the proportion of observations left outside should be the same in each of the tails. Consequently, prediction intervals are compared in terms of average coverage and length and the proportion of observations lying out to the left and to the right of the intervals through all Monte Carlo replicates.

Computations have been carried out in a HP-UX C360 workstation, using Fortran 77 and the corresponding subroutines of Numerical Recipes by Press, Flannery, Teuklosky, and Vetterling (1986).

4.2. Results of the Monte Carlo experiments

The results of the Monte Carlo experiments for the AR(1) model in Eq. (17) with Gaussian innovations appear in Table 1. First of all, as measured by interval coverage, the STD1 intervals appear to be about as accurate as STD2 and PRR at the 95% level. However, the interval coverage is somewhat misleading because STD1 generate rather biased one-sided prediction intervals; see the average of observations lying out to the left and to the right of the intervals. Additionally, notice that the accuracy of the STD1

Table 1
Monte Carlo results for AR(1) model with Gaussian errors

Lead time	Sample size	Method	Average coverage	Coverage below/above	Average length
1	n	Empirical	95%	2.5%/2.5%	2.04
	50	STD1	94.96 (0.03)	0.33/4.71	2.12 (2.81)
		STD2	94.23 (0.03)	2.90/2.86	2.09 (2.77)
		STD3	93.94 (0.03)	4.00/2.06	2.20 (2.93)
		CB	92.43 (0.04)	3.77/3.80	2.02 (2.74)
		PRR	93.45 (0.03)	3.32/3.23	2.06 (2.76)
	100	STD1	95.24 (0.02)	0.22/4.54	2.09 (2.68)
		STD2	94.63 (0.02)	2.67/2.70	2.06 (2.65)
		STD3	94.34 (0.02)	3.76/1.90	2.18 (2.79)
		CB	93.83 (0.03)	3.08/3.09	2.05 (2.71)
		PRR	94.10 (0.02)	2.92/2.98	2.04 (2.62)
3	n	Empirical	95%	2.5%/2.5%	3.44
	50	STD1	94.15 (0.04)	0.04/5.80	3.80 (5.64)
		STD2	93.11 (0.04)	3.47/3.42	3.65 (5.38)
		STD3	92.38 (0.04)	5.54/2.08	4.21 (6.34)
		CB	92.31 (0.04)	3.79/3.90	3.55 (5.31)
		PRR	92.87 (0.04)	3.56/3.57	3.50 (5.18)
	100	STD1	94.69 (0.03)	0.00/5.31	3.68 (4.82)
		STD2	94.05 (0.03)	2.93/3.01	3.54 (4.62)
		STD3	93.34 (0.03)	4.91/1.74	4.07 (5.38)
		CB	93.65 (0.03)	3.07/3.28	3.48 (4.58)
		PRR	93.78 (0.03)	3.03/3.19	3.40 (4.32)

Quantities in parenthesis are Monte Carlo standard deviations.

intervals does not improve with sample size. Therefore, the symmetric STD1 intervals seem to be not adequate to predict future values of transformed variables. This effect was also observed by [Collins \(1991\)](#) in the context of regression models. Furthermore, notice that the average length of the STD1 intervals is systematically larger than the empirical length. As an illustration, [Fig. 1](#) plots the prediction density corresponding to the STD1 intervals together with the empirical density for a particular series of size $T=100$ generated by model (17) with Gaussian errors. It is rather obvious that the symmetric density is not adequate to represent the empirical density of X_{T+k} .

Next, analyzing the behavior of the intervals based on Eqs. (9) and (10) in [Table 1](#), it is interesting to note that the use of the bias-corrected STD3 intervals do not improve in any case the results of the STD2 intervals. They have larger average length than STD2, and the average observations left out on the right and on the left are clearly asymmetric. This means that using the debiasing factor (designed to obtain a better estimation of the conditional expectation in the

original scale) for correcting the bias of the prediction intervals does not seem to work for the cases considered in this paper.

Comparing PRR and STD2 intervals, it is possible to observe that they have similar performance in terms of both average coverage and length. The reason for the good behavior of STD2 seems clear. In this case, when taking logarithms of the original observations, the resulting transformed series has normal errors, and therefore, the usual [Box and Jenkins \(1976\)](#) and the PRR intervals have similar performance; see [Pascual et al. \(2004\)](#). Consequently, when going back to the original metric this similar behavior remains. In [Fig. 1](#), we also plot the density of X_{T+k} corresponding to retransforming Y_{T+k} , as it is done when constructing the STD2 intervals. Notice that although this density is closer to the empirical density than the one based on STD1 intervals, the shape is still slightly different.

Finally, we concentrate on the comparison of PRR with respect to CB intervals which does not incorporate the parameter uncertainty variability. The results reported in [Table 1](#) show that CB intervals have lower

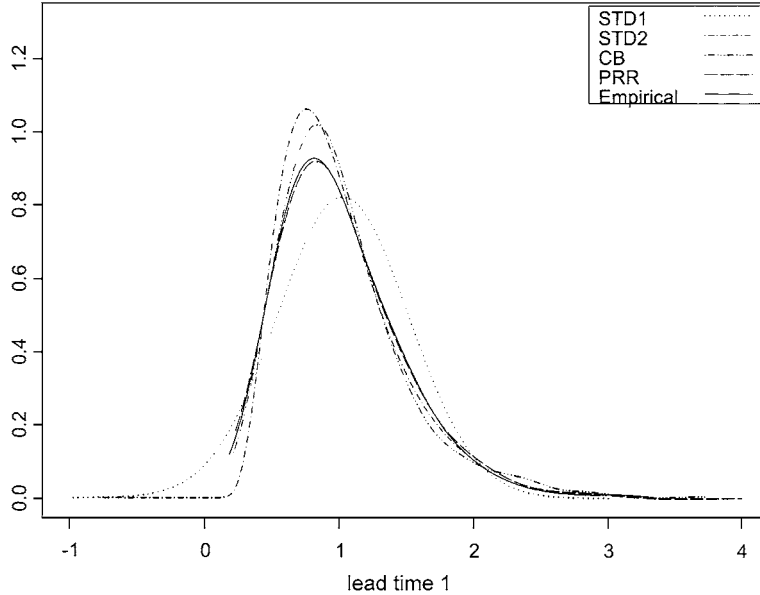


Fig. 1. Densities of one-step ahead predictions of one series of size 100 generated by AR(1)model with Gaussian innovations.

average coverage than PRR, the latter having average coverage closer to the nominal value. Therefore, it seems to be important to include the uncertainty due to parameter estimation in prediction intervals in order to obtain coverages close to the nominal values. The necessity of using PRR is more evident for the smallest sample size. As expected, inasmuch as the conditional QML estimator is consistent, CB and PRR intervals get closer in terms of coverage and length as the sample size increases. The conclusions are essentially the same for predictions made one and three steps ahead. In Fig. 1, it is rather clear that the PRR prediction density is closer to the empirical density than the CB and STD2 densities which are rather close. Consequently, it seems that the improvement of PRR over STD2 intervals is not due to the distribution of the forecast errors but to the inclusion of the variability due to parameter estimation.

When predictions are made three steps ahead into the future, the average coverages of all procedures are relatively larger with respect to the empirical average than when predictions are made one step ahead. However, it is important to observe that the relative increase of the average length of the nonbootstrap intervals is much larger than the relative increase of the bootstrap intervals. For example, if the sample size is 50, the average lengths of the STD2 and PRR one-

step ahead intervals are 2.5% and 1% larger than the corresponding empirical length. However, when the prediction horizon is 3, the STD2 interval is 6% larger, while the PRR interval is only 1.7% larger.

The results for other alternative coverages are similar or slightly more favorable towards bootstrap intervals than the ones reported in Table 1 for 95% intervals.

To analyze whether having a leptokurtic, although symmetric distribution of the errors in the transformed variable, affects the results reported above, Table 2 shows the Monte Carlo results for the 80% prediction intervals for log-transformed series generated by the AR(1) model with innovations generated by the Student- t distribution. We report results for the 80% intervals because the differences are clearer than when 95% intervals are considered. The conclusions with respect to the comparisons between STD1, STD3 and CB intervals are the same as before. In this table, it is also possible to observe the improvement of PRR with respect to STD2 intervals. In this case, the average coverage and lengths of the STD2 intervals are larger than nominal values, and what is even more important is that this bad behavior does not improve as the sample size increases. Remember that the STD2 intervals are built assuming that the transformed variable Y_t is normal. Therefore, as soon as this

Table 2
Monte Carlo results for AR(1) model with Student-5 errors

Lead time	Sample size	Method	Average coverage	Coverage below/above	Average length
1	n	Empirical	80%	10%/10%	1.12
	50	STD1	84.82 (0.06)	6.01/9.17	1.55 (2.75)
		STD2	82.40 (0.06)	8.76/8.83	1.47 (2.58)
		STD3	81.89 (0.06)	11.15/6.97	1.55 (2.76)
		CB	78.06 (0.06)	10.91/11.03	1.30 (2.29)
		PRR	79.15 (0.06)	10.35/10.50	1.31 (2.28)
	100	STD1	85.55 (0.04)	5.64/8.81	1.34 (1.89)
		STD2	83.14 (0.04)	8.38/8.48	1.28 (1.78)
		STD3	82.63 (0.04)	10.74/6.63	1.35 (1.90)
		CB	79.21 (0.04)	10.23/10.55	1.14 (1.56)
		PRR	79.54 (0.04)	10.15/10.31	1.14 (1.59)
3	n	Empirical	80%	10%/10%	1.89
	50	STD1	85.52 (0.08)	3.85/10.63	2.93 (7.47)
		STD2	78.51 (0.08)	10.68/10.81	2.50 (6.22)
		STD3	77.05 (0.08)	15.47/7.48	2.93 (7.49)
		CB	76.44 (0.08)	11.66/11.89	2.38 (6.11)
		PRR	77.54 (0.08)	11.14/11.32	2.31 (6.01)
	100	STD1	87.24 (0.05)	3.01/9.75	2.36 (3.43)
		STD2	80.18 (0.05)	9.86/9.96	2.05 (2.88)
		STD3	78.75 (0.05)	14.59/6.65	2.37 (3.45)
		CB	78.36 (0.05)	10.72/10.91	1.95 (2.77)
		PRR	78.38 (0.05)	10.75/10.87	1.90 (2.62)

Quantities in parenthesis are Monte Carlo standard deviations.

assumption is not satisfied, the intervals do not have the usual properties.

The results are even worse when the innovations have an asymmetric distribution. Table 3 reports the results for the 95% prediction intervals for log-transformed series generated by the ARMA(1,1) model with exponential innovations. The nonbootstrap methods have, in general, average coverage and length over nominal values, and as the sample size increases, this bad behavior tends to be even worse. They are not able to cope with the asymmetry of the transformed series Y_t . Notice that the PRR intervals have an adequate performance, and additionally, as the sample size gets larger, its average coverage and length measures get closer to nominal values, supporting the asymptotic properties mentioned in Section 3. Also, notice that, in this case, the necessity of introducing the variability due to parameter estimation by the use of PRR is crucial even for large sample sizes. For the asymmetric distribution considered, CB intervals have lower average coverage than PRR. Fig. 2 plots kernel estimates of the densities obtained for a particular replicate in this case. It is possible to

observe that the nonbootstrap densities do not resemble the shape of the empirical density while the PRR density is able to mimic it.

Finally, Table 4 reports the Monte Carlo results for the ARI(1, 1) model in Eq. (19) with Student- t errors when the intervals have a nominal coverage of 80%. Notice that, in this case, the average length of the empirical density changes with the sample size. Table 4 illustrates that the bootstrap procedure proposed also works adequately in the presence of unit roots. The results in Table 4 are very similar to the ones reported in Table 2 for the AR(1) model with a large autoregressive parameter. However, it is important to mention that the parameters of the particular ARI(1, 1) model chosen in these simulations are rather small, and therefore, the predictions of future values are dominated by the unit root. Consequently, the CB and PRR are rather similar. The uncertainty due to parameter estimation should be more important when predicting with models with larger parameters. Furthermore, in Table 4, the asymmetry of the STD1 intervals is not as pronounced as it is in Table 2. However, this asymmetry increases the larger the prediction horizon is.

Table 3
 Monte Carlo results for ARMA(1, 1) model with exponential errors

Lead time	Sample size	Method	Average coverage	Coverage below/above	Average length	
1	n	Empirical	95%	2.5%/2.5%	1.99	
		50	STD1	99.67 (0.03)	0.06/0.27	4.43 (2.35)
		STD2	94.10 (0.03)	5.81/0.09	4.02 (1.88)	
		STD3	91.98 (0.03)	8.00/0.02	5.38 (3.29)	
		CB	90.90 (0.09)	3.79/5.31	2.07 (0.78)	
		PRR	94.27 (0.06)	3.44/2.28	2.28 (0.74)	
		100	STD1	99.99 (0.01)	0.01/0.00	4.41 (1.84)
		STD2	94.44 (0.02)	5.56/0.00	4.05 (1.58)	
		STD3	92.25 (0.02)	7.75/0.00	5.29 (2.41)	
		CB	93.18 (0.06)	3.10/3.72	2.07 (0.70)	
		PRR	94.91 (0.05)	3.02/2.07	2.15 (0.68)	
	3	n	Empirical	95%	2.5%/2.5%	2.58
			50	STD1	99.28 (0.03)	0.01/0.71
			STD2	94.29 (0.04)	5.53/0.19	4.80 (2.15)
		STD3	91.78 (0.03)	8.17/0.05	7.12 (5.49)	
		CB	91.33 (0.07)	3.56/5.11	2.60 (0.77)	
		PRR	93.48 (0.05)	3.33/3.19	2.77 (0.78)	
		100	STD1	99.77 (0.01)	0.00/0.23	5.38 (2.15)
		STD2	94.83 (0.02)	5.14/0.03	4.75 (1.59)	
		STD3	92.19 (0.02)	7.81/0.00	6.75 (3.20)	
		CB	93.06 (0.05)	3.03/3.91	2.59 (0.63)	
		PRR	93.93 (0.04)	2.97/3.09	2.67 (0.63)	

Quantities in parenthesis are Monte Carlo standard deviations.

Summarizing, PRR intervals perform as well as STD2 intervals when the innovations of the transformed data are well approximated by a normal

distribution and outperform the existing procedures when this distribution differs from the Gaussian one, a situation frequently found when working with real

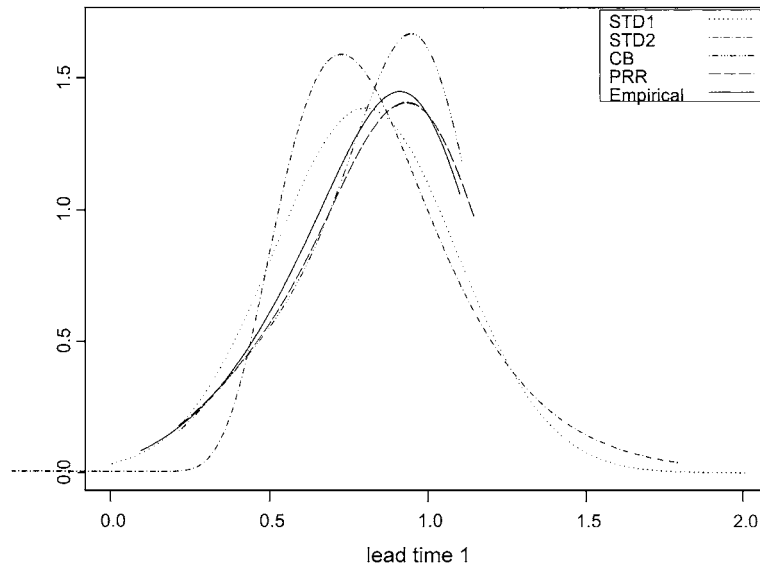


Fig. 2. Densities of one-step ahead predictions of one series of size 100 generated by AR(1) model with exponential innovations.

Table 4
Monte Carlo results for ARI(1, 1) model with Student-5 errors

Lead time	Sample size	Method	Average coverage	Coverage below/above	Average length
1	n	Empirical	80%	10%/10%	0.79
	50	STD1	82.27(.06)	8.48/9.25	0.90(1.96)
		STD2	82.15(.06)	8.85/9.01	0.90(1.95)
		STD3	82.12(.06)	9.33/8.55	0.90(1.96)
		CB	78.03(.07)	10.88/11.09	0.80(1.75)
		PRR	78.80(.06)	10.44/10.76	0.82(1.79)
1	n	Empirical	80%	10%/10%	1.02
	100	STD1	83.06(.04)	8.15/8.79	1.15(3.26)
		STD2	82.94(.04)	8.51/8.55	1.15(3.25)
		STD3	82.91(.04)	9.00/8.09	1.15(3.26)
		CB	79.13(.04)	10.40/10.47	1.05(2.97)
		PRR	79.41(.04)	10.24/10.35	1.06(3.01)
3	n	Empirical	80%	10%/10%	1.81
	50	STD1	80.43(.08)	8.99/10.58	2.01(4.63)
		STD2	79.81(.08)	10.04/10.14	1.98(4.55)
		STD3	79.67(.08)	11.23/9.10	2.01(4.64)
		CB	78.19(.08)	10.80/11.01	1.89(4.29)
		PRR	78.39(.07)	10.70/10.90	1.89(4.34)
3	n	Empirical	80%	10%/10%	2.35
	100	STD1	81.50(.06)	8.51/9.99	2.56(7.49)
		STD2	80.84(.05)	9.60/9.57	2.53(7.39)
		STD3	80.71(.05)	10.79/8.49	2.57(7.50)
		CB	79.12(.05)	10.38/10.51	2.42(7.10)
		PRR	79.18(.05)	10.33/10.49	2.43(7.11)

Quantities in parenthesis are Monte Carlo standard deviations.

data; see, for example, [Harvey and Newbold \(2003\)](#). Furthermore, the symmetric intervals based on Eq. (8) are shown to have poor properties even when the transformed data are Gaussian. The bias-correcting factor for the end of the prediction intervals in Eq. (9) proposed by [Guerrero \(1993\)](#) is also shown not to improve the properties of the noncorrected intervals. We have also shown that including the uncertainty due to estimation of the parameters of the model in the bootstrap prediction intervals may be crucial depending on the distribution of the transformed data.

Finally, it is important to point out that, for the models and sample sizes considered in this paper and given the computer facilities available nowadays, the differences between the computer times needed for the bootstrap and nonbootstrap procedures to get prediction intervals are irrelevant. For a particular series, it takes just a few seconds to obtain the

corresponding prediction intervals. Therefore, it seems that, as soon as the Gaussianity assumption is not adequate, it is worth using the proposed bootstrap intervals.

5. Real data applications

In this section, we illustrate empirically the use of the suggested bootstrap method to construct prediction intervals for transformed variables. We start considering the Sales Data, studied first by [Chatfield and Prothero \(1973\)](#) and latter by [Pankratz and Dudley \(1987\)](#) and [Guerrero \(1993\)](#) among others. The series, plotted in the top panel of [Fig. 3](#), consists of 77 observations of the monthly sales of an engineering product with a marked trend and a strong seasonal pattern. Inasmuch as the size of the seasonal effect increases with the mean level of sales, [Chatfield](#)

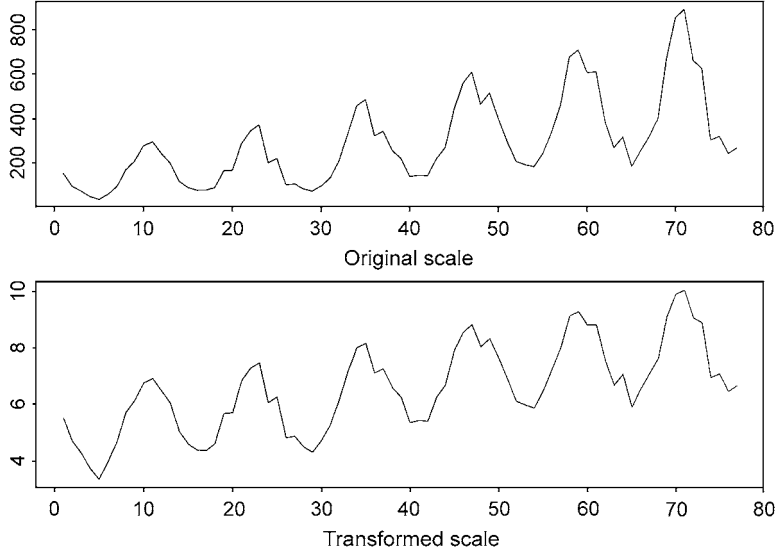


Fig. 3. Observations of monthly sales of an engineering product both in the original (top panel) and in the transformed scale (bottom panel).

and Prothero originally used the log transformation. However, this transformation was criticized by Wilson (1973) who found by maximum likelihood that a more convenient power transformation was $\hat{\lambda}=0.34$; see also Guerrero (1993). Therefore, we will consider $\lambda=1/3$ as known. The bottom panel of Fig. 3 represents the transformed observations. The model finally fitted to the transformed data is

$$(1 - \phi L)\Delta\Delta_{12}y_t = (1 - \Theta L^{12})a_t \quad (20)$$

where $y_t=X_t^{1/3}$ and X_t denotes the original series. The first 65 observations of the series corresponding to the period from January 1965 up to May 1970 are used to estimate the parameters of model (20). The last 12 observations corresponding to the period from June 1970 up to May 1971 are used to assess the predictive performance of the STD2, STD3 and PRR prediction intervals. Note that, for this particular transformation, STD1 prediction intervals cannot be computed inasmuch as no formulas for the variance of the prediction error of X_t are available.

The QML estimates are $\hat{\phi}=-0.5437$ and $\hat{\Theta}=0.5466$. The skewness coefficient of the corresponding residuals is -0.2946 , and the excess kurtosis is 0.065 , with the former significantly different from 0. Therefore, the empirical distribution of the residuals has a long tail to the left.

Then, we implement the bootstrap procedure described in Section 3 to construct prediction densities of the transformed variable y_{65+k} for $k=1, \dots, 12$. The estimated density for lead time 1 together with the normal density appears in Fig. 4 where it can be observed that the bootstrap density is asymmetric to the left, as observed in the residuals distribution. Finally, we implement the PRR procedure to construct bootstrap prediction densities and intervals for future values of the variable in the original scale. In Table 5, which reports the interval lengths for some selected horizons, it can be observed that the bootstrap intervals are always thinner than the STD2 intervals, with both intervals having the same coverages.

In this case, using the mean or the median of the bootstrap density does not improve the MSE of the predictions over the retransformed point predictions.

Next, we analyze two economic time series, the U.S. Dollar–Pound real exchange rate (RXR) and the ratio of nonborrowed total reserves (NBRX). These series are studied by Kilian (1998) in the context of VAR models, who shows that the residuals clearly reject the normality assumption. Both series consist of 197 observations, where the first 173 are used to estimate the parameters of the ARIMA model fitted to the log-transformed data, and the last 24 observations are used to assess the predictive performance of the methods considered in this section.

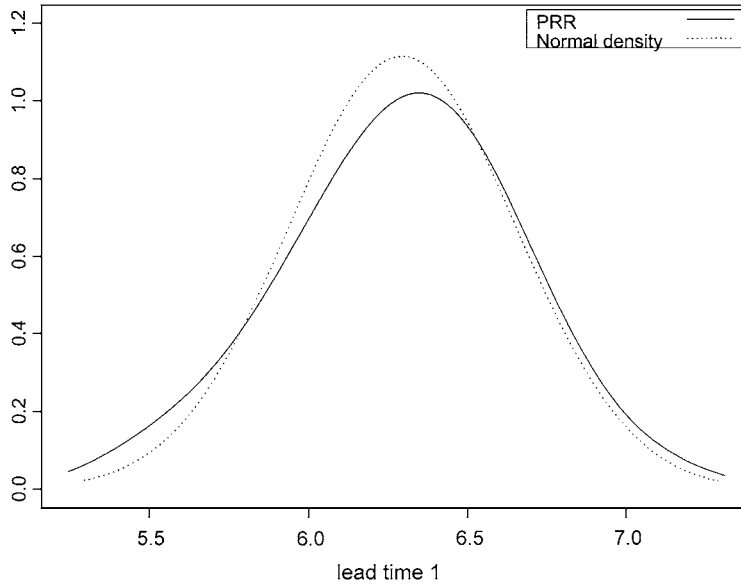


Fig. 4. Estimated density of one-step ahead predictions of the transformed monthly sales constructed by PRR method and normal density.

The final model fitted to the log-RXR series y_t is

$$\Delta y_t = (1 + 0.401L)\hat{a}_t \quad (21)$$

The skewness coefficient of the residuals \hat{a}_t is 0.2112, and the excess kurtosis is 0.37, with the former being different from 0. Therefore, as concluded by Kilian (1998), the residuals have an asymmetric distribution. The PRR and STD2 prediction intervals at 80%, plotted in Fig. 5, are asymmetric, although the upper and lower bounds of the bootstrap intervals are under the corresponding bounds of the STD2 intervals. As a consequence, the bootstrap intervals are able to include one observation left out by the STD2 intervals.

Finally, the model fitted to the log-NBRX data y_t is

$$\hat{y}_t = -0.031 + 0.8481\hat{y}_{t-1} \quad (22)$$

In this case, the skewness coefficient of the residuals is -0.9071 , and the excess kurtosis is 6.50, both significantly different from 0. Once more, the assumption of normality is clearly rejected. Fig. 6 shows how the bootstrap prediction intervals capture the asymmetry and kurtosis inherent in the residuals and consequently in the prediction densities. It is important to note that the length of the bootstrap intervals are shorter than the STD2 ones having the same coverages.

6. Summary and conclusions

This paper extends the bootstrap technique proposed by Pascual et al. (2004) to construct prediction intervals for a variable after a linear ARIMA model is fitted to its power transformation. In this situation, there is no generally accepted method of computing prediction intervals. The proposed resampling scheme does not assume any distribution for the errors neither in the original nor in the transformed metrics and, at the same time, allows to incorporate the variability due to parameter estimation. By means of Monte

Table 5
Naive (STD2) and bootstrap (PRR) interval lengths for Sales data

Nominal	Method	Forecast horizon					
		1	2	4	6	8	12
80%	STD2	108.28	132.86	268.63	400.23	399.38	258.86
	PRR	105.30	132.51	257.79	388.62	376.98	258.08
95%	STD2	166.18	203.94	412.48	614.75	614.31	401.13
	PRR	174.73	203.48	394.92	604.09	608.20	389.36
99%	STD2	221.09	271.38	549.16	818.80	819.73	540.45
	PRR	217.56	260.74	523.04	753.21	801.73	558.21

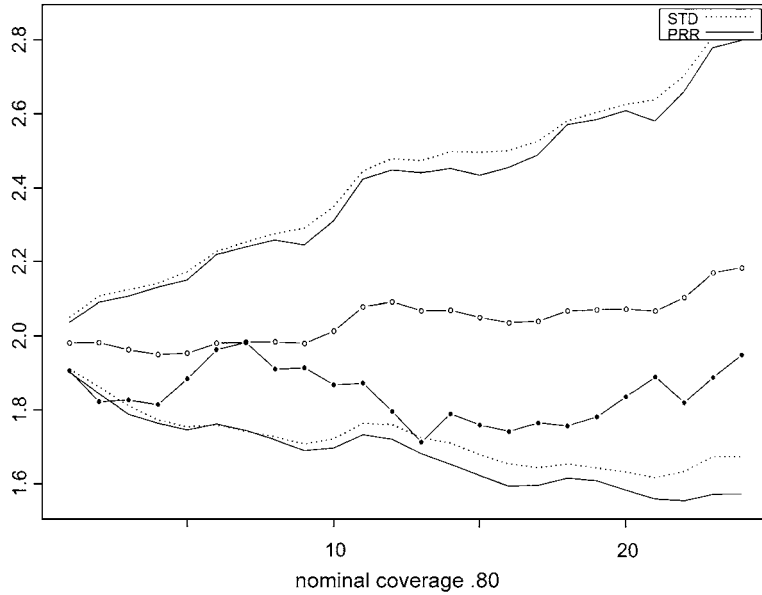


Fig. 5. Observed levels of RRX series (●) and point predictions (○). 80% intervals constructed by STD2 and PRR procedures.

Carlo experiments, we compare the finite sample performance of alternative methods previously proposed in the literature to construct prediction intervals for power-transformed series with the bootstrap approach proposed in this paper. There are two main

alternatives. The first one based on [Granger and Newbold \(1976\)](#) assumes a symmetric distribution for both the original and the transformed variable and can only be implemented for logarithmic and root squared transformations. These prediction intervals ignore the

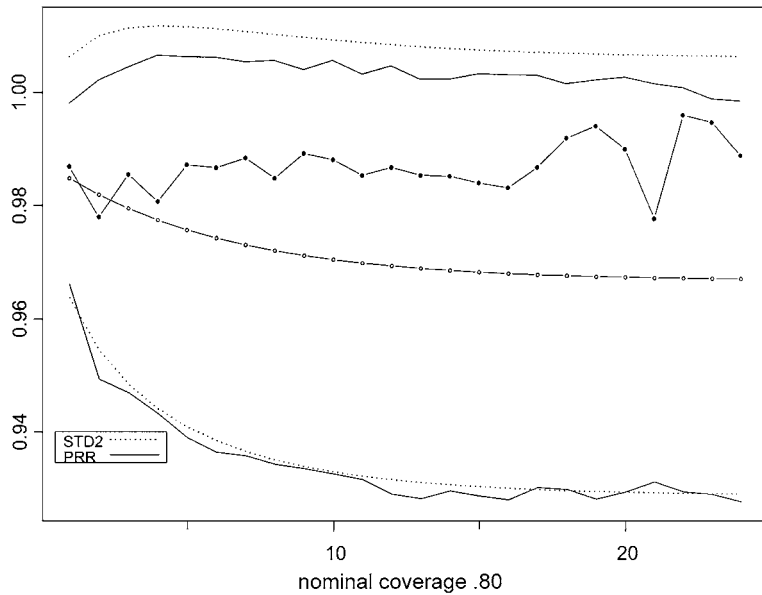


Fig. 6. Observed levels of NBRX series (●) and point predictions (○). 80% intervals constructed by STD2 and PRR procedures.

skewness and all higher moments of the variable of interest. As a result, this approach will generate biased one-sided prediction intervals.

The second alternative is based on retransforming the ends of the prediction intervals for the transformed variable. In this case, only the Gaussianity of the transformed variable is needed. None of these intervals are able to take into account the uncertainty due to estimation of the parameters. The intervals constructed simply by retransforming the upper and lower values of the usual prediction intervals for the transformed series have only good properties when the transformed series has normal errors. In this case, the usual Box–Jenkins intervals for the variable in the transformed metric have very good properties, and therefore, when going back to the original scale, this good behavior remains. The results show that, for nonnormal innovations, these prediction intervals can be heavily distorted. The bias correction proposed by Guerrero (1993) does not improve the results for the cases considered in this paper. The bootstrap intervals seem to have appropriate properties.

We also analyze how the coverage and length of prediction intervals are affected by not taking into account the variability due to parameter estimation. We show that the average coverage of the intervals is closer to the nominal value when intervals are constructed incorporating parameter uncertainty. As expected, inasmuch as we are considering consistent estimators, the effects of parameter estimation are particularly important for small sample sizes. Furthermore, these effects are more important when the error distribution is not Gaussian; see also Pascual et al. (2001).

The bootstrap approach presented in this paper seems to have reasonable properties when prediction intervals are required for a variable after a power transformation is taken to its original values. This approach gives prediction intervals with a reasonable finite sample performance in terms of average coverage and average length with both normal and non-normal distributions of the innovations. As expected, its behavior improves as the sample size increases. Additionally, this method not only gives prediction intervals but also provides estimates of the prediction density function of the variable in its original scale; see, for example, Tay and Wallis (2000) for a survey on the importance of prediction density.

The behavior of the PRR technique is illustrated with the analysis of three real-time series. It is shown that the PRR intervals are shorter than the retransformed intervals having at the same time better coverage properties.

One interesting generalization of the bootstrap procedure proposed in this paper is to consider a procedure to construct prediction intervals that also incorporate the uncertainty associated with the transformation parameter λ which could be estimated by QML methods as proposed, among others, by Robinson (1991) or by semiparametric methods as proposed by Foster, Tian, and Wei (2001). However, in this case, it is not obvious which is the best inverse transformation to be used to obtain prediction intervals in the original scale. This problem is left for further research.

Acknowledgments

The authors are very grateful to Lutz Kilian for providing the RXR and NBRX series and to the associate editor Antonio García Ferrer and two anonymous referees for valuable comments that have been very helpful to improve the paper. Financial support was provided by projects DGES PB96-0111 and BEC 2002-03720 from the Spanish Government and Catedra de Calidad from BBV.

References

- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, B-26*, 211–243.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Cao, R., Febrero-Bande, M., González-Manteiga, W., Prada-Sánchez, J. M., & García-Jurado, I. (1997). Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Communications in Statistics. Simulation and Computation, 26*, 961–978.
- Chatfield, C. (1993). Calculating interval forecasts (with discussion). *Journal of Business and Economic Statistics, 11*, 121–135.
- Chatfield, C., & Prothero, D. L. (1973). Box Jenkins seasonal forecasting: Problems in a case study. *Journal of the Royal Statistical Society, A-136*, 295–315.
- Collins, S. (1991). Prediction techniques for Box-Cox regression models. *Journal of Business and Economic Statistics, 9*, 267–277.

- Foster, A. M., Tian, L., & Wei, L. J. (2001). Estimation for the Box-Cox transformation model without assuming parametric error distribution. *Journal of the American Statistical Association*, *96*, 1097-1101.
- Granger, C. W. J., & Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society, B*:38, 189-203.
- Granger, C. W. J., & Sin, C. -Y. (2000). Modelling the absolute returns of different stock indices: Exploring the forecastability of an alternative measure of risk. *Journal of Forecasting*, *19*, 277-298.
- Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, *12*, 37-48.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge, UK: Cambridge University Press.
- Harvey, A. C. (1993). *Time series models*. Harvester Wheatsheaf.
- Harvey, D. I., & Newbold, P. (2003). The non-normality of some macroeconomic forecast errors. *International Journal of Forecasting*, *19*, 635-653.
- Kilian, L. (1998). Confidence intervals for impulse responses under departures from normality. *Econometric Reviews*, *17*, 1-29.
- Pankratz, A., & Dudley, U. (1987). Forecasts of power-transformed series. *Journal of Forecasting*, *6*, 239-248.
- Pascual, L., Romo, J., & Ruiz, E. (2001). Effects of parameter estimation on prediction densities: A bootstrap approach. *International Journal of Forecasting*, *17*, 83-103.
- Pascual, L., Romo, J., & Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis*, *25*, 449-485.
- Press, W. H., Flannery, B. P., Teuklosky, S. A., & Vetterling, W. T. (1986). *Numerical recipes*. Cambridge, UK: Cambridge University Press.
- Robinson, P. M. (1991). Best nonlinear three-stage least squares estimator of certain econometric models. *Econometrica*, *59*, 755-786.
- Tay, A. S., & Wallis, K. F. (2000). Density forecasting: A survey. *Journal of Forecasting*, *19*, 235-254.
- Taylor, J. M. G. (1986). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association*, *81*, 114-118.
- Tukey, J. W. (1957). On the comparative anatomy of transformations. *Annals of Mathematical Statistics*, *28*, 602-632.
- Wilson, G. T. (1973). Discussion of Chatfield and Prothero (1973). *Journal of the Royal Statistical Society, A-136*, 315-319.

Biographies: Lorenzo PASCUAL received his PhD in Mathematical Engineering from Universidad Carlos III de Madrid in 2001. He is Chief Analyst at the Energy Management Department of HidroCantábrico.

Juan ROMO received his PhD in Mathematics from Texas A&M University. He has been an Associate Professor both at Universidad Complutense and Universidad Carlos III de Madrid, where he is Professor of Statistics and Operations Research.

Esther RUIZ received her PhD in Statistics from the London School of Economics in 1992. She is an Associate Professor of Economics at Universidad Carlos III de Madrid.