Bootstrap Prediction Intervals for Autoregressive Time Series

Michael P. Clements
Department of Economics
University of Warwick, UK

Jae H. Kim*
Department of Econometrics and Business Statistics
Monash University
Australia

Keywords: Autoregressive models, Bootstrapping, Bias-Correction.

First version December 2005,
This version September 2006.

Abstract

This paper is concerned with the calculation of interval forecasts for highly-persistent autoregressive (AR) time series using the bootstrap. Three methods are considered for countering the small-sample bias of least squares estimation for processes which have roots close to the unit circle: a bootstrap bias-corrected OLS estimator; the use of the Roy-Fuller estimator in place of OLS; and the use of the Andrews-Chen estimator in place of OLS. All three methods of bias correction yield superior results to the bootstrap in the absence of bias correction. Of the three correction methods, the bootstrap prediction intervals based on the Roy-Fuller estimator are generally superior to the other two. The small sample performance of bootstrap prediction intervals based on the Roy-Fuller estimator are investigated when the order of the AR model is unknown, and has to be determined using an information criterion.

* Corresponding author: Department of Econometrics and Business Statistics Monash University, Caulfield East, Victoria 3145, Australia. Tel. (61 3) 99031596, Fax. (61 3) 99032007, Email: Jae.Kim@Buseco.monash.edu.au

We are grateful to Lutz Kilian, two anonymous referees, and editor Patrick Groenen for helpful comments. The Gauss computer code used to perform the calculations is available from the corresponding author on request.

1. Introduction

In this paper we build on the successful application of the bootstrap to the calculation of prediction intervals for stationary autoregressions by Thombs and Schucany (1990) by focusing on processes which are highly persistent. Many macroeconomic time-series data appear to be either near-unit root processes or to possess a unit root. See, for example, Nelson and Plosser (1982), and the literature on whether US GNP contains a unit root or can be better characterised as a trend stationary processes (see, e.g., Christiano and Eichenbaum, 1990). Although it may often be difficult to distinguish between the unit root hypothesis, and trend-stationarity, a number of studies have shown that difference-stationary (DS) and trend-stationary (TS) models may imply very different point forecasts (e.g., Diebold and Senhadji, 1996, Clements and Hendry, 2001).

We supplement the literature on the point-forecast properties of near-unit root models with an analysis of the calculation of prediction intervals for models with linear trends and roots close to one using the bootstrap. Thombs and Schucany (1990) and subsequent authors have established the relative advantages of the bootstrap (see Efron, 1979) over classical methods for calculating prediction intervals for stationary processes. Authors using bootstrap methods include Masarotto (1990), Kabaila (1993), Breidt, Davis and Dunsmuir (1995), Grigoletto (1998), Kim (1999, 2001, 2002), Pascual, Romo and Ruiz (2001, 2006), Clements and Taylor (2001), and Alonso and Pena (2002). The `classical methods' are those of Box and Jenkins (1970). These generally require that the error distribution is normal, or at least that it is known, and fail to take into account the sampling variability of the estimated coefficients: Chatfield (1993) provides a critique. In this paper we examine whether bootstrap methods also work well for near-unit root models.

The bootstrap prediction intervals involve repeated generation of pseudo-data sets using the estimated model and re-sampling the residuals. The large-sample validity of the bootstrap prediction intervals obtained in this way is established for stationary processes by

Thombs and Schucany (1990, Theorem 3.1) when the pseudo data sets are generated using the backward representation of the AR model. Pascual, Romo and Ruiz (2001) and Clements and Taylor (2001) show that the asymptotic validity extends to re-sampling based on the forward-looking AR model for stationary processes. However, the bootstrap prediction intervals will be adversely affected by the small-sample bias of OLS when sample sizes are small and the process is highly persistent. Other estimators, such as maximum likelihood and Yule-Walker estimators, are also biased in small samples. We restrict our attention to the LS estimator in this paper, as the LS estimator is the most widely used in practice.

Without some form of bias correction, the small-sample bias of LS will impart a "double bias" to the bootstrapped prediction intervals: the estimated model used to generate bootstrap replicates of the data will be affected by small-sample bias, as will the parameter estimates of the model on the bootstrap replicates. Kim (2001) and Clements and Taylor (2001) use a "pre-bootstrap" to bias correct the parameter estimator, following the lead of Kilian (1998a), giving rise to the bootstrap-after-bootstrap. We use PI_B to denote prediction intervals calculated using the bootstrap-after-bootstrap, and PI to refer to prediction intervals calculated using a standard bootstrap (i.e., without bias correction). PI_B has been found to yield substantial improvements in the small-sample properties of bootstrap prediction intervals, especially for near-unit root processes.

There are alternatives to the bootstrap as a means of bias-correction, which we consider in this paper. The bootstrap is computationally expensive, so that the biases associated with the AR parameter estimators obtained from the pseudo-data sets are typically approximated by those obtained from the original time series. In addition, the bootstrap bias-correction can push stationary estimates into the non-stationary region of the parameter space. Kilian (1998a) suggests an adjustment to prevent this from happening, but the adjustment lacks a theoretical basis. For the two alternatives to PI_B that we consider, the prediction intervals will still be calculated by a bootstrap, but instead of employing the bootstrap to bias

correct we use the estimators of Roy and Fuller (2001) and Andrews and Chen (1994). We denote the resulting intervals by PI_{RF}, and PI_{AC}, respectively. These estimators are designed to be approximately median-unbiased for the AR model with a possible unit root. Thus, whilst bootstrap bias-correction seeks to make the estimator mean unbiased (which is traditionally what is meant when an estimator is termed unbiased), these estimators seek to set the median of the corrected estimator to the true value. The Roy-Fuller estimator has an edge over the other two methods (Andrews-Chen and PI_B) in terms of computational efficiency, but we will explore the relative merits of the three in terms of the small-sample properties of the prediction intervals.

Recent work by Hansen (1999) on a grid bootstrap method has been used by Gospodinov (2002) to obtain median unbiased forecasts for highly persistent processes. Although computationally intensive, the grid bootstrap may work well in the present context. We leave investigation of this to future research.

We carry out an extensive Monte Carlo study of the small-sample properties of the alternative bias-corrected prediction intervals for autoregressive models. Our experimental design will shed light on the behaviour of the methods for different sample sizes, forecast horizons, and degrees of persistence, including near-unit root and unit root processes. We consider first-order and higher-order autoregressions (the leading example of which is obviously an AR(2)). Whilst the validity of the bootstrap for unit root processes has not been formally established in the literature, we include unit root processes in our simulations to see how well the methods perform as the root gets arbitrarily close to unity. We note that although Inoue and Kilian (2002) establish the validity of the bootstrap for higher-order unit-root processes for impulse response analysis, the extension to (conditional) forecasting is non-trivial.

The remainder of the paper is as follows. Section 2 explains the three bias-correction methods. Section 3 describes the application of these methods to bootstrapping prediction

intervals for autoregressions. Section 4 outlines the experimental design, and section 5 presents the results of the Monte Carlo. Section 6 describes an empirical application of the methods studied in this paper, and section 7 concludes.

2. The AR Model and Bias-Corrected Parameter Estimators

We consider autoregressive models (see, e.g., Box and Jenkins, 1970, for a full treatment). The AR(p) model takes the following form:

$$Y_t = \mu' + \beta' t + v_t,\ v_t = \alpha_1 v_{t-1} + \ldots + \alpha_p v_{t-p} + u_t, \qquad (1)$$

where u_t is independently and identically distributed (iid), i.e., $u_t \sim$ iid $(0,\sigma^2)$. We will focus on the case of normal (gaussian) innovations, but will briefly consider the robustness of the results to other assumptions about the error distribution. We assume the order of the process, p, is known to begin with, but then consider the impact of uncertainty over the model lag order. In (1), μ' and β', and the α_i, are fixed parameters (as are μ and β in the following) whereas Y_t and v_t are random variables. The model can be re-written in the following way:

$$Y_t = \mu + \beta t + \alpha_1 Y_{t-1} + \ldots + \alpha_p Y_{t-p} + u_t, \qquad (2)$$

where $\mu = \mu'(1-\alpha) + (\alpha_1 + 2\alpha_2 + \ldots + p\alpha_p)\beta'$, $\beta = \beta'(1-\alpha)$ and $\alpha = \alpha_1 + \ldots + \alpha_p$. The parameter α measures the degree of persistence of the AR model, and is equal to one if the model has a unit root. When $\alpha = 1$, the trend coefficient β becomes zero and the model becomes a random walk with drift. Equation (2) can also be written as

$$Y_t = \mu + \beta t + \alpha Y_{t-1} + \psi_1 \Delta Y_{t-1} + \ldots + \psi_{p-1} \Delta Y_{t-p+1} + u_t, \qquad (3)$$

to bring out the role of α. Here, the first difference operator is $\Delta = 1 - L$, where L is the lag operator, such that $L^s y_t = y_{t-s}$. Note that the parameters of equations (2) and (3) are related by:

$$\alpha_1 = \alpha + \psi_1,\ \alpha_j = -\psi_{j-1} + \psi_j \text{ for } 2 \le j \le p\text{-}1,\ \alpha_j = -\psi_{j-1} \text{ for } j = p. \qquad (4)$$

Given the observed time series $\{Y_t\}_{t=1}^n$, the least-squares (LS) estimator for $\gamma = (\mu, \beta, \alpha_1, \ldots, \alpha_p)$ in equation (2) can be obtained by regressing Y_t on $(1, t, Y_{t-1}, \ldots, Y_{t-p})$. The LS

estimators and the associated residuals are denoted as $\hat{\gamma} = (\hat{\mu}, \hat{\beta}, \hat{\alpha}_1, ..., \hat{\alpha}_p)$ and $\{\hat{u}_t\}_{t=p+1}^{n}$ respectively. The LS estimators for the parameters in equation (3) can be obtained in a similar way, or by noting that (4) holds for the estimators as well as the population values. Note that $\hat{\alpha} = \hat{\alpha}_1 + ... + \hat{\alpha}_p$.

2.1 The Bootstrap Bias-Corrected Estimator

The bootstrap (see, e.g., Efron, 1979, and Efron and Tibshirani, 1993) is a computer-intensive method of approximating the unknown sampling distribution of a statistic, as it involves repeated re-sampling of the observed data. It has been applied widely in econometrics and time series analysis, and has been found to be a useful alternative to the conventional methods of estimation, forecasting, and statistical inference (see Li and Maddala, 1996; Berkowitz and Kilian, 2000; MacKinnon, 2002). In the context of the AR model, the bootstrap procedure involves the generation of a large number of pseudo-data sets using the estimated coefficients and re-sampled residuals. In this way, the pseudo-data sets inherit the dependence structure present in the observed time series.

The biases of AR parameter estimators can be estimated as follows using the bootstrap. Generate a pseudo-data set $\{Y_t^*\}_{t=1}^{n}$ as:

$$Y_t^* = \hat{\mu} + \hat{\beta}t + \hat{\alpha}_1 Y_{t-1}^* + \ldots \hat{\alpha}_p Y_{t-p}^* + u_t^*, \qquad (5)$$

using $\{Y_t\}_{t=1}^{p}$ as starting values, where u_t^* is a random draw with replacement from $\{\hat{u}_t\}_{t=p+1}^{n}$, where the LS residuals have been appropriately scaled (see e.g., Thombs and Schucany, 1990). B sets of pseudo-data are generated, each of which gives a bootstrap parameter estimate for γ, so that we obtain $\{\gamma^*(j)\}_{j=1}^{B}$. A typical $\gamma^* = (\mu^*, \beta^*, \alpha_1^*, \ldots, \alpha_p^*)$ is obtained by regressing Y_t^* on $(1, t, Y_{t+1}^*, ..., Y_{t+p}^*)$. The bias of $\hat{\gamma}$ is estimated as $Bias(\hat{\gamma}) = \bar{\gamma}^* - \hat{\gamma}$, where $\bar{\gamma}^*$ is the sample mean of $\{\gamma^*(j)\}_{j=1}^{B}$. Using this bootstrap estimator of the bias, the bias-corrected estimator for γ is given by $\hat{\gamma}^B = \hat{\gamma} - Bias(\hat{\gamma}) = 2\hat{\gamma} - \bar{\gamma}^*$. MacKinnon and Smith

(1998) refer to this as the *constant-bias-correcting* (or CBC) estimator, because one can think of it as being founded on the assumption that the bias of the LS estimator is a constant function of the true value of the parameter vector being estimated. This suggests that one might use their *linear-bias-correcting* estimator (LBC) which assumes that the bias is a linear function of the true parameter. MacKinnon and Smith show that linearity of the bias function may a reasonable assumption for much of the stationary region of the AR(1) parameter space. We leave an investigation into the usefulness of the LBC and related estimators for bootstrapping prediction intervals for future research, and instead focus on the CBC as our bootstrap bias-corrected estimator.

In performing the bootstrap bias-correction, we may wish to ensure that the correction does not push stationary estimates into the non-stationary region of the parameter space. Kilian (1998a) suggests the following algorithm. If the modulus of the largest root of the characteristic equation formed from the autoregressive parameters in $\hat{\gamma}$ exceeds one (non-stationary), then $\hat{\gamma}^B = \hat{\gamma}$. That is, the estimates are not adjusted. If the original estimates imply stationarity, and the bias-corrected estimates also imply stationarity, then the correction stands. The algorithm becomes active when the bias correction causes stationary estimates to become non-stationary. Let $\delta_1 = 1$, $\Delta_1 = Bias(\hat{\gamma})$ and $\hat{\gamma}_i^B = \hat{\gamma} - \Delta_i$. Set $\Delta_{i+1} = \delta_i \Delta_i$, $\delta_{i+1} = \delta_i - 0.01$ for $i = 1, 2, 3, \ldots$. Iterate until $\hat{\gamma}_i^B$ satisfies the condition of stationarity and set $\hat{\gamma}^B = \hat{\gamma}_i^B$. For example, in the AR(1) case, if $\hat{\alpha}_1 = 0.95$ and $\hat{\alpha}_1^B = 1.05$, then $\hat{\alpha}_1^B$ is adjusted to 0.99. This yields the bias-corrected estimator for γ based on the bootstrap method $\hat{\gamma}^B = (\hat{\mu}^B, \hat{\beta}^B, \hat{\alpha}_1^B, ..., \hat{\alpha}_p^B)$. When combined with the bootstrap method of calculating prediction intervals, we have the PI_B method.

2.2 Roy-Fuller Estimator

The Roy-Fuller estimator provides a simple modification to the LS estimator for α in equation (3). Let $\hat{\alpha}^{RF} = \min(\ddot{\alpha}, 1)$, where $\ddot{\alpha} = \dot{\alpha} + [C_p(\hat{\tau}_1) + C_{-p}(\hat{\tau}_{-1})]\hat{\sigma}_1$. $\dot{\alpha}$ is the LS estimator for the coefficient of $\hat{y}_{t-1}$ in the regression of $\hat{y}_t$ on $\hat{y}_{t-1}, \Delta\hat{y}_{t-1}, ..., \Delta\hat{y}_{t-p+1}$, where $\hat{y}_t$ is the LS residual from the regression of Y_t on (1, t), and $\hat{\sigma}_1$ is the standard error of $\dot{\alpha}$. Note that $\hat{\tau}_1 = (\dot{\alpha} - 1)\hat{\sigma}_1^{-1}$ is the unit root test statistic, and $\hat{\tau}_{-1}$ is the statistic to test for the hypothesis that $\alpha = -1$. The functions $C_p(\hat{\tau}_1)$ and $C_{-p}(\hat{\tau}_{-1})$ control the way in which bias-correction is conducted, and are constructed so that $\hat{\alpha}^{RF}$ is approximately median-unbiased at $\alpha = 1$ and -1 respectively.

Roy and Fuller (2001, p.488) suggested the following form of $C_p(\hat{\tau}_1)$ function, which is related to the bias expression for $\hat{\alpha}$ they derived:

$$
\begin{aligned}
C_p(\hat{\tau}_1) &= -\tau_{MED} + d_n(\hat{\tau}_1 - \tau_{MED}) && \hat{\tau}_1 > \hat{\tau}_{MED}, \\
&= I_p(n^{-1}\hat{\tau}_1) - 3[\hat{\tau}_1 + k(\hat{\tau}_1 + K)]^{-1} && -K < \hat{\tau}_1 \le \hat{\tau}_{MED}, \\
&= I_p(n^{-1}\hat{\tau}_1) - 3(\hat{\tau}_1)^{-1} && -(3n)^{0.5} < \hat{\tau}_1 \le -K, \\
&= 0 && \hat{\tau}_1 \le -(3n)^{0.5},
\end{aligned}
$$

where I_p is the integer part of $0.5(p+1)$, τ_{MED} (=-2.18) is the median of the limiting distribution of $\hat{\tau}_1$ when $\alpha = 1$, and $k = [3n - \tau_{MED}^2(I_p + n)][\tau_{MED}(K + \tau_{MED})(I_p + n)]^{-1}$. The values of K and d_n are set to 5 and 0.290, as suggested by Roy and Fuller (2001). The function $C_p(\hat{\tau}_1)$ determines the magnitude of bias-correction, as a smooth and increasing function of the unit root statistic $\hat{\tau}_1$. The function $C_{-p}(\hat{\tau}_{-1})$ is constructed in a similar manner. However, as nearly all economic and business time series take the value of α equal to one or close to one, the value of $C_{-p}(\hat{\tau}_{-1})$ is practically zero for economic and business time series. Further details on the $C_p(\hat{\tau}_1)$ and $C_{-p}(\hat{\tau}_{-1})$ functions are given in Roy and Fuller (2001).

Once the value of $\hat{\alpha}^{RF}$ is found, (μ, β, ψ_1,…, ψ_{p-1}) can be estimated by regressing $Y_t - \hat{\alpha}^{RF} Y_{t-1}$ on (1, t, ΔY_{t-1},…, ΔY_{t-p+1}). The value of β is restricted to be zero when $\hat{\alpha}^{RF} = 1$. The parameter estimators for model (3) based on the Roy-Fuller estimator can be converted to those for model (2) using the relationships given in (4). They are denoted as $\hat{\gamma}^{RF} = (\hat{\mu}^{RF}, \hat{\beta}^{RF}, \hat{\alpha}_1^{RF}, ..., \tilde{\alpha}_p^{RF})$.

2.3 Andrews-Chen Estimator

The Andrews-Chen estimator attempts to find the approximately median-unbiased estimator for α in equation (3). Let $m(\alpha)$ be the median function of $\hat{\alpha}$ which is strictly increasing over the parameter space $\alpha \in (-1,1]$. Then, the Andrews-Chen estimator $\hat{\alpha}^{AC}$ of α is given by

$$\begin{aligned} \hat{\alpha}^{AC} &= 1 && \text{if } \hat{\alpha} > m(1), \\ &= m^{-1}(\hat{\alpha}) && \text{if } m(-1) < \hat{\alpha} \le m(1), \\ &= -1 && \text{if } \hat{\alpha} \le m(-1), \end{aligned}$$

where $m(-1) = \lim_{\alpha \to -1} m(\alpha)$ and m^{-1} is the inverse function of $m(\alpha)$. This median function can be computed using the Monte Carlo simulation method as described in Andrews and Chen (1994, p.192). As the simulation method calculates the $m(\alpha)$ function over a grid of α values in the interval (−1,1], a linear interpolation is conducted to approximate the value of $m^{-1}(\hat{\alpha})$, as suggested by Andrews (1993, p146). Note that this estimator is exactly median-unbiased for the AR(1) model, but approximately median-unbiased for the AR(p) model with $p > 1$ (for the details, see Andrews, 1993, and Andrews and Chen, 1994).

Once the value of $\hat{\alpha}^{AC}$ is found, (μ, β, ψ_1,…, ψ_{p-1}) are estimated by regressing $Y_t - \hat{\alpha}^{AC} Y_{t-1}$ on (1, t, ΔY_{t-1},…, ΔY_{t-p+1}). Similarly to the Roy-Fuller estimator, the value of β is set to zero when $\hat{\alpha}^{AC} = 1$. The above procedure can be iterated until convergence, as described in Andrews and Chen (1994, p.191). The parameter estimators for model (3) based on the Andrews-Chen estimator obtained in this way can be converted to those for model (2) using the relationships given in (4). They are denoted as $\hat{\gamma}^{AC} = (\hat{\mu}^{AC}, \hat{\beta}^{AC}, \hat{\alpha}_1^{AC}, ..., \tilde{\alpha}_p^{AC})$.

3 Bootstrap Prediction Intervals

This section outlines the construction of bootstrap prediction intervals using the bias-corrected estimators described in section 2. Since the approach is similar for each of the three approaches, we consider $\hat{\gamma}^c$, where $\hat{\gamma}^c \in \{\hat{\gamma}^B, \hat{\gamma}^{RF}, \hat{\gamma}^{AC}\}$, outlining any differences that arise. The procedure consists of the following steps.

Stage 1.

Given a realization $\{Y_t\}_{t=1}^n$ generated by equation (2), calculate the bias-corrected estimator $\hat{\gamma}^c$ using one of the three methods outlined in section 2. Let the corresponding residuals be denoted by $\{\tilde{u}_t\}_{t=p+1}^n$.

Stage 2.

Using $\hat{\gamma}^c$, generate a pseudo-data set recursively from:

$$Y_t^* = \hat{\mu}^c + \hat{\beta}^c t + \hat{\alpha}_1^c Y_{t-1}^* + \ldots \hat{\alpha}_p^c Y_{t-p}^* + u_t^*, \tag{6}$$

where the p starting values are set equal to the first p values of the original series, $Y_t^* = Y_t, t = 1,..., p$, and u_t^* is a random draw from $\{\tilde{u}_t\}_{t=p+1}^n$ made with replacement. Using this pseudo-data set $\{Y_t^*\}_{t=1}^n$, the parameters of the model given by (2) are re-estimated using the bias-corrected estimator adopted in Stage 1. The estimates so obtained are denoted by $\hat{\gamma}^{c*} = (\hat{\mu}^{c*}, \hat{\beta}^{c*}, \hat{\alpha}_1^{c*}, ..., \hat{\alpha}_p^{c*})$. Based on these estimates, the bootstrapped forecasts of $Y_{n+h,}$ made at period n, are generated recursively from:

$$Y_n^*(h) = \hat{\mu}^{c*} + \hat{\beta}^{c*}(n+h) + \hat{\alpha}_1^{c*} Y_n^*(h-1) + ... + \hat{\alpha}_p^{c*} Y_n^*(h-p) + u_{n+h}^*, \tag{7}$$

where $Y_n^*(j) = Y_{n+j}^* = Y_{n+j}$ for $j \le 0$ and u_{n+h}^* is a random draw from $\{\tilde{u}_t\}$ with replacement. Setting $Y_n^*(j) = Y_{n+j}$ for $j \le 0$ conditions the forecasts on the relevant history of the process. A number of bootstrap forecasts are generated by repeating the above: pseudo data sets are generated recursively from (6) with drawings of the errors from Stage 1, and for each of these

forecasts are calculated from (7) with the bias-corrected estimates. Carrying out this step B times gives rise to the bootstrap forecast distribution $\{Y_n^*(h;i)\}_{i=1}^B$.

Stage 3.

The prediction intervals for Y_{n+h} are based on the Efron percentile interval (e.g., Efron and Tibshirani, 1993), as this is widely used in the literature. Alternative methods include the Hall percentile interval (see e.g., Hall, 1992, or Davison and Hinkley, 1997), as well as percentile-*t* methods (see Efron and Tibshirani, 1993; p.160). The use of the latter is complicated by the need to calculate standard errors of the forecasts. Asymptotic formulae can be used, but their reliability is questionable with roots of unity or near-unity.

The Efron interval uses the percentiles of the empirical distributions (estimated at Stage 2) to define the upper and lower values of the prediction intervals, and is an intuitive approach to adopt. Thus, an interval with nominal coverage rate of 100(1-θ) per cent is given by $[Y_n^*(h,\tau), Y_n^*(h,1-\tau)]$, where $Y_n^*(h,\tau)$ is the 100τth percentile of the bootstrap distribution $\{Y_n^*(h;i)\}_{i=1}^B$, and $\tau = 0.5\theta$.

If we use OLS instead of a bias corrected estimator for the parameters of (7), the resulting interval is denoted by *PI*. PI_B, PI_{AC} and PI_{RF} denote intervals obtained using the bootstrap bias-corrected estimator, the Andrews-Chen, and the Roy-Fuller estimators, respectively.

When the Roy-Fuller estimator is used, $\hat{\gamma}^{c*}$ in Stage 2 is relatively simple to compute. But the computational burden can be substantial for the bootstrap bias-correction and the Andrews-Chen estimator. To lessen the burden, for the bootstrap bias-correction we follow Kilian (1998a) and use the estimate of the bias obtained at Stage 1, $Bias(\hat{\gamma})$, as an approximation to the biases at Stage 2. So on each of the B bootstrap samples the Stage 1 bias estimate is used to correct the estimated parameters.

For the Andrews-Chen estimator, we follow a similar strategy. Note that the median functions for $\{Y_t\}_{t=1}^n$ and each of the B $\{Y_t^*\}_{t=1}^n$ are different because of the presence of nuisance parameters (see Andrews and Chen, 1994). To reduce the computational burden, the median function associated with $\{Y_t\}_{t=1}^n$ is used to approximate those associated with each set of $\{Y_t^*\}_{t=1}^n$. The AR(1) model is a special case, however, because $\{Y_t\}_{t=1}^n$ and each set of $\{Y_t^*\}_{t=1}^n$ share the same median function, so that no such approximation is necessary.

If $\hat{\gamma}^c$ implies a unit root in Stage 1, the unit-root restriction is imposed on the parameter estimators $\hat{\gamma}^{c*}$ in Stage 2 for both the Andrews-Chen and Roy-Fuller estimators. As a consequence, the bootstrap replicates of AR forecasts given by (7) are generated from a model with a unit-root restriction when the observed data is found to have a unit root. This is in accordance with the bootstrap principle that the pseudo-data sets should mimic the original time series in their dependence structure. The treatment of the bootstrap bias-corrected estimator is a little different. OLS estimation of the model on the original data (Stage 1) rarely results in roots in the non-stationary region, so that whether these are imposed at Stage 2 makes little difference to the results. In the simulations, when non-stationary roots result at Stage 1, the bias correction is set to zero, and the Stage 2 estimator is simply the OLS estimator for each of the B bootstrap samples.

Because we allow for non-stationarity, the forward AR model is used to generate bootstrap replicates of the data (equation (6)). This contrasts the approach of Thombs and Schucany (1990) and Kim (2001). These authors use the backward AR model to generate the pseudo-data samples, so that the last p observations of the samples are equal to those of the actual data. This results in estimates of the parameters and forecasts that are conditioned on the actual values of the data for the last p observations. For non-stationary processes, the backward representation does not exist, and instead we condition the forecasts on the last p

observations by using (7) (but note that the parameter estimates are not conditioned on these observations).

In practice the order of the AR model will not be known, and will need to be selected based on the observed data. The uncertainty associated with the estimation of the lag order can be incorporated into the bootstrap interval estimation procedure by using the endogenous lag order bootstrap algorithm of Kilian (1998b). This entails selecting the lag order for each pseudo-data set. The three stage bootstrap procedure is adapted as follows. In Stage 1, the unknown order p is estimated from $\{Y_t\}_{t=1}^{n}$ using an appropriate order selection criterion, and is denoted as p^*. In Stage 2, we either generate the pseudo-data sets $\{Y_t^*\}_{t=1}^{n}$ using the estimated order p^*, which is then referred to as the exogenous lag order bootstrap algorithm, or the selection algorithm is applied anew, giving rise to "endogenous lag-order selection". Letting p^{**} denote the order selected on the bootstrap replicate, the bootstrap forecasts in equation (7) are then generated using either p^* or p^{**}. We use the AIC (Akaike, 1974) for lag order, following Kilian (1998b p.542). The maximum allowed lag order is set to 8, so that p^* and p^{**} are chosen so that the *AIC* is minimized over $1 \leq p \leq 8$. Setting the minimum order to one was not restrictive for the persistent processes that are the focus of our study.

4 Experimental Design

We consider AR models of orders 1 and 2. In terms of equation (1), the values of the AR(1) and AR(2) models' parameters are $a \in \{0.5, 0.9, 0.975, 1\}$ in $(1 - aB)\, v_t = u_t$ and $(1 - 0.5B)(1 - aB)\, v_t = u_t$. Initially, we assume that the disturbances are standard normal, $u_t \sim$ iid N(0,1). The intercept and trend are set to $\mu' = \beta' = 1$, although our formulation implies that the trend is excluded when $a = 1$, ruling out a quadratic trend. The number of Monte Carlo trials is set to 10,000, and the sample sizes simulated are 50, 100 and 200, with forecast horizons h ranging from 1 to 8. We assess the performance of the three methods at nominal coverage rates $(1-\theta)$ of 0.8 and 0.95. The methods are assessed in terms of the Monte Carlo estimates of the

average coverage rates of the prediction intervals. One run through of Stages 1 to 3 in section 3 constitutes a single Monte Carlo trial. The quantities of interest are estimated by averaging over Monte Carlo trials.

To calculate the coverage rates, we generate 500 sets of continuations (of length 8), or true future values, for each of the $\{Y_t\}_{t=1}^{n}$ generated at Stage 1, where each continuation is conditioned on the last *p* observations $\{Y_t\}_{t=n-p+1}^{n}$. For that Monte Carlo trial the coverage rate is estimated as the proportion of the continuations belonging to the prediction interval. For the *i*-th trial, we denote this proportion by C_i^* .The final estimate is then obtained by averaging these proportions over the Monte Carlo trials.

For the bootstrap bias-corrected estimator, the number of bootstrap iterations for bias-correction is set to 500. For each of the three methods, we set the number of bootstrap iterations to 500. The number of Monte Carlo trials to calculate the median function for the Andrews-Chen estimator is set to 100. The maximum number of iterations for the Andrews-Chen estimator is set to 10, which should be sufficient to achieve convergence.

5 Simulation Results

We first present results for when the model order is known. The section after discusses our findings when the lag order is unknown.

5.1. Known Model Order

Tables 1 and 2 present the results on coverage rates of 80% and 95% nominal intervals for AR(2) and AR(1) models. We begin with the results for the AR(2) model (Table 1). Looking at the first three panels for stationary models (a=0.5, 0.9, 0.975), it is evident that the actual coverage rates of the *PI* intervals are less than the nominal, and that the degree of under-prediction is increasing in the horizon (for a given sample size and degree of persistence), and in the degree of persistence, but is decreasing in the sample size. This is

because the small-sample bias increases with the persistence of the process, but is decreasing in the sample size. As expected, the standard bootstrap intervals (*PI*) coverage rates are consistent; for n=200 and h=1 the actual coverage rates are less than one and a half percentage points from the nominal for both intervals. The coverage rates of the PI_B intervals are generally markedly better than those of the *PI*, especially at long horizons for persistent processes, when the under-coverage of *PI* intervals is most acute. The PI_B intervals tend to over-estimate coverage at the nominal 80%, whilst the 95% intervals are generally well calibrated. For the more persistent of the three stationary cases, the PI_{RF} coverage rates are close to the nominal except at the smallest sample size for the longer forecasts. In these cases, the under-estimation of PI_{AC} is more marked. In general, the PI_{RF} intervals have the best coverage rates over the three cases considered.

For the unit-root process (Table 1, a=1), we find a similar pattern of results, except that the coverage rates of the PI_B intervals are now closer to the nominal for the 80% intervals. The PI_B intervals are best overall for the unit-root, although the PI_{RF} intervals are on a par except for n=50 and large h. The under-prediction of PI_{RF} in these circumstances is again exacerbated for PI_{AC}. The results suggest that the bootstrap works well for unit root AR(2) processes, even though we cannot formally establish its validity. We also experimented with PI_B without the ad hoc adjustment to prevent the bias-correction putting estimates into the non-stationary region of the parameter space. This leads to coverage rates in excess of the nominal for the 80% intervals when n=50 (e.g., for h=8, the actual coverage was 85%), but small improvements for the 95% intervals.

Table 1. Monte Carlo Estimates for Coverage Rates for AR(2) with Normal Errors

Nominal coverage = 80%												
	$n = 50$				$n = 100$				$n = 200$			
H	1	2	5	8	1	2	5	8	1	2	5	8
$a = 0.5$												
PI	75.5	73.2	68.7	67.6	77.7	76.6	74.1	73.7	78.8	78.3	76.9	76.7
PI_B	77.3	78.2	80.0	81.1	78.0	78.3	79.3	80.3	78.9	78.9	79.4	79.8
PI_{RF}	80.5	80.5	81.7	83.0	80.2	80.3	81.5	82.4	80.1	80.3	81.1	81.6
PI_{AC}	79.8	78.8	78.6	79.6	79.8	79.4	79.5	79.9	79.9	79.8	79.8	80.1
$a = 0.9$												
PI	74.8	71.8	63.0	57.2	77.6	76.3	71.5	67.7	78.7	78.1	75.6	73.4
PI_B	82.0	83.4	83.5	82.4	80.5	81.9	83.8	84.3	79.2	79.6	80.6	81.3
PI_{RF}	81.1	80.4	79.4	79.1	80.7	80.5	80.8	81.3	80.1	80.1	80.6	81.1
PI_{AC}	80.7	79.2	76.4	75.3	80.3	79.8	78.9	78.7	79.9	79.7	79.5	79.7
$a = 0.975$												
PI	74.5	71.0	61.5	54.8	77.4	75.8	70.4	65.4	78.6	77.9	75.0	72.0
PI_B	82.9	84.1	83.7	82.1	83.1	84.9	86.5	86.3	81.6	83.1	85.4	86.2
PI_{RF}	80.9	80.0	78.4	77.3	80.7	80.4	80.2	80.2	80.4	80.4	80.7	81.1
PI_{AC}	80.4	78.7	75.2	73.0	80.5	79.8	78.5	77.6	80.3	80.1	79.8	79.8
$\alpha = 1$												
PI	77.4	74.6	71.0	69.5	78.5	77.1	75.2	74.6	79.1	78.3	77.2	76.9
PI_B	79.1	78.8	79.2	80.0	79.0	78.8	79.0	79.7	79.2	78.9	78.8	79.1
PI_{RF}	80.8	79.9	78.2	76.9	80.6	80.2	79.8	79.5	80.3	80.2	80.2	80.3
PI_{AC}	80.3	78.5	74.8	72.5	80.4	79.7	78.1	76.9	80.2	79.9	79.2	78.7
Nominal coverage = 95%												
	$n = 50$				$n = 100$				$n = 200$			
H	1	2	5	8	1	2	5	8	1	2	5	8
$a = 0.5$												
PI	91.6	90.2	86.9	86.1	93.3	92.8	91.1	90.9	94.0	93.8	93.0	93.0
PI_B	92.8	93.3	93.6	93.8	93.4	93.7	94.2	94.6	94.1	94.2	94.4	94.6
PI_{RF}	94.7	94.9	95.1	95.4	94.7	95.0	95.4	95.8	94.8	95.0	95.4	95.6
PI_{AC}	94.3	94.0	93.5	93.8	94.5	94.4	94.4	94.5	94.7	94.7	94.8	94.9
$a = 0.9$												
PI	91.2	89.2	81.7	76.0	93.2	92.5	89.2	86.0	94.0	93.7	92.2	90.7
PI_B	94.7	95.1	94.0	92.6	94.7	95.3	95.7	95.4	94.3	94.5	95.0	95.2
PI_{RF}	94.9	94.7	93.7	92.8	95.0	95.1	95.1	95.1	94.8	95.0	95.2	95.4
PI_{AC}	94.7	93.9	91.9	90.9	94.9	94.7	94.1	93.7	94.7	94.7	94.6	94.7
$a = 0.975$												
PI	91.0	88.6	80.2	73.2	93.1	92.3	88.3	83.9	94.0	93.6	91.8	89.6
PI_B	95.1	95.3	93.9	92.1	95.8	96.3	96.3	95.6	95.5	96.0	96.5	96.5
PI_{RF}	94.8	94.5	93.1	91.8	95.0	95.0	94.8	94.5	95.0	95.1	95.3	95.3
PI_{AC}	94.5	93.7	91.2	89.4	94.9	94.7	93.8	93.1	95.0	95.0	94.8	94.7
$a = 1$												
PI	92.7	91.1	88.6	87.2	93.7	93.1	92.0	91.5	94.2	93.9	93.3	93.1
PI_B	93.8	93.7	93.7	93.8	94.0	94.0	94.1	94.4	94.2	94.2	94.2	94.3
PI_{RF}	94.7	94.4	93.0	91.6	94.9	94.9	94.6	94.1	94.9	95.0	95.0	95.0
PI_{AC}	94.4	93.6	90.1	89.1	94.8	94.6	93.6	92.7	94.9	94.8	94.5	94.2

Table 2. Monte Carlo Estimates for Coverage Rates for AR(1) with Normal Errors

Nominal coverage = 80%												
	$n = 50$				$n = 100$				$n = 200$			
h	1	2	5	8	1	2	5	8	1	2	5	8
$a = 0.5$												
PI	76.7	74.7	73.6	73.5	78.2	77.1	76.6	76.7	79.0	78.5	78.2	78.2
PI_B	77.1	77.5	78.9	79.6	78.3	78.5	79.1	79.4	79.1	79.2	79.5	79.5
PI_{RF}	80.1	80.8	82.4	82.9	80.2	80.7	81.6	81.9	80.1	80.4	81.0	81.0
PI_{AC}	79.8	80.4	81.7	82.4	79.8	79.9	80.3	80.4	80.0	80.0	80.1	80.2
$a = 0.9$												
PI	76.3	72.0	64.7	60.8	78.1	75.7	71.0	68.5	79.0	77.7	75.0	73.4
PI_B	81.9	82.8	82.9	82.6	80.8	81.0	82.4	83.2	79.3	79.5	80.2	80.8
PI_{RF}	80.9	81.2	81.9	82.4	80.6	81.0	82.0	82.9	80.1	80.4	81.1	81.9
PI_{AC}	80.3	79.7	78.5	78.1	80.1	80.1	79.7	79.7	79.9	79.8	79.7	79.7
$a = 0.975$												
PI	75.9	71.3	62.7	57.5	77.9	75.3	69.4	65.4	78.9	77.5	74.0	71.3
PI_B	83.3	84.1	83.4	82.2	83.1	84.6	85.7	85.6	81.7	83.1	85.0	85.6
PI_{RF}	80.8	80.9	80.6	80.3	80.7	80.9	81.4	81.6	80.4	80.7	81.4	81.9
PI_{AC}	80.1	79.2	76.8	75.3	80.2	79.9	79.0	78.2	80.2	80.3	80.2	80.1
$a = 1$												
PI	77.7	77.6	77.3	77.1	78.7	78.7	78.6	78.6	79.2	79.3	79.2	79.2
PI_B	78.1	78.4	79.1	79.6	78.8	79.0	79.2	79.5	79.3	79.3	79.4	79.5
PI_{RF}	80.8	80.7	80.2	79.6	80.6	80.8	80.8	80.8	80.3	80.5	80.7	80.9
PI_{AC}	80.1	79.0	76.2	74.3	80.1	79.8	78.5	77.3	80.2	80.1	79.6	79.1
Nominal coverage = 95%												
	$n = 50$				$n = 100$				$n = 200$			
h	1	2	5	8	1	2	5	8	1	2	5	8
$a = 0.5$												
PI	92.3	91.2	90.5	90.4	93.5	93.0	92.7	92.7	94.2	93.9	93.8	93.8
PI_B	92.6	92.9	93.6	93.8	93.6	93.8	94.1	94.2	94.2	94.3	94.5	94.5
PI_{RF}	94.4	94.8	95.5	95.6	94.7	95.1	95.5	95.6	94.8	95.1	95.4	95.4
PI_{AC}	94.3	94.7	95.7	96.0	94.6	94.7	94.9	95.0	94.7	94.8	94.9	94.9
$a = 0.9$												
PI	92.1	89.3	83.2	79.6	93.5	92.1	88.9	86.8	94.1	93.6	91.9	90.8
PI_B	94.9	94.8	93.9	93.1	94.6	95.0	95.1	95.0	94.3	94.5	94.8	95.0
PI_{RF}	94.8	95.0	94.8	94.6	94.9	95.3	95.6	95.7	94.9	95.1	95.4	95.7
PI_{AC}	94.4	94.2	93.1	92.5	94.7	94.8	94.4	94.1	94.7	94.8	94.7	94.6
$a = 0.975$												
PI	91.8	88.6	81.2	75.9	93.3	91.8	87.4	83.9	94.1	93.4	91.1	89.1
PI_B	95.4	95.2	93.6	92.4	95.8	96.1	95.8	95.2	95.5	96.0	96.3	96.2
PI_{RF}	94.7	94.9	94.2	93.4	94.9	95.2	95.3	95.1	95.0	95.3	95.6	95.7
PI_{AC}	94.3	93.9	92.2	90.9	94.7	94.7	94.0	93.3	94.9	95.0	95.0	94.7
$a = 1$												
PI	92.9	93.2	93.2	93.1	93.8	94.0	94.1	94.1	94.3	94.4	94.4	94.4
PI_B	93.1	93.6	94.2	94.5	93.8	94.1	94.3	94.5	94.3	94.4	94.5	94.6
PI_{RF}	94.7	94.8	94.0	93.1	94.9	95.1	95.0	94.7	95.0	95.2	95.2	95.2
PI_{AC}	94.3	93.8	91.9	90.3	94.7	94.6	93.8	92.9	94.8	94.9	94.7	94.3

The results for the AR(1) in Table 2 match in many respects the results for the AR(2). One difference is that the tendency of the PI_{RF} coverage rates to understate the nominal at small n is no longer apparent. The results for the unit-root case indicate that the bootstrap again provides a good approximation to the small-sample coverage. Overall, the results for the AR(1) favour the PI_{RF} intervals.

Also of interest is the variability of the coverage rates over Monte Carlo trials, which we measure by the Monte Carlo standard deviation (MCSD),

$MCSD = \sqrt{\sum_{i=1}^{R}(C_i^* - \overline{C}^*)^2 /(R-1)}$, where $\overline{C}^* = \sum_{i=1}^{R} C^*_i / R$ is the Monte Carlo estimate of the interval coverage rate, with C_i^* the estimate of the coverage rate on the ith replication. A high MCSD indicates high variability across replications: the $\overline{C}^*$ may be close to the nominal rate (of 80 or 90%), but the interval coverages vary more across replications – are less precise – than intervals generated by methods with low MCSDs. To save space, we do not report MCSDs for all the experiments, but comment on a number of cases of interest.

For the AR(2) DS process for small sample sizes (n=50), the MCSDs associated with the PI_B 80% intervals are smaller than for the PI intervals (ranging from 7.2 for h=1 to 16.4 for h=8, compared to 7.9 to 20.4), so that the bootstrap bias-correction yields intervals which are more accurate (average coverage closer to nominal) and more precise (smaller MCSDs). The dispersion of PI_{RF} and PI_{AC} interval coverage rates is similar to that of the PI_B, at 6.8 to 15.9, and 7.1 to 15.9, respectively, for 1 to 8-step forecasts. For n=100 the MCSDs of the PI_B and PI 80% intervals are lower at 4.8 to 10.7, and 5.0 to 12.5, and the MCSDs of PI_{RF} and PI_{AC} are again similar to those of PI_B.

By way of contrast, the MCSDs for the stationary AR(2) (with a=0.9) for n=50 are: PI, 9.4 to 21.4; PI_B, 10.4 to 21.8; PI_{RF}, 4.0 to 12.2; and PI_{AC}, 7.5 to 19.3 (for h=1 to 8, and 80% coverage). The greater precision of the Roy-Fuller bootstrap intervals suggests that such intervals are likely to be more reliable in practice. The superiority of PI_{RF} over PI_{AC} is related

to the finding by Roy and Fuller (2001, Table 3) that the mean squared error of their estimator is smaller than that of the Andrew-Chen estimator for the AR(1) model, especially when the value of the AR(1) coefficient is close to 1.

Table 3. Monte Carlo Estimates for Coverage Rates for AR(1) and AR(2) models with non-normal (Chi-squared (5)) errors

Nominal coverage = 80%												
	n = 50				n = 100				n = 200			
h	1	2	5	8	1	2	5	8	1	2	5	8
AR(2)												
a= 0.975												
PI	76.6	74.6	69.8	69.3	79.2	79.3	79.1	80.5	80.1	81.3	84.2	87.7
PI_B	85.3	87.0	88.9	90.0	85.6	88.1	92.1	94.3	83.9	86.6	92.4	96.1
PI_{RF}	83.5	84.8	89.7	94.1	83.1	85.3	92.6	98.3	82.8	85.4	93.9	99.6
PI_{AC}	83.0	83.6	87.4	92.5	82.9	84.8	91.7	97.7	82.7	85.2	93.6	99.6
a = 1												
PI	79.8	78.6	79.4	82.4	80.7	81.0	84.5	89.3	81.0	82.1	86.9	93.0
PI_B	80.1	80.8	84.8	88.5	80.7	81.8	86.9	92.1	80.9	82.4	87.9	94.0
PI_{RF}	83.5	84.9	90.5	96.1	83.1	85.4	93.3	98.9	82.8	85.5	94.4	99.8
PI_{AC}	83.0	83.7	88.5	95.2	82.8	84.9	92.5	98.8	82.7	82.5	94.0	99.9
AR(1)												
a = 1												
PI	80.1	81.9	86.6	89.5	80.8	82.8	88.8	94.1	81.2	83.4	90.1	96.4
PI_B	79.9	82.0	87.5	91.4	80.7	82.9	89.1	94.6	81.1	83.3	90.1	96.5
PI_{RF}	83.6	86.0	93.1	98.0	83.1	86.1	94.7	99.5	82.8	86.0	95.4	100.0
PI_{AC}	82.8	84.3	90.9	97.3	82.8	85.2	93.5	99.3	82.6	85.5	94.9	99.8
Nominal coverage = 95%												
	n = 50				n = 100				n = 200			
h	1	2	5	8	1	2	5	8	1	2	5	8
AR(2)												
a = 0.975												
PI	92.4	90.9	87.0	86.3	93.9	93.9	93.5	93.9	94.3	94.9	96.2	97.4
PI_B	96.1	96.4	96.5	96.7	96.6	97.3	98.3	98.8	96.2	97.1	98.7	99.5
PI_{RF}	95.7	96.2	97.6	98.5	95.5	96.6	98.8	99.7	95.2	96.6	99.2	100.0
PI_{AC}	95.3	95.5	96.7	98.1	95.4	96.3	98.6	99.6	95.1	96.5	99.1	100.0
a = 1												
PI	93.9	93.3	94.2	95.5	94.4	94.7	96.5	98.3	94.4	95.2	97.4	99.1
PI_B	94.2	94.3	96.1	97.3	94.4	95.1	97.2	98.8	94.4	95.3	97.7	99.2
PI_{RF}	95.6	96.2	97.8	99.1	95.4	96.6	98.9	99.9	95.1	96.6	99.2	100.0
PI_{AC}	95.3	95.6	97.0	98.5	95.3	96.3	98.7	99.8	95.0	96.4	99.1	100.0
AR(1)												
a = 1												
PI	93.7	94.9	97.4	98.6	94.1	95.5	98.1	99.4	94.3	95.7	98.4	99.6
PI_B	93.6	94.9	97.5	98.8	94.1	95.4	98.1	99.4	94.2	95.7	98.4	99.6
PI_{RF}	95.4	96.5	98.6	99.9	95.2	96.8	99.2	99.9	95.1	96.8	99.4	100.0
PI_{AC}	95.0	95.8	97.9	99.4	95.0	96.4	98.9	100.0	94.9	96.5	99.3	100.0

The results reported so far are obtained assuming normal innovations. As a check on the sensitivity of the results to this assumption, we re-ran the simulations with drawings of the innovations from a (centred) chi-squared distribution with five degrees of freedom, as an example of a non-symmetric distribution. The results are displayed in Table 3 for a stationary but persistent AR(2) (a=0.975), and for the AR(2) and AR(1) with unit roots. For the stationary AR(2) the coverage rate of PI is increasing in n, especially for large h. The degree of under-coverage is less than for normal errors at small n, but for n=200 the coverage at large h exceeds the nominal. The bootstrap is not consistent for the longer-horizon interval coverage rates. The PI_{RF} and PI_{AC} intervals exacerbate the over-coverage at the longer horizons, as do the PI_B but to a lesser extent. For the unit-root cases, the short-horizon intervals are essentially correctly sized for PI and PI_B, but the coverage is in excess of the nominal at the longer horizons, and this tendency is extenuated as the sample size increases. The PI_{RF} and PI_{AC} are over-sized throughout.

Other departures from normality could be considered (e.g., fat tails), but these results warn against the use of the bootstrap for highly persistent processes and skewed disturbances. We note that the Andrews-Chen estimator requires the normality of the disturbances, so that comparisons involving non-normal disturbances may not be warranted. That said, Andrews and Chen (1994, p.199) suggest that their estimator may be reasonably robust to departures from non-normality of the innovations.

5.2 Unknown Model Order

To allow for the possibility of under-specifying the model order, as well as choosing too many lags, we restrict our simulations to second-order data generating processes. The maximum lag order is set to 8, and the selection criterion is AIC. The results when the model order is known generally favour Roy-Fuller. This finding, and the relatively low computational costs of Roy-Fuller (coupled with the high costs of performing a Monte Carlo

evaluation when there is endogenous lag selection) cause us to confine our calculations to the Roy-Fuller estimator.

Table 4. Monte Carlo Estimates for Coverage Rates for AR(2) model using Roy-Fuller lag order selection

Nominal coverage = 80%												
	$n = 50$				$n = 100$				$n = 200$			
h	1	2	5	8	1	2	5	8	1	2	5	8
$a = 0.5$												
PI_{RF}	80.5	80.5	81.7	83.0	80.2	80.3	81.5	82.4	80.1	80.3	81.1	81.6
PI_{RFn}	79.5	78.0	77.7	79.0	79.5	78.8	79.3	80.2	79.9	79.6	79.9	80.3
PI_{RFx}	78.0	76.4	76.3	77.9	79.1	78.4	78.8	79.7	79.7	79.5	79.7	80.2
$a = 0.975$												
PI_{RF}	80.9	80.0	78.4	77.3	80.7	80.4	80.2	80.2	80.4	80.4	80.7	81.1
PI_{RFn}	79.8	77.8	73.6	70.8	80.1	79.3	77.6	76.4	80.1	79.8	79.2	78.9
PI_{RFx}	78.5	76.5	72.3	69.6	79.6	78.9	77.1	75.8	79.9	79.5	78.9	78.6
$a = 1$												
PI_{RF}	80.8	79.9	78.2	76.9	80.6	80.2	79.8	79.5	80.3	80.2	80.2	80.3
PI_{RFn}	79.8	77.5	73.1	70.2	80.0	79.2	77.3	75.8	80.0	79.6	78.8	78.2
PI_{RFx}	78.4	76.2	71.9	69.0	79.5	78.7	76.8	75.3	79.8	79.5	78.6	77.9
Nominal coverage = 95%												
	$n = 50$				$n = 100$				$n = 200$			
h	1	2	5	8	1	2	5	8	1	2	5	8
$a = 0.5$												
PI_{RF}	94.7	94.9	95.1	95.4	94.7	95.0	95.4	95.8	94.8	95.0	95.4	95.6
PI_{RFn}	94.2	93.5	93.0	93.5	94.4	94.1	94.3	94.7	94.7	94.7	94.8	95.0
PI_{RFx}	93.3	92.5	92.1	92.8	94.1	93.9	94.0	94.5	94.6	94.6	94.7	94.9
$a = 0.975$												
PI_{RF}	94.8	94.5	93.1	91.8	95.0	95.0	94.8	94.5	95.0	95.1	95.3	95.3
PI_{RFn}	94.3	93.2	90.1	87.7	94.7	94.4	93.4	92.4	94.8	94.8	94.5	94.3
PI_{RFx}	93.4	92.3	89.1	86.7	94.4	94.2	93.1	92.0	94.7	94.6	94.4	94.1
$a = 1$												
PI_{RF}	94.7	94.4	93.0	91.6	94.9	94.9	94.6	94.1	94.9	95.0	95.0	95.0
PI_{RFn}	94.3	93.0	89.8	87.2	94.7	94.3	93.2	92.1	94.8	94.7	94.3	93.9
PI_{RFx}	93.4	92.2	88.8	86.2	94.4	94.1	92.9	91.7	94.7	94.6	94.2	93.7

The results are summarised in Table 4, where we also repeat the findings from table 1 for the Roy-Fuller with known order, to aid comparison and bring out the impact of lag selection. We consider a clearly stationary case (a=0.5), the near-unit root case and a DS model. As one might expect, selecting the model order from the data, as opposed to using the true specifications, leads to a deterioration in average coverage rates of the derived intervals. This is especially so if no allowance is made for the order being estimated in the bootstrap

algorithm (contrast the rows labelled PI_{RF} and PI_{RFx}). The effects are then larger at the lower nominal coverage rate of 80%, for the smaller size and at longer horizons. For example, when n=50 and h=8, the coverage rates of the nominal 80% Roy-Fuller intervals are just under 70% compared to 77% when the model order is known (and a=0.975). These differences are also apparent when n=100, and are found for the unit root case, but the effects of model-order specification are small at the higher coverage rate when the sample size is larger than 50.

How successful is the process of endogenizing selection in the bootstrap algorithm? Comparing the rows labelled PI_{RFn} and PI_{RFx} in Table 4 indicates only small improvements, raising the average coverage rates by the order of one percentage point or so. While the marginal improvements are modest, it is worth recalling that even with unknown lag order the Roy-Fuller intervals are generally superior to the benchmark *PI* intervals when the model order is known.

We also consider the performance of the three methods for a higher-order AR process. We take as our example the AR(6) model of industrial production of Andrews and Chen (1994, Table 4). The model includes a deterministic trend, and has a root which is close to unity. In Table 5 we record the coverage rates when the lag order is assumed known, and for exogenous and endogenous lag order selection for the Roy-Fuller method. There are a number of noteworthy findings. Firstly, the small-sample performance of *PI* deteriorates dramatically relative to the low order AR(1) and AR(2) models, especially when the estimation sample is small (n=50). Secondly, the performance of the three methods PI_B, PI_{AC} and PI_{RF} is little affected by the higher dimensionality of the process. All three methods have reasonable coverage rates for samples of 50 and 100 (we omit results for n=200 to save space). Lastly, the degree of under-coverage that results from estimating the lag order for Roy-Fuller (PI_{RFx}) is less acute than for the AR(2) processes considered hitherto, and the offsetting improvements from endogenising the lag order selection within the bootstrap (PI_{RFn}) are a little larger, particularly for the lower coverage 80% intervals. In summary, our limited results

with higher-order processes (Table 5) do not suggest any reasons to doubt that our findings for low-order processes (in Tables 1 to 4) are likely to be useful more generally, at least for the three methods that take account of the small-sample biases of the least-squares estimator.

Table 5. Estimates for Coverage Rates for AR(6) data generating process for industrial production (taken from Andrews and Chen, 1994, Table 4).

Nominal coverage = 80%								
	$n = 50$				$n = 100$			
PI	70.5	66.8	59.8	57.5	75.6	73.8	69.8	68.3
PI_B	77.7	79.0	79.3	78.9	79.6	81.2	83.2	83.8
PI_{AC}	81.1	80.2	78.1	78.3	80.7	80.5	79.8	79.9
PI_{RF}	81.0	81.0	80.9	81.8	80.8	81.1	81.7	82.6
PI_{RFn}	79.1	77.6	75.9	77.0	79.2	78.3	77.7	79.3
PI_{RFx}	77.1	75.5	73.9	75.3	78.5	77.6	76.8	78.3
Nominal coverage = 95%								
	$n = 50$				$n = 100$			
PI	88.3	85.3	78.9	77.0	92.1	90.9	87.9	86.7
PI_B	92.4	92.4	91.4	90.9	94.2	94.7	95.0	94.9
PI_{AC}	95.0	94.3	92.5	92.5	95.0	95.0	94.4	94.2
PI_{RF}	94.9	94.9	94.4	94.5	95.1	95.3	95.5	95.7
PI_{RFn}	94.0	93.2	91.7	91.8	94.2	93.9	93.4	93.8
PI_{RFx}	92.7	91.8	90.3	90.7	93.8	93.4	92.8	93.3

6 Empirical Application

As an application, the industrial production data (annual from 1860 to 1988 with 129 observations) used in the previous section is again adopted. In order to compare alternative prediction intervals in a general setting, we have used a rolling window of 73 observations, while generating 1-step and 8-step ahead forecasts from each window. That is, we start from the 7th observation (ignoring the first six data points, 1860 - 1865), and take a sample of size 73 to the 79th observation. The AR model is specified and estimated, and used to generate forecasts for the 80th and 87th observations (1 and 8 step ahead forecasts, respectively). We continue this process to the last window that consists of the sample from the 49th to the 121st observations, generating forecasts for the 122nd and 129th observations. This provides 43 1-step and 8-step ahead forecasts. The order of the AR model is selected afresh on each window

of data using AIC. We calculate prediction intervals with nominal coverage rates of 80% and 95%.

Table 6: Mean and standard deviation of the lengths, and actual coverage rates, of alternative prediction intervals (industrial production data)

	PI	PI_B	PI_{RFx}	PI_{RFn}	PI_{AC}
			Nominal Coverage 80%		
h=1					
Mean	0.28	0.28	0.31	0.31	0.30
SD	0.02	0.02	0.02	0.02	0.02
Coverage	90.7	88.4	88.4	93.0	88.4
h=8					
Mean	0.53	0.84	0.86	0.89	0.79
SD	0.03	0.12	0.09	0.12	0.08
Coverage	100	100	100	100	100
			Nominal Coverage 95%		
h=1					
Mean	0.46	0.48	0.49	0.50	0.48
SD	0.02	0.02	0.02	0.03	0.02
Coverage	97.7	100	97.7	100	95.3
h=8					
Mean	0.83	1.31	1.33	1.38	1.23
SD	0.05	0.19	0.19	0.23	0.19
Coverage	100	100	100	100	100

We present the results in terms of the means and standard deviations of the lengths, and of the actual coverage rates, of the alternative prediction intervals in Table 6. When the nominal coverage is 80%, and for 1-step ahead forecasts, the intervals from the different methods all possess similar mean lengths (although those based on the Roy-Fuller and Andrews-Chen estimators are a little wider) and similar coverage rates (a difference of 2.3% in the coverage rate corresponds to one extra interval containing the actual). All intervals show similar degrees of variability. For 8-step ahead prediction intervals, however, the standard bootstrap intervals (*PI*) are notably shorter than the others, with smaller variability. All the 8-step intervals have coverage rates of 100%. These features are also evident when 95% prediction intervals are compared.

If we consider the nominal 80% 1-step ahead intervals, then standard binomial theory suggests an approximate 95% interval of (67.8,92.2) for the percentage coverage rate (assuming the nominal coverage rate of each interval is 80%, and independence across trials). The width of the interval reflects the small number of forecasts (43). The actual coverage rates (of all but PI_{RFn}) are just below the upper bound. For the 8-step ahead prediction intervals the actual coverage rates of 100% are clearly in excess of the nominal 80% level. A likely reason for the departure of the actual and nominal coverage rates is that the assumption that the simple autoregressive models are correctly-specified does not hold throughout the sample, and that the models' error terms may not be serially uncorrelated (as assumed in our re-sampling schemes). The period of 1860 to 1988 is likely to be characterised by structural breaks when modelled by simple autoregressions for the growth in industrial production. Clements and Hendry (2006) review the effects of structural breaks on forecast performance. This is outside our present concerns, but our empirical illustration serves to indicate that fixing estimation bias may not improve forecast performance when the model is mis-specified and when there are structural breaks (see also Clements and Hendry, 1999).

The empirical comparison indicates that at longer horizons differences between the various ways of bias-correcting are generally of a small magnitude compared to the difference between bias-correcting or not. Unfortunately, the 100% coverage rates in all instances prevent us from drawing any firm conclusions from this as to which is best.

7 Concluding Remarks

We have examined the small sample properties of a number of bootstrap prediction intervals for highly-persistent autoregressive models. All the methods we employ make use of the bootstrap, and should be superior to the traditional Box-Jenkins method, as shown by Thombs and Schucany (1990), *inter alia*.

We consider methods of generating intervals based on two estimators, the Andrews-Chen and Roy-Fuller estimators, and a bootstrap method, which are designed to counter the small-sample bias of least-squares estimators of persistent autoregressive processes. All three methods are successful in varying degrees at generating intervals with average coverage rates closer to the nominal. Our main finding is that the bootstrap prediction interval based on the Roy-Fuller estimator tends to perform well relative to the alternative intervals across a range of cases, including when the process contains a unit root.

We then present a further set of Monte Carlo simulations for the Roy-Fuller intervals which estimate the size of the gains from endogenizing lag-order selection in the bootstrap procedure when the model order is unknown. There are gains, but they turn out to be rather modest.

Table Legends

Table 4

RF_n denotes endogenous lag order selection (lag order selected on each of the bootstrap replicates) and RF_x exogenous selection.

Table 5

The model is $y_t = 0.9678\, y_{t-1} - 0.1566 y_{t-2} + 0.0949\, y_{t-3} - 0.0740\, y_{t-4} - 0.1566\, y_{t-5} + 0.2298\, y_{t-6} + 0.0758 + 0.0035t$.
RF_n denotes endogenous lag order selection (lag order selected on each of the bootstrap replicates) and RF_x exogenous selection.

Table 6

RF_n denotes endogenous lag-order selection (lag order selected on each of the bootstrap replicates) and RF_x exogenous selection.
The entries in coverage are in percentages.

References

Akaike, H., 1974. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, AC-19, 716-723.

Alonso, A. M., Pena, D., Forecasting Time Series with Sieve Bootstrap. Journal of Statistical Planning and Inference, 100, 1-11.

Andrews, D. W. K., 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica, 61, 139-165.

Andrews, D. W. K., Chen, H.-Y., 1994. Approximate median-unbiased estimation of autoregressive models. Journal of Business & Economic Statistics, 12, 187-204.

Berkowitz, J., Kilian, L., 2000. Recent developments in bootstrapping time series. Econometric Reviews, 19, 1-48.

Box, G. E.P., Jenkins, G. M., 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Breidt, F. J., Davis, R. A., Dunsmuir, W. T., 1995. Improved Bootstrap Prediction Intervals for Autoregressions. Journal of Time Series Analysis, 16, 177-200.

Chatfield, C., 1993. Calculating Interval Forecasts. Journal of Business & Economic Statistics, 11, 121-135.

Christiano, L. J., Eichenbaum, M., 1990. Unit Roots in Real GNP: Do we know and do we care? Carnegie Rochester Conference Series on Public Policy, 32, 7-82.

Clements, M. P., Hendry, D. F., 1999. Forecasting Non-Stationary Economic Time Series. The MIT Press, Cambridge, Massachusetts.

Clements, M. P., Hendry, D. F., 2001. Forecasting with Difference-Stationary and Trend-Stationary models. Econometrics Journal, 4, S1-S19.

Clements, M. P., Hendry, D. F., 2006. Forecasting with Breaks. In Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24, (eds) Elliott, G. and Granger, C.W.J and Timmermann, A. 605-657, Elsevier, North-Holland.

Clements, M. P., Taylor, N., 2001. Bootstrapping Prediction Intervals for Autoregressive Models. International Journal of Forecasting, 17, 247-267.

Davison, A.,Hinkley, D., 1997. Bootstrap Methods and their Applications. Cambridge University Press, Cambridge.

Diebold, F. X., Senhadji, A. S., 1996. The Uncertain Unit-Root in US GNP: Comment. American Economic Review, 86, 1291-1298.

Efron, B., 1979. Bootstrap methods: another look at the jacknife. Annals of Statistics, 9, 1218-1228.

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap.Chapman & Hall, New York.

Gospodinov, N., 2002. Median unbiased forecasts for highly persistent autoregressive processes. Journal of Econometrics, 111, 85-101.

Grigoletto, M., 1998. Bootstrap Prediction Intervals for Autoregressions: Some Alternatives. International Journal of Forecasting, 14, 447-456.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Hansen, B. E., 1999. The grid bootstrap and the autoregressive model. Review of Economics and Statistics, 81, 594-607.

Inoue, A., Kilian, L., (2002. Bootstrapping Autoregressive Processes with Possible Unit Roots. Econometrica, 70(1), 377-391.

Kabaila, P., 1993. On Bootstrap Predictive Inference for Autoregressive Processes. Journal of Time Series Analysis, 14,_ 473-484.

Kilian, L., 1998a. Small sample confidence intervals for impulse response functions. The Review of Economics and Statistics, 80, 218 - 230.

Kilian, L., 1998b. Accounting for Lag-Order Uncertainty in Autoregressions: the Endogenous Lag Order Bootstrap Algorithm. Journal of Time Series Analysis, 19, 531-528.

Kim, J. H., 1999. Asymptotic and Bootstrap Prediction Regions for Vector Autoregression. International Journal of Forecasting, 15, 393-403.

Kim, J. H., 2001. Bootstrap-after-Bootstrap Prediction Intervals for Autoregressive Models. Journal of Business & Economic Statistics, 19, 117-128.

Kim, J. H., 2002. Bootstrap Prediction Intervals for Autoregressive Models of Unknown or Infinite Lag Order. Journal of Forecasting, 21, 265-280.

Li, H., Maddala, G. S., 1996. Bootstrapping time series models. Econometric Reviews, 15, 115-158.

MacKinnon, J. G., 2002. Bootstrap inference in econometrics. Canadian Journal of Economics, 35, 615-645.

MacKinnon, J. G., Smith, A. A., Jr., 1998. Approximate bias correction in econometrics. Journal of Econometrics, 85, 205-230.

Masarotto, G., 1990. Bootstrap Prediction Intervals for Autoregression. International Journal of Forecasting, 6, 229-329.

Nelson, C. R., Plosser, C. I., 1982. Trends and random walks in macroeconomic time series. Journal of Monetary Economics, 10, 129-162.

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., 2006. Bootstrap prediction for returns and volatilities in GARCH models. Computational Statistics and Data Analysis, 50, 2293-2312.

Roy, A., Fuller, W. A., 2001. Estimation for autoregressive time series with a root near one. Journal of Business & Economic Statistics, 19(4), 482-493.

Thombs, L. A., Schucany, W. R., 1990. Bootstrap prediction intervals for autoregression. Journal of the American Statistical Association, 95, 486-492.