ELSEVIER

Journal of Econometrics 111 (2002) 85–101

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Median unbiased forecasts for highly persistent autoregressive processes

Nikolay Gospodinov*

Department of Economics, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, Canada H3G 1M8

Abstract

This paper considers the construction of median unbiased forecasts for near-integrated autoregressive processes. It derives the appropriately scaled limiting distribution of the deviation of the forecast from the true conditional mean. The dependence of the limiting distribution on nuisance parameters precludes the use of the standard asymptotic and bootstrap methods for bias correction. We propose a bootstrap method that generates samples backward in time and approximates the median function of the predictive distribution on a grid of values for the nuisance parameter. The method can be easily adapted to approximate any quantile of the conditional predictive distribution. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C12; C15; C22

Keywords: Near-integrated AR process; Conditional predictive inference; Local-to-unity asymptotics; Bootstrap

1. Introduction

The autoregressive (AR) models provide a flexible parametric framework for approximating the dynamic behavior of many economic time series. Despite their simple structure, some serious inference problems arise when the largest autoregressive root is near or on the unit circle. The last two decades witnessed a surge of research efforts for developing a full-blown statistical theory for parameter inference in the unit root case. In some cases, however, the ultimate goal of the analysis is not estimation and inference on the AR parameters but construction of point and interval predictors of the variable of interest y conditional on its realizations up to time T. Although most of the empirical studies are concerned with the conditional mean of y, obtaining the

* Tel.: +1-514-848-3935; fax: +1-514-848-4536.
E-mail address: gospodin@vax2.concordia.ca (N. Gospodinov).

PII: S0304-4076(02)00123-9

complete conditional predictive distribution of y would provide very valuable information for policy analysis and business decisions. For example, computing the median of the distribution would give us a more robust point predictor whereas some relevant quantiles would help us assess the uncertainty associated with the point forecast. In this paper, we propose a method for conditional median unbiased forecasting of nearly nonstationary AR processes. We discuss in full detail the zero-mean AR(1) model and derive the asymptotic and bootstrap approximations of the conditional distribution of the forecasts. These approximations are used for constructing median unbiased forecasts.

The development of the method draws on two strands of literature. The first line of research deals with the construction of median unbiased parameter estimates when the largest AR parameter is close to unity. Stock (1991), Andrews (1993) and Hansen (1999) show how to obtain median unbiased estimates of the largest AR root and the sum of all AR coefficients by inverting the median function of the asymptotic and bootstrap distributions of the OLS estimator and the t-statistic. Stock (1996, 1997) introduced a general framework for constructing unbiased long-run forecasts for highly persistent AR processes based on median unbiased estimates and unit root pretests. However, the interval predictors that he discussed appeared to suffer from strong conditional bias for some combinations of the AR parameter and the last value of the process.

The second literature is concerned with the simulation of conditional bootstrap sample paths for strictly stationary AR(p) processes that pass through the last p observations of the original series. This allows us to study the effect of the parameter variability on the conditional prediction by preserving the dependence of the parameter estimates on the data used for forecasting. Some efficient algorithms for improved bootstrap prediction inference in stationary Gaussian and non-Gaussian AR models are discussed in Thombs and Schucany (1990), Breidt et al. (1992, 1995) and Kabaila (1993) among others.

The paper is organized as follows. Section 2 discusses the problem of conditional predictive inference in the context of a zero-mean, nearly integrated AR(1) model. It derives the conditional limiting representation of the deviation of the forecast from its true mean. The analytical results motivate the use of the grid bootstrap method (Hansen, 1999) for construction of median unbiased forecasts and prediction intervals. In Section 3, we discuss different procedures for generating conditional bootstrap samples and suggest a consistent bootstrap estimator of the predictive distribution. The results from a small-sample simulation study are presented in Section 4. We apply the procedure for median unbiased forecasting to one-month T-bill yields. Section 6 concludes.

2. Predictive inference for nearly integrated AR(1) processes

2.1. Model and notation

Consider the zero-mean AR(1) model

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where $|\alpha| \leqslant 1$, $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ and y_0 is assumed fixed.[1] It is well documented that the AR representation of many economic time series has a root near or on the unit circle. Therefore, in the subsequent analysis we will adopt the local-to-unity framework that reparameterizes the AR coefficient as $\alpha_T = 1 + c/T$, where $c \leqslant 0$ is a finite constant and T is the sample size (Chan and Wei, 1987; Phillips, 1987).

In this study, we are interested in constructing forecasts for the future value y_{T+k} conditional on the last observation y_T when α is near one. Due to the recursive first-order Markov structure of the AR(1) model, the conditional predictive distribution of the future value y_{T+k} depends only on the last observation of the process. The unconditional predictive inference ignores the information contained in the last realized value of the series and averages over all possible values of y_T to obtain the predictive distribution. Moreover, when the conditional predictive distribution is asymmetric, integrating out the value of y_T symmetrizes the shape of the distribution and the forecast appears to be unbiased.

Throughout the paper we use the following notation. The true k-step ahead future value of y is denoted by $y_{T+k} = \alpha^k y_T + \sum_{j=0}^{k-1} \alpha^j \varepsilon_{T+k-j}$, the true conditional mean by $\bar{y}_{T+k|T} = \mathrm{E}(y_{T+k} \mid y_T) = \alpha^k y_T$ and the predictive conditional mean by $\hat{y}_{T+k|T} = \hat{\alpha}^k y_T$, where $\hat{\alpha}$ is the OLS estimator $\sum_{t=2}^{T} y_t y_{t-1} / \sum_{t=2}^{T} y_{t-1}^2$. The latter delivers the linear predictor of y that minimizes the mean square forecast error. The forecast error is given by $\hat{e}_{T+k|T} = \hat{y}_{T+k|T} - y_{T+k} = (\hat{\alpha}^k - \alpha^k) y_T - \sum_{j=0}^{k-1} \alpha^j \varepsilon_{T+k-j}$, and the deviation of the forecast from the true conditional mean is $\bar{e}_{T+k|T} = \hat{y}_{T+k|T} - \bar{y}_{T+k|T} = (\hat{\alpha}^k - \alpha^k) y_T$.

In AR models, the dependence between the data used for estimation and the data used for prediction introduces some difficulties in assessing the effects of the parameter variability and the properties of the estimators on the forecasts (Phillips, 1979). Unlike the stationary case, the dependence between $\hat{\alpha}$ and y_T does not vanish asymptotically for AR processes with a root on or near the unit circle. This can be seen from the joint limiting representations of the OLS estimator of α and the last observed value of the series which is given by (Phillips, 1987)

$$(T(\hat{\alpha} - \alpha), T^{-1/2} y_T) \Rightarrow \left(\frac{1}{2} \frac{J_c(1)^2 - J_c(0)^2 - 2c \int_0^1 J_c(r)^2 \, \mathrm{d}r - 1}{\int_0^1 J_c(r)^2 \, \mathrm{d}r}, J_c(1) \right) \tag{2}$$

where $\alpha = 1 + c/T$, $\{J_c(r) \colon r \in [0,1]\}$ is a homogeneous Ornstein–Uhlenbeck process generated by the stochastic differential equation $\mathrm{d}J_c(r) = cJ_c(r) + \mathrm{d}W(r)$, $W(r)$ is the standard Brownian motion and $\Rightarrow$ denotes weak convergence.

Let $g_T = T^{-1/2}(\hat{y}_{T+k|T} - \bar{y}_{T+k\,|\,T})$ and $h_T = T^{-1/2}(\hat{y}_{T+k|T} - y_{T+k})$ denote the normalized deviation of the OLS forecast from its true conditional mean and the normalized forecast error, respectively. For the purposes of this paper, we need to approximate the distributions of g_T and h_T conditional on the value taken by y_T. Since y_T is $\mathrm{O_p}(T^{1/2})$, it diverges as $T \to \infty$. Thus, it will be more appropriate to condition on

[1] The asymptotic results derived in this paper are not affected if the initial condition is assumed to be $\mathrm{o_p}(T^{1/2})$.

the rescaled terminal value $T^{-1/2}y_T$. Define $G_T(z\,|\,x) = \Pr(g_T \leqslant z\,|\,T^{-1/2}y_T = x)$ and $H_T(z\,|\,x) = \Pr(h_T \leqslant z\,|\,T^{-1/2}y_T = x)$ to be the conditional distributions of g_T and h_T. The goal of the paper is to obtain the asymptotic and bootstrap approximations of these exact finite-sample conditional distributions.

The derivation of the asymptotic limits of the conditional distributions $G_T(z\,|\,x)$ and $H_T(z\,|\,x)$, however, poses some practical problems. One approach is to derive the conditional distribution of $(T(\hat{\alpha}-\alpha)\,|\,T^{-1/2}y_T)$ from the joint unconditional distribution of $(T(\hat{\alpha}-\alpha), T^{-1/2}y_T)$ in (2) which is a mixture of a non-standard and a Gaussian distribution. This would involve inversion of the joint characteristic function of $T(\hat{\alpha}-\alpha)$ and $T^{-1/2}y_T$ and then conditioning on $T^{-1/2}y_T = x$ to obtain the conditional density of interest.

Alternatively, the conditioning on the last value of the time series arises naturally if we consider the backward (time-reversed) representation of the process. For the zero-mean AR(1) process in Eq. (1) the backward representation has the form

$$y_t = \rho y_{t+1} + \xi_t = \rho^{T-t} y_T + \sum_{i=0}^{T-t-1} \rho^i \xi_{t+i}, \quad t = T-1, T-2, \ldots, 1, \tag{3}$$

where ξ_t denotes the backward noise. Thus, we can employ the functional central limit theorem and the relationship between the OLS estimators of ρ and α, $\hat{\rho} = \hat{\alpha}[1-(y_T^2 - y_1^2)/\sum y_t^2]$, to obtain the conditional distribution of $(\hat{\alpha}-\alpha)$ for a fixed x. Instead, we take a third approach and focus on the limiting approximation of the product $(\hat{\alpha}^k - \alpha^k)y_T$ conditional on $T^{-1/2}y_T$ taking a particular value.

2.2. Asymptotic representations of forecast errors

In this section, we derive the conditional asymptotic representation of the appropriately normalized forecast errors $\bar{e}_{T+k|T}$ and $\hat{e}_{T+k|T}$. The limiting behavior of these quantities is formalized in Stock (1996), Phillips (1998) and Kemp (1999). Lemma 1 restates these results in the context of our model.

Assumption 1. *In model $y_t = \alpha_T y_{t-1} + \varepsilon_t$ with $\alpha_T = 1 + c/T$, the innovations sequence ε_t is iid with $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t^2) = 1$ and $\mathrm{E}|\varepsilon_t|^\eta < \infty$ for some $\eta > 2$.*

Lemma 1. *Let $k = [\lambda T]$, where λ is a fixed constant and $[.]$ denotes the integer part of the argument. Then, under Assumption 1 and as $T \to \infty$,*

$$(g_T\,|\,T^{-1/2}y_T = x) \Rightarrow (g(c,\lambda,x)\,|\,x), \tag{4}$$

$$(h_T\,|\,T^{-1/2}y_T = x) \Rightarrow (h(c,\lambda,x)\,|\,x), \tag{5}$$

where

$$g(c,\lambda,x) = \exp(\lambda c)\left\{\exp\left(\frac{\lambda}{2}\,\frac{x^2 - J_c(0)^2 - 2c\int_0^1 J_{c|x}(r)^2\,\mathrm{d}r - 1}{\int_0^1 J_{c|x}(r)^2\,\mathrm{d}r}\right) - 1\right\}x,$$

$$h(c,\lambda,x) = g(c,\lambda,x) + \mathrm{N}\left(0, \frac{\exp(2c\lambda) - 1}{2c}\right),$$

$J_c(r) = \exp(cr)J_c(0) + \exp(cr)\int_0^r \exp(-cs)\,\mathrm{d}W(s)$ *is the Ornstein–Uhlenbeck process,* $J_{c|x}(r)$ *is the path consistent with the scaled terminal value* x.

Proof. See Appendix A.

A few remarks on the results of Lemma 1 are worth mentioning. In the lemma, we treat the lead time k as a fraction of the sample size, i.e. it approaches infinity at the same rate as T. Letting the forecast horizon grow with the sample size is better suited for long-term forecasting. Also, this treatment appears to be more convenient for the asymptotic analysis of the forecast error. Finally, this parameterization allows the asymptotic approximation to preserve the parameter estimation uncertainty present in the finite sample distribution whereas in the fixed forecast horizon case this uncertainty vanishes asymptotically as k/T goes to 0.[2]

Lemma 1 shows that the conditional asymptotic representations of $(\hat{y}_{T+k|T} - \bar{y}_{T+k|T})$ depend on the values of c, λ and the last observation y_T.[3] Since c is not consistently estimable, we can simulate the conditional limiting distribution of $(\hat{y}_{T+k|T} - \bar{y}_{T+k|T})$ on a grid of values for c for any given λ and y_T. The dependence of the conditional representation on the data through y_T precludes a prior tabulation of the asymptotic representations in Lemma 1 and requires their computation after the data (and y_T) become available. Since this may be impractical in applied work, we use the grid bootstrap method, proposed by Hansen (1999), whose properties are discussed in Section 3.

2.3. Median unbiased forecasts

The normalized deviation of the OLS forecast from its true conditional mean $g_T = T^{-1/2}(\hat{y}_{T+k|T} - \bar{y}_{T+k|T})$ is a function of the data Y, the forecast horizon k and the local-to-unity parameter c. Since the data and the forecast horizon are predetermined at the time of the forecast, we parameterize g_T only as a function of the parameter c and denote it by $g(c)$.

Let $G_T(x\,|\,c) = \Pr\{g(c) \leqslant x\,|\,c\}$ denote the sampling distribution of $g(c)$. Suppose that the 0.5th quantile $q_{0.5}(c)$ of the distribution function $G_T(x\,|\,c)$, which is given

[2] Sampson (1991) suggests a more flexible framework by setting $T = (1/\lambda)(k^\tau)$ with $\tau \geqslant 0$. This framework nests our specification when $\tau = 1$.

[3] If σ is not normalized to 1, we can obtain the same limiting representations as in Lemma 1 by replacing y_T with the standardized quantity y_T/σ.

by $\Pr\{g(c) \leqslant q_{0.5}(c) \mid c\} = 0.5$, is uniquely defined and monotonically increasing in c. Then, the median unbiased forecast is obtained by inverting the median function $\hat{y}^{\mathrm{MU}}_{T+k|T} = \hat{y}^{\mathrm{OLS}}_{T+k|T} - T^{1/2} q_{0.5}^{-1}(g_T)$ such that[4]

$$\Pr\{\bar{y}_{T+k|T} \leqslant \hat{y}^{\mathrm{MU}}_{T+k|T}\} = \Pr\{\bar{y}_{T+k|T} \geqslant \hat{y}^{\mathrm{MU}}_{T+k|T}\} = 0.5.$$

The median unbiased forecasts possess the impartiality property that the probability of underprediction is equal to the probability of overprediction. The median forecast is the optimal predictor under symmetric "linlin" loss function (Granger, 1969). The procedure can be easily accommodated to fit any asymmetric "linlin" loss function by substituting the corresponding quantile q_θ of the prediction distribution for the median such that $\Pr\{\bar{y}_{T+k|T} \leqslant q_\theta^{-1}[g(c)]\} = \theta$.

Finally, it will be instructive to compare the (classical) method proposed in this paper to the Bayesian approach to predictive inference. The Bayesian approach implicitly incorporates the model uncertainty without the necessity of repeated sampling. The predictive density of y_{T+k} is given by

$$\begin{aligned} p(y_{T+k} \mid y_T) &= \int p(y_{T+k}, \alpha, \sigma^2 \mid Y)\, \mathrm{d}\alpha\, \mathrm{d}\sigma^2 \\ &= \int p(y_{T+k} \mid \alpha, \sigma^2, Y)\, p(\alpha, \sigma^2 \mid Y)\, \mathrm{d}\alpha\, \mathrm{d}\sigma^2, \end{aligned} \tag{6}$$

where the integration is over $\sigma^2 > 0$ and the parameter space of α, $Y = (y_1, \ldots, y_T)$ and $p(\alpha, \sigma^2 \mid Y)$ is the posterior distribution of the parameters.

The form of the posterior $p(\alpha, \sigma^2 \mid Y)$ is determined by the choice of a prior distribution of the parameters $p(\alpha, \sigma^2)$. A typical choice of a prior is the flat prior (Zellner, 1971, Chapter 7) which is considered uninformative about the parameters. This prior implies a posterior proportional to the normal-inverted gamma distribution which facilitates the numerical evaluation of the integral in (6) for $k > 1$. However, Phillips (1991) argued that a flat prior on the parameters in AR models is informative in some regions of the parameter space and suggested the use of Jeffreys' prior. Unfortunately, Jeffreys' prior suffers from other drawbacks summarized in Bauwens et al. (1999) and the debate on an appropriate prior in AR models is still not fully resolved.

3. Bootstrap approximation of predictive distribution

3.1. Conditional bootstrap samples

This section discusses some bootstrap methods that retain the last observation of the original series used for prediction in AR(1) models across all samples. The conventional bootstrap methods do not preserve the dependence between $\hat{\alpha}$ and y_T and would not be appropriate for conditional predictive inference. Therefore, the bootstrap methods need

[4] For more details, see Stock (1991), Andrews (1993) and the references therein.

to generate conditional sample paths that pass through the last value of the original series, i.e. $y_T^* = y_T$ where $\{y_t^*\}_{t=1}^T$ is the bootstrap sample. For strictly stationary time series, there are several methods for generating conditional bootstrap samples. For instance, we can use the properties of the forward and backward representations of a time series to obtain approximate bootstrap samples conditional on the last observation.

If $|\alpha| < 1$, the autocovariance generating function of the AR(1) process y in (1) is given by $\gamma_y(L) = (1-\alpha L)^{-1}(1-\alpha L^{-1})^{-1}\sigma^2$. The above expression shows that the processes $(1-\alpha L)y_t$ and $(1-\alpha L^{-1})y_t$ have identical second-order properties and if the errors are normally distributed, the distributions of the forward and backward errors are identical. This observation has led Thombs and Schucany (1990) to suggest a direct resampling of the backward residuals. However, if the errors are non-Gaussian, reversing time would produce uncorrelated but not independent backward disturbances. Since the naive bootstrap treats the errors as if they were *iid*, resampling the backward residuals would not be valid for non-Gaussian processes. Breidt et al. (1992, 1995) proposed an alternative algorithm that generates a stationary sequence of backward residuals as a function of the resampled forward residuals. The procedure, however, is developed for strictly stationary processes and its properties are not known when the AR root is near or exactly on the unit circle.

As pointed out in the introductory remarks, it is a stylized fact that many economic time series are highly persistent processes and the appropriate parameterization of the largest AR root would be $\alpha = 1 + c/T$. Define the forward and backward quasi-differences as $\Delta_{t,t-1}^c = y_t - (1+c/T)y_{t-1}$ and $\Delta_{t,t+1}^c = y_t - (1+c/T)y_{t+1}$. The corresponding restricted and recentered residuals are $\bar{\varepsilon}_t = \Delta_{t,t-1}^c - \bar{\Delta}_{t,t-1}^c$ and $\bar{\xi}_t = \Delta_{t,t+1}^c - \bar{\Delta}_{t,t+1}^c$ with distribution functions $F_{\bar{\varepsilon}}$ and $F_{\bar{\xi}}$, respectively, where $\bar{\Delta}$ denotes the mean of Δ. Let $\{\varepsilon_t\}_{t=1}^T$ denote the true (unobserved) forward innovations with a population distribution F_ε and $\{\bar{\xi}_t^*\}_{t=1}^T$ is a bootstrap sequence of backward residuals. Then, we have the following result.

Theorem 1. *Under Assumption* 1 *and as* $T \to \infty$,

$$d_2(F_{\bar{\xi}^*}, F_{-\varepsilon}) \to 0,$$

where $d_2(.)$ *is the Mallows metric*[5] *of degree* 2, *defined as* $d_2(F_X, F_Z) = \inf(\mathrm{E}|X - Z|^2)^{1/2}$ *over all joint distributions for the random variables* X *and* Z *with marginal distributions* F_X *and* F_Z.

Proof. See Appendix A. □

Theorem 1 establishes the asymptotic validity of the bootstrap resampling of the backward residuals in nearly integrated AR models. This result also suggests that for near-integrated processes, the resampling of the backward residuals is asymptotically equivalent to the forward residuals resampling. To ensure the validity of the bootstrap over the entire parameter space, however, we need to employ a block bootstrap

[5] For the properties of the Mallows metric, see Section 8 in Bickel and Freedman (1981).

procedure that would preserve the higher-order dependence when the parameter is away from the nonstationary boundary.

3.2. Consistent bootstrap estimator of predictive distribution

Let $F_0(x) = \Pr\{y_t \leqslant x\}$ denote the population distribution of the random sample $y_1, y_2, \ldots, y_T$ that belongs to a family of distribution functions Ξ, and $F_T(x) = (1/T) \sum_{t=1}^{T} \mathbf{1}\{y_t \leqslant x\}$ be its empirical distribution function. The quantity of interest is $g_T = T^{-1/2}(\hat{y}_{T+k|T} - \bar{y}_{T+k|T})$ with asymptotic distribution $G_\infty(z, F_0)$ given in (4). The bootstrap estimator of the distribution of g, $G_T^*(z, F_T)$, approximates the exact finite sample distribution $G_T(z, F_0)$ by replacing the unknown distribution function F_0 with an estimator. The general conditions for the consistency of the bootstrap distribution $G_T^*(z, F_T)$ are given in Beran and Ducharme (1991) and Horowitz (2001).

The consistency of the bootstrap estimator requires a continuity of the mapping from the sampling distribution of the data to the limiting distribution of the test statistic. It is well known that the distribution of the AR parameter is discontinuous at $\alpha = 1$ which violates this condition and invalidates the conventional bootstrap (Basawa et al., 1991a). Even if we recast the AR parameter in the local-to-unity framework that provides a smooth transition from the Gaussian approximation to the Dickey–Fuller distribution, the bootstrap based on the OLS estimates is still invalid since the limiting representation depends on the local-to-unity parameter c which is not consistently estimable. In order to obtain a consistent bootstrap estimator, Hansen (1999) suggested a method that approximates the distribution of the AR parameter on a grid of points for α. A similar remedy is proposed by Basawa et al. (1991b) in the unit root case that imposes the null hypothesis of $\alpha = 1$ in generating the bootstrap data. The following theorem shows the asymptotic validity of the bootstrap predictive distribution of autoregressive processes with a root near or on the unit circle.

Theorem 2. *Under Assumption* 1, *for any fixed* $\delta > 0$ *and* $F_0 \in \Xi$,

$$\lim_{T\to\infty} \Pr\left\{\sup_z |G_T^*(z, F_T) - G_\infty(z, F_0)| > \delta\right\} = 0.$$

Proof. See Appendix A. □

Theorem 2 and the continuity of the local-to-unity framework ensure the consistency of the grid bootstrap based on the backward restricted residuals. Theorem 2 shows that the quantile predictors obtained by the grid bootstrap are first-order asymptotically correct. The consistent estimation of the quantiles of the bootstrap predictive distribution implicitly assumes that the number of bootstrap replications and the number of grid points approach infinity. Some recommendations about the choice of the number of bootstrap repetitions can be found in Andrews and Buchinsky (2000) and Davidson and MacKinnon (2000). The computationally intensive construction of a fine grid for the AR parameter can be avoided by smoothing nonparametrically the quantile functions (Hansen, 1999) or efficient algorithms for inverting the quantile function such as the

bisection and the Robbins–Monro search processes. Here, we propose the following procedure for practical implementation of the grid bootstrap for constructing unbiased prediction estimators.

Step 1. Estimate the AR model by OLS and construct a grid of J values for the largest AR root on an interval around the OLS estimate, i.e. α_j for $j = 1, 2, \ldots, J$.

Step 2. Calculate the recentered restricted residuals at each grid point.

Step 3. Use the restricted residuals from *Step 2* and the corresponding value of α_j to simulate conditional bootstrap samples.

Step 4. Using the bootstrap data, re-estimate the AR model and re-compute the quantity $(\alpha_j^*)^k y_T$ B times, and then calculate the θth prediction quantile by taking the $[(B+1)\theta]$th element of the sorted forecasts.

Step 5. Repeat the *Steps* 2–4 at each point on the grid and connect the points for the θth quantile. This gives us an approximation of the θth quantile function of the predictive distribution.

Step 6. The intersection of the OLS forecast function and the θth (or 0.5th) quantile function computed by linear interpolation produces our θth quantile (or median) unbiased forecasts.

The asymptotic and the bootstrap methods suggested above can be extended to more realistic cases such as nonzero-mean and higher-order AR processes. For details, see Gospodinov (2001).

4. Small-sample experiment

4.1. Experimental design and description of methods

To assess the finite sample performance of the median unbiased forecasts obtained via grid bootstrap, we conducted a small Monte Carlo study. In the simulation experiment, the data are generated from an AR(1) zero-mean Gaussian process with $T = 100$, four values of the AR parameter, $\alpha = 0.9$, 0.95, 0.98 and 0.99, forecast horizon of $k = 10$ and 5,000 Monte Carlo repetitions. The data are generated from the conditional density of the process backward in time using a rejection–acceptance algorithm described in Kabaila (1993) and Breidt et al. (1995).

To evaluate the properties of the bootstrap median unbiased forecasts, we used the OLS and random walk forecasts as benchmarks. The OLS method as well as all bootstrap methods estimate the AR parameter from a regression with an intercept but in the forecasting stage the mean of the series is set equal to its true value of 0. The random walk method involves no estimation and imposes a unit root and a zero mean on the dynamics of the process. Thus, the k-period ahead forecast is given by the last observation of the series.

The bootstrap-based procedures for obtaining median unbiased forecast have the following characteristics. The bootstrap samples are generated backward in time conditional on the last value of the original series. The conventional bootstrap estimator (C-BOOT) estimates the median of the bootstrap distribution of $\hat{\alpha}^k y_T$, and then corrects for the bias by computing $\hat{\alpha}^k y_T - \{Med[(\alpha^*)^k y_T] - \hat{\alpha}^k y_T\}$. For the grid

bootstrap (G-BOOT), we follow the steps described in Section 3.2 with $\alpha_j \in [\hat{\alpha} - s.e.(\hat{\alpha}), \max(1, \hat{\alpha} + 2s.e.(\hat{\alpha}))]$ for $j = 1, 2, \ldots, 6$ and $B = 399$ and calculate the median forecast $Med[(\alpha_i^*)^k y_T]$ at each of these points. Then, we invert the bootstrap median function to obtain the median unbiased forecast.

In the simulations we also evaluate the performance of the asymptotic expression of the conditional predictive distribution derived in Lemma 1 and the Bayesian predictor. The limiting distribution is simulated for $T = 1,000$ and $c = 0, -1, \ldots, -21$ using 20,000 replications. The asymptotic median unbiased forecast is constructed by inverting the 0.5th quantile of the simulated asymptotic representation. The computation of the Bayes predictor involves the choice of a prior distribution for (α, σ^2). Since our approach of focusing on the near-nonstationarity region implies a strong prior knowledge about the properties of the process, we consider an informative normal-inverted gamma prior on the parameters, $(\alpha, \sigma^2) \propto \mathrm{N}(\alpha|\alpha_0, B_0^{-1}) IG(\sigma^2|\nu_0/2, \eta_0/2)$ with $\alpha_0 = 0.95$, $B_0^{-1} = (0.1)^2$ and $\nu_0 = \eta_0 = 0$. The corresponding posteriors $p(\sigma^2 \mid Y)$ and $p(\alpha \mid \sigma^2, Y)$ are given in Bauwens et al. (1999) and Albert and Chib (1993). We follow Albert and Chib (1993) and use the Gibbs sampler to obtain the posteriors and the predictive distribution of y_{T+k}. Some other choices of an informative prior for the slope parameter such as the beta prior are also possible.[6]

4.2. Simulation results

In the Monte Carlo simulations for the median unbiased forecasts, we compute the probability that the forecast is below the true conditional forecast mean (probability of underprediction), $\Pr\{\hat{y}_{T+k|T} < \bar{y}_{T+k|T}\}$, and the mean absolute deviation (MAD) of the forecast from the true conditional mean, $\frac{1}{1000}\sum_{i=1}^{1000} |\hat{y}_{T+k|T,i} - \bar{y}_{T+k|T,i}|$. Since the forecast accuracy is measured in terms of deviations from the true conditional mean, it tends to overstate the magnitude of the relative forecast improvement reported in this section.

The results for the AR(1) process with the last observation set equal to 0.5, 1 and 3 are presented in Table 1. The OLS forecast underpredicts significantly the true conditional mean for all parameter and terminal values considered. As expected, imposing a unit root on the series removes the parameter variability and reduces the forecast errors for large values of the parameter but it always overpredicts the true conditional mean. The Bayesian forecasts are better centered than the OLS forecasts for $\alpha \leqslant 0.95$ but their properties deteriorate as the AR parameter approaches unity. By contrast, the asymptotic representation appears to provide a good approximation when the AR root is close to the unit circle boundary although it tends to overpredict as we move towards the interior of the stationarity region.

The results for the conventional bootstrap method clearly show that as the AR parameter approaches one, the number of forecasts that are below the true conditional

[6] We also experimented with a flat prior on the AR parameter and simulated the predictive density as in Thompson and Miller (1986). Since the posterior for the AR parameter under a flat prior is centered at $\hat{\alpha}$, it inherits the downward bias of the OLS estimator and the results are very similar to the OLS forecasts and thus not reported.

Table 1
10-step ahead forecasts from AR(1) model

	$\alpha = 0.9$		$\alpha = 0.95$		$\alpha = 0.98$		$\alpha = 0.99$	
Method	Prob	MAD	Prob	MAD	Prob	MAD	Prob	MAD
$y_T = 0.5$								
OLS	0.8038	0.0745	0.9052	0.1208	0.9636	0.1624	0.9670	0.1803
RW	0.0000	0.3257	0.0000	0.2006	0.0000	0.0915	0.0000	0.0478
BAYES	0.6440	0.0493	0.8352	0.0851	0.9576	0.1215	0.9962	0.1398
ASYMPTOTIC	0.3820	0.1124	0.4748	0.1074	0.4736	0.0995	0.4988	0.0912
C-BOOT	0.6028	0.0754	0.6416	0.1025	0.6400	0.1198	0.6594	0.1299
G-BOOT	0.5030	0.0865	0.4990	0.1087	0.4802	0.1007	0.5148	0.0925
$y_T = 1$								
OLS	0.7916	0.1479	0.8890	0.2373	0.9634	0.3287	0.9596	0.3590
RW	0.0000	0.6513	0.0000	0.4013	0.0000	0.1829	0.0000	0.0956
BAYES	0.6452	0.0975	0.8356	0.1695	0.9584	0.2475	0.9940	0.2694
ASYMPTOTIC	0.3730	0.2287	0.4732	0.2145	0.5012	0.2010	0.5092	0.1837
C-BOOT	0.5944	0.1537	0.6364	0.2067	0.6664	0.2477	0.6554	0.2671
G-BOOT	0.5042	0.1773	0.4994	0.2195	0.5146	0.2085	0.5206	0.1898
$y_T = 3$								
OLS	0.6902	0.3926	0.8252	0.6258	0.9432	0.9193	0.9562	1.0412
RW	0.0000	1.9539	0.0000	1.2038	0.0000	0.5488	0.0000	0.2869
BAYES	0.4894	0.2741	0.7442	0.3947	0.9424	0.6233	0.9954	0.7477
ASYMPTOTIC	0.3106	0.6822	0.3868	0.6788	0.4468	0.5933	0.4802	0.5286
C-BOOT	0.5592	0.4615	0.6106	0.6363	0.6750	0.7734	0.6848	0.8448
G-BOOT	0.4988	0.5582	0.4974	0.7305	0.4812	0.6887	0.5074	0.6246

Note: Prob $= \Pr\{\hat{y}_{T+k} \leqslant \bar{y}_{T+k}\}$; MAD=mean absolute deviation of forecast from conditional mean; RW=random walk; BAYES=Bayesian predictor with a normal-inverted gamma prior; ASYMPTOTIC=local-to-unity asymptotics; C-BOOT=conventional bootstrap; G-BOOT=grid bootstrap.

mean increases and they underpredict 66–69% of the time for $\alpha = 0.99$. As we pointed out above, this is caused by the inconsistent estimates of the local-to-unity parameter and hence the inconsistency of the conventional bootstrap estimator. The numerical results also reveal that when the AR parameter is close to unity, the bias-corrected bootstrap significantly improves the accuracy of the point forecasts over the OLS method.

The small-sample performance of the grid bootstrap methods supports the theoretical discussion in Section 3. The grid bootstrap method produces forecasts that are very close to the nominal level of 0.5 over all parameter and terminal values under consideration. The grid bootstrap forecast errors are lower than the conventional bootstrap methods when the AR parameter is near the unit circle but larger for $\alpha = 0.9$ and 0.95.[7]

[7] Numerical results for forecasts with estimated mean, higher-order AR models and prediction intervals can be found in Gospodinov (2001).

5. Application: forecasting interest rates

In this empirical example, we study the prediction accuracy of the median unbiased forecasts for 1-month U.S. T-bill yields. The interest rates are highly persistent processes and typically the unit root hypothesis can not be rejected. But the presence of an exact unit root is inconsistent with the finance theory of bond pricing and its imposition cannot be justified on theoretical grounds. On the other hand, the OLS forecasts suffer from a strong bias towards zero and systematically underpredict the absolute value of the true conditional mean. This makes our bootstrap method a natural candidate for obtaining correctly centered and impartial forecasts of interest rates. In this application, we use data for 1-month spot rate r_t and k-month forward rate $f_{k,t}$ for $k = 3, 6, 12$ and 18 from the McCulloch-Kwon data set for the period January 1952–February 1991.

The forecasting exercise compares the performance of the grid bootstrap forecasts to the OLS forecasts and is conducted as follows. We estimate a nonzero-mean AR(1) model for the spot rate on a rolling window of 120 observations and compute the k-period ahead forecast $\hat{r}_{T+k|T}$. This gives us 349 forecasts that are used to construct the measures for prediction evaluation. The rolling window estimation is expected to be more robust to a potential structural shift that may have occurred in the 1979–82 period due to the change of the Fed operating procedures.

In addition to the conventional statistical measures for forecast accuracy such as the mean absolute error, we evaluate the forecasting performance of the different methods using a profit-based rule (see Leitch and Tanner, 1991). For the profit-based experiment, we assume that the investor can sell and buy forward contracts at the specified forward rate and take a position every period. Since the transaction costs do not affect the relative performance of the different methods, they are assumed to be zero. The trading strategy is the following: if the k-step ahead forecast of the spot rate is above the k-month forward rate, the investor sells a \$1 k-month forward contract and if the forecast is below the forward rate, the investor buys a \$1 k-month forward contract. The profit is $|\exp(-r_{T+k}) - \exp(-f_{k,T})|$, if $(\hat{r}_{T+k|T} - f_{k,T})/(r_{T+k} - f_{k,T}) \geqslant 0$, and $-|\exp(-r_{T+k}) - \exp(-f_{k,T})|$ otherwise.

Table 2 reports the mean absolute errors, the profits from the profit-based rule and the average directional accuracy of four grid bootstrap forecasts relative to the OLS forecast. The first three grid bootstrap forecasts are based on the median, the 0.25th and the 0.75th quantiles of the distribution of $T^{-1/2}(\hat{r}_{T+k|T} - \bar{r}_{T+k|T})$ conditional on r_T and the last one is their combined forecast calculated as a simple average of the three quantile predictions.

Several interesting observations are in order. First, the forecasts obtained by the grid bootstrap rarely produce a lower mean absolute error than the OLS forecast. In terms of profits and directional accuracy, however, the grid bootstrap forecasts outperform the OLS methods for almost all cases. The gains in terms of profits for the grid bootstrap are nontrivial and for longer forecasting horizons exceed 25%. Compared to the other quantile forecasts, the median unbiased forecast is dominated by the 0.75th quantile forecast which is the point forecast that is performing best in terms of all three criteria. Also, the combination forecast that summarizes the information in the different quantiles of the predictive distribution appears to produce significant forecasting improvements.

Table 2
k-step ahead forecast comparison for 1-month T-bill spot rates

Method	MAE	PBR	ADA	MAE	PBR	ADA
	$k=3$			$k=6$		
G-BOOT (*Med*)	0.9735	1.0831	1.0314	0.9724	1.0787	1.0294
G-BOOT (0.25Q)	1.0618	1.0186	1.0078	1.0657	0.9793	0.9958
G-BOOT (0.75Q)	0.9487	1.0956	1.0431	0.9185	1.1060	1.0798
G-BOOT (*COMB*)	0.9778	1.0431	1.0235	0.9749	1.0551	1.0168
	$k=12$			$k=18$		
G-BOOT (*Med*)	1.0036	1.1168	1.0290	1.0263	1.1558	1.0272
G-BOOT (0.25Q)	1.0537	1.0688	1.0100	1.0181	1.1255	1.0326
G-BOOT (0.75Q)	0.9594	1.2766	1.0746	0.9947	1.2688	1.0707
G-BOOT (*COMB*)	0.9717	1.2443	1.0647	0.9906	1.2872	1.0652

Note: All entries are relative to the OLS forecast. PBR=profit-based rule, ADA=average directional accuracy, *Med*=median, x$Q=x$th quantile, *COMB*=simple average of the 0.25th, 0.5th and 0.75th quantiles (combination forecast) and MAE=$\frac{1}{349}\sum_{i=1}^{349}|\hat{r}_{i,T+k|T}-r_{i,T+k}|$ is the mean absolute error of the forecasts.

6. Conclusion

The median unbiasedness of conditional forecast estimators proves to be a very desirable property in many practical situations. This paper develops a bootstrap-based procedure for computing median unbiased forecasts for near-integrated AR processes. The OLS and the conventional bootstrap methods underpredict the value of the true conditional mean way too often and are not appropriate when impartiality, captured by the loss function, is a crucial feature of the decision-making process. The paper derives the suitably normalized limiting distribution of the deviation of the point forecast from the true conditional mean. The dependence of the asymptotic representation of the predictive distribution on nuisance parameters precludes the use of the standard asymptotic and bootstrap methods for bias correction. To overcome this problem, we propose a bootstrap method that approximates the median function of the predictive distribution on a grid of points of the unknown parameter. We show that the method could be further extended to different quantile point forecasts, prediction intervals, nonzero-mean and higher-order AR processes. Finally, the potential usefulness of the method in applied work is illustrated in the context of forecasting the 1-month T-bill rate.

Acknowledgements

This paper is based on Chapter 2 of my Ph.D. thesis in the Department of Economics at Boston College. I am deeply indebted to Bruce Hansen and Serena Ng for their insightful advice and numerous stimulating discussions. I would like to thank Jushan Bai, Pierluigi Balduzzi, Lutz Kilian, Arthur Lewbel, Arnold Zellner (the editor) and two anonymous referees for helpful comments and suggestions.

Appendix A. Mathematical proofs

A.1. Proof of Lemma 1

Appealing to the functional central limit theorem yields (Phillips, 1987)

$$(\hat{c}-c) \Rightarrow \left[\int_0^1 J_c(r)\,\mathrm{d}W(r)\right]\left[\int_0^1 J_c(r)^2\,\mathrm{d}r\right]^{-1}, \tag{7}$$

$$T^{-1/2}y_T \Rightarrow J_c(1), \tag{8}$$

where $J_c(r)=\exp(cr)J_c(0)+\exp(cr)\int_0^r \exp(-cs)\,\mathrm{d}W(s)$ is the Ornstein–Uhlenbeck process. Using the relationship $\int_0^1 J_c(r)\,\mathrm{d}J_c=\int_0^1 J_c(r)\,\mathrm{d}W+c\int_0^1 J_c(r)^2\,\mathrm{d}r$ and applying Ito's lemma to $\int_0^1 J_c(r)\,\mathrm{d}J_c$ gives that $\int_0^1 J_c(r)\,\mathrm{d}W=\frac{1}{2}(J_c(1)^2-J_c(0)^2-1)-c\int_0^1 J_c(r)^2\,\mathrm{d}r$.

If k is a fraction of the sample size, $k=[\lambda T]$, then

$$\alpha^k=\left(1+\frac{c}{T}\right)^k=\left(1+\frac{c}{T}\right)^{[\lambda T]}\to\exp(c\lambda).$$

Therefore, $\hat{\alpha}^k-\alpha^k=\alpha^k(1+T(\hat{\alpha}-\alpha)/T\alpha)^k-\alpha^k\to\exp(\lambda c)[\exp(\lambda(\hat{c}-c))-1]$ or, equivalently, $\hat{\alpha}^k-\alpha^k\to\exp(\lambda\hat{c})-\exp(\lambda c)=\exp(\lambda c)[\exp(\lambda(\hat{c}-c))-1]$.

Thus, combining these expressions, we have

$$\begin{aligned}(T^{-1/2}(\hat{y}_{T+k}-\bar{y}_{T+k})|T^{-1/2}y_T=x)&=((\hat{\alpha}^k-\alpha^k)[T^{-1/2}y_T]|T^{-1/2}y_T=x)\\ &\Rightarrow\exp(\lambda c)\left\{\exp\left[\frac{\lambda}{2}\frac{x^2-J_c(0)^2-1-2c\int_0^1 J_{c|x}(r)^2\,\mathrm{d}r}{\int_0^1 J_{c|x}(r)^2\,\mathrm{d}r}\right]-1\right\}x\end{aligned} \tag{9}$$

as desired.

The expression for the conditional forecast error of a zero-mean AR(1) process is given by

$$\hat{e}_{T+k|T}=(\hat{\alpha}^k-\alpha^k)y_T-\sum_{i=0}^{k-1}\alpha^i\varepsilon_{T+k-i},$$

where ε_{T+k-i} is independent of y_T.

The asymptotic behavior of the first term is shown in (9). The second term in the expression can be rewritten as (see Stock, 1996)

$$T^{-1/2}\sum_{i=0}^{k-1}\alpha^i\varepsilon_{T+k-i}=T^{-1/2}\sum_{i=1}^{T+k}\alpha^{T+k-i}\varepsilon_i-\alpha^kT^{-1/2}\sum_{i=1}^{T}\alpha^{T-i}\varepsilon_i.$$

Then, using $k=[\lambda T]$ and applying the functional central limit theorem yields

$$T^{-1/2}\sum_{i=1}^{[T(1+\lambda)]}\alpha^{[T(1+\lambda)]-i}\varepsilon_i \Rightarrow J_c(1+\lambda), T^{-1/2}\sum_{i=1}^{T}\alpha^{T-i}\varepsilon_i \Rightarrow J_c(1) \text{ and } \alpha^k \to \exp(c\lambda).$$

Also, since $J_c(r)=\exp(cr)J_c(0)+\int_0^r \exp[c(r-s)]\,\mathrm{d}W(s)$,

$$J_c(1+\lambda)-\exp(c\lambda)J_c(1)=\int_1^{1+\lambda}\exp[c(1+\lambda-s)]\,\mathrm{d}W(s).$$

Finally, by applying the FCLT and using the results that $\mathrm{E}(\int_{r_0}^{r_1} G\,\mathrm{d}W(s))=0$ and $Var(\int_{r_0}^{r_1} G\,\mathrm{d}W(s))=\int_{r_0}^{r_1}\mathrm{E}(G)^2\,\mathrm{d}s$, it follows that

$$T^{-1/2}\sum_{i=0}^{k-1}\alpha^i\varepsilon_{T+k-i}\sim \mathrm{N}\left(0,\frac{\exp(2c\lambda)-1}{2c}\right). \qquad \square$$

A.2. Proof of Theorem 1

Suppose for simplicity that $\bar{\Delta}^c_{t,t-1}=0$ and $\bar{\Delta}^c_{t,t+1}=0$. Then, the relationship between the backward and forward residuals in the near-integrated AR(1) model with $\alpha=1+c/T$ is given by

$$\bar{\xi}_t=-\alpha\bar{\varepsilon}_{t+1}+(1-\alpha^2)\sum_{i=1}^{t}\alpha^{t-i}\bar{\varepsilon}_i$$

$$=-\bar{\varepsilon}_{t+1}-c\frac{\bar{\varepsilon}_{t+1}}{T}+\frac{c}{\sqrt{T}}\left(\frac{c}{T}+2\right)\frac{1}{\sqrt{T}}\sum_{i=1}^{t}\left(1+\frac{c}{T}\right)^{t-i}\bar{\varepsilon}_i.$$

Since

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{[rT]}\left(1+\frac{c}{T}\right)^{[rT]-i}\bar{\varepsilon}_i \Rightarrow J_c(r)=\mathrm{O_p}(1)$$

it follows immediately from the above expression that $d_2(F_{\bar{\xi}^*},F_{-\bar{\varepsilon}^*})^2\to 0$ as $T\to\infty$, where $\{\bar{\varepsilon}_t^*\}$ are the bootstrap forward residuals. Then,

$$d_2(F_{\bar{\xi}^*},F_{-\varepsilon})^2 = inf(\mathrm{E}|\bar{\xi}^*-(-\varepsilon)|^2)\leqslant \mathrm{E}|\bar{\xi}^*-(-\varepsilon)|^2$$

$$\leqslant \mathrm{E}|\bar{\xi}^*-(-\bar{\varepsilon}^*)|^2+\mathrm{E}|\varepsilon-\bar{\varepsilon}^*|^2\to 0 \text{ as } T\to\infty.$$

since $\mathrm{E}|\varepsilon-\bar{\varepsilon}^*|^2\to 0$ (Heimann and Kreiss, 1996). The same result holds true for non-zero means of the forward and backward restricted residuals. $\square$

A.3. Proof of Theorem 2

Consider model (1) with $\alpha = (1 + c/T)$. Let the bootstrap processes $\{y_t^*\}_{t=1}^T$ and $\{y_t^{**}\}_{t=1}^T$ be generated by

$$y_t^* = (1 + c/T)y_{t-1}^* + \bar{\varepsilon}_t^* \tag{10}$$

and

$$y_t^{**} = (1 + c/T)y_{t+1}^{**} + \bar{\xi}_t^*, \qquad t = T, T-1, \ldots, 1 \text{ and } y_T^{**} = y_T, \tag{11}$$

where $\bar{\varepsilon}_t = y_t - (1 + c/T)y_{t-1} - (T-1)^{-1}\sum_{t=1}^{T-1}[y_t - (1 + c/T)y_{t-1}]$, $\bar{\xi}_t = y_t - (1 + c/T)y_{t+1} - (T-1)^{-1}\sum_{t=1}^{T-1}[y_t - (1 + c/T)y_{t+1}]$ and $\{\bar{\varepsilon}_t^*\}_{t=1}^T$ and $\{\bar{\xi}_t^*\}_{t=1}^T$ are random samples from the empirical distribution functions of $\bar{\varepsilon}_t$ and $\bar{\xi}_t$, respectively.

Recursions (10) and (11) are used to construct the processes

$$S_T^*(r) = \exp\left(c\frac{[rT]}{T}\right)\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]}\exp\left(-c\frac{j}{T}\right)\bar{\varepsilon}_j^* + \mathrm{o_p}(1), \quad 0 \leqslant r \leqslant 1,$$

$$S_T^{**}(r) = \exp\left(-c\frac{[rT]}{T}\right)\frac{1}{\sqrt{T}}\sum_{j=1}^{[rT]}\exp\left(c\frac{j}{T}\right)\left(-\bar{\xi}_j^*\right) + o_{\mathrm{p}}(1), \quad 0 \leqslant r \leqslant 1.$$

From the results in Theorem 1, $S_T^*(r) - S_T^{**}(r) = \mathrm{o_p}(1)$. Then, following the arguments in Theorem 2.2 in Basawa et al. (1991b) and Theorem 2 in Hansen (1999), we get that $S_T^*(r) \Rightarrow J_c(r)$ for $r \in [0,1]$ and

$$T(\alpha^* - \alpha) \Rightarrow \frac{1}{2}\,\frac{J_c(1)^2 - J_c(0)^2 - 2c\int_0^1 J_c(r)^2\,\mathrm{d}r - 1}{\int_0^1 J_c(r)^2\,\mathrm{d}r} \tag{12}$$

which is the same as the limiting representation of $T(\hat{\alpha} - \alpha)$, where α^* and $\hat{\alpha}$ denote the bootstrap and the OLS estimators. Consider now the bootstrap forecasts $y_{T+k|T}^* = (\alpha^*)^k y_T$ and parameterize the forecast horizon as $k = [\lambda T]$. Then, applying the arguments from Lemma 1 delivers the desired consistency result. □

References

Albert, J.H., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business and Economic Statistics 11, 1–15.

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

Andrews, D.W.K., Buchinsky, M., 2000. A three-step method for choosing the number of bootstrap repetitions. Econometrica 68, 23–51.

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991a. Bootstrapping unstable first-order autoregressive processes. Annals of Statistics 19, 1098–1101.

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991b. Bootstrap test of significance and sequential bootstrap estimation for unstable first-order autoregressive processes. Communications in Statistics-Theory and Methods 20, 1015–1026.

Bauwens, L., Lubrano, M., Richard, J.-F., 1999. Bayesian Inference in Dynamic Econometric Models. Oxford University Press, Oxford.

Beran, R., Ducharme, G.R., 1991. Asymptotic Theory for Bootstrap Methods in Statistics, Les Publications CRM, Centre de recherches mathematiques. Universite de Montreal, Montreal.

Bickel, P.J., Freedman, D., 1981. Some asymptotic theory for the bootstrap. Annals of Statistics 9, 1196–1217.

Breidt, F.J., Davis, R.A., Dunsmuir, W.T.M., 1992. On backcasting in linear time series models. In: Brillinger, D., Caines, P., Geweke, J., Parzen, E., Rosenblatt, M., Taqqu, M. (Eds.), New Directions in Time Series Analysis, Part I. Springer, New York, pp. 25–40.

Breidt, F.J., Davis, R.A., Dunsmuir, W.T.M., 1995. Improved bootstrap prediction intervals for autoregressions. Journal of Time Series Analysis 16, 177–200.

Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.

Davidson, R., MacKinnon, J.G., 2000. Bootstrap tests: how many bootstraps? Econometric Reviews 19, 55–68.

Gospodinov, N., 2001, Median unbiased forecasts for highly persistent autoregressive processes. Working paper, Department of Economics, Concordia University.

Granger, C.W.J., 1969. Prediction with a generalized cost of error function. Operational Research Quarterly 20, 199–207.

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

Heimann, G., Kreiss, J.-P., 1996. Bootstrapping general first order autoregression. Statistics and Probability Letters 30, 87–98.

Horowitz, J.L., 2001. The bootstrap. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, Vol. 5. North-Holland, Amsterdam.

Kabaila, P., 1993. On bootstrap predictive inference for autoregressive processes. Journal of Time Series Analysis 14, 473–484.

Kemp, G.C.R., 1999. The behavior of forecast errors from a nearly integrated AR(1) model as both sample size and forecast horizon become large. Econometric Theory 15, 238–256.

Leitch, G., Tanner, J.E., 1991. Economic forecast evaluation: profit versus the conventional error measures American Economic Review 81, 580–590.

Phillips, P.C.B., 1979. The sampling distribution of forecasts from a first-order autoregression. Journal of Econometrics 9, 241–261.

Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.

Phillips, P.C.B., 1991. To criticize the critics: an objective Bayesian analysis of stochastic trends Journal of Applied Econometrics 6, 333–364.

Phillips, P.C.B., 1998. Impulse response and forecast error variance asymptotics in nonstationary VARs. Journal of Econometrics 83, 21–56.

Sampson, M., 1991. The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. Journal of Applied Econometrics 6, 67–76.

Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. Journal of Monetary Economics 28, 435–459.

Stock, J.H., 1996. VAR, error-correction and pretest forecasts at long horizons. Oxford Bulletin of Economics and Statistics 58, 685–701.

Stock, J.H., 1997. Cointegration, long-run comovements and long horizon forecasting. In: Kreps, D., Willis, K.F. (Eds.), Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society. Cambridge University Press, Cambridge, pp. 34–60.

Thombs, L.A., Schucany, W.R., 1990. Bootstrap prediction intervals for autoregression. Journal of American Statistical Association 85, 486–492.

Thompson, P.A., Miller, R.B., 1986. Sampling the future: a Bayesian approach to forecasting from univariate time series models Journal of Business and Economic Statistics 4, 427–436.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.