The Grid Bootstrap and the Autoregressive Model

Bruce E. Hansen*
University of Wisconsin†

www.ssc.wisc.edu/~bhansen

August 1998
Revised: May 1999

Abstract

A "grid" bootstrap method is proposed for confidence interval construction, which has improved performance over conventional bootstrap methods when the sampling distribution depends upon the parameter of interest. The basic idea is to calculate the bootstrap distribution over a grid of values of the parameter of interest, and form the confidence interval by the no-rejection principle. Our primary motivation is given by autoregressive models, where it is known that conventional bootstrap methods fail to provide correct first-order asymptotic coverage when an autoregressive root is close to unity. In contrast, the grid bootstrap is first-order correct globally in the parameter space. Simulation results verify these insights, suggesting that the grid bootstrap provides an important improvement over conventional methods. Gauss code which calculates the grid bootstrap intervals, and replicates the empirical work reported in this paper, is available from the author's webpage.

*This research was supported by a grant from the National Science Foundation and a Sloan Foundation Research Fellowship. I thank Nikolay Gospodinov for excellent research assistance, Yuichi Kitamura and Joel Horowitz for helpful conversations and insights, and Frank Diebold and two referees for constructive comments on a previous draft.

†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.

1 Introduction

For motivation, consider the autoregressive (AR) model of order 1 with trend, which may be written as

$$Y_t = \mu_0 + \mu_1 t + y_t \tag{1}$$

$$y_t = \alpha y_{t-1} + e_t, \tag{2}$$

$t = 1, ..., n$, with and e_t independent and identically distributed (iid) with unknown distribution function $P(\cdot)$, $Ee_t = 0$, and $Ee_t^2 < \infty$. Correct inference in the AR model (1)-(2) is the first step towards correct inference in more complicated time series models. The standard method to estimate the model is by ordinary least squares (OLS) on the single equation

$$Y_t = \mu_0' + \mu_1' t + \alpha Y_{t-1} + e_t.$$

Let $\hat{\alpha}$ and $s(\hat{\alpha})$ denote the OLS estimate of α and its standard error.

Our goal is to construct an $\beta\%$ confidence interval for the AR parameter α. The conventional asymptotic interval is based on the asymptotic $N(0,1)$ approximation to the t-statistic

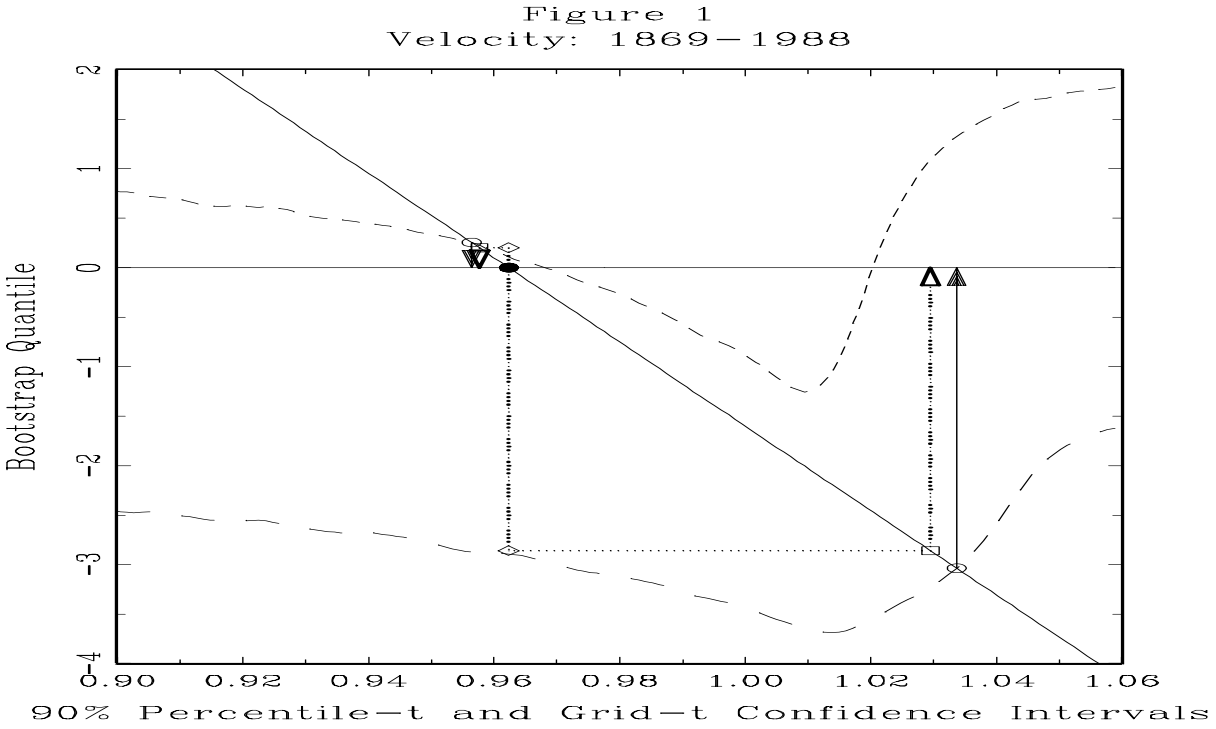
$$t(\alpha) = \frac{\hat{\alpha} - \alpha}{s(\hat{\alpha})}, \tag{3}$$

which is valid in the AR(1) model when $|\alpha| < 1$. Unfortunately, the normal approximation is quite poor in practice[1], especially when $|\alpha|$ is large. To see the extent of this distortion, the dashed lines in Figure 1 plot the 5% and 95% quantile functions of the bootstrap distribution of $t(\alpha)$ for the velocity series for 1869-1988 from Schotman and van Dijk (1991), an extension of the Nelson-Plosser (1982) dataset. These functions mark the quantiles of the sampling distribution of the t-statistic $t(\alpha)$ assuming that the data was generated from (2) as α is varied over the range $[.90, 1.06]$. The precise manner in which these functions are calculated will be discussed in section 3.1. What is important for our purposes is that the normal approximation suggests that these quantile functions should be constant at -1.645 and 1.645, respectively, and it is striking how different the functions are from these values.

When conventional asymptotic approximations are poor, many researchers turn to the bootstrap. The bootstrap replaces the asymptotic sampling distribution by an exact distribution which acts as if the empirical distribution of the sample is the population distribution. In particular, the parametric percentile-t bootstrap constructs a confidence interval for a parameter α by evaluating the sampling distribution of the t-statistic $t(\alpha)$ assuming that the

[1]This is well known. For early discussions see Fuller (1996, p. 411) and Phillips (1977).

data was generated from (2) with the true value of α equalling the OLS estimate $\hat{\alpha}$. We can read the percentile-t bootstrap interval for α from the information provided in Figure 1, assisted by the dotted lines. From the OLS estimate $\hat{\alpha} = .962$ the dotted lines move vertically to the 95% and 5% bootstrap quantile functions, the intersections marked by the open diamonds. From these points the dotted lines move horizontally to the t-statistic function $t(\alpha)$. The points of intersection, marked by open rectangles, are the percentile-t bootstrap endpoints. They are projected onto the x-axis and marked by the white arrowheads. This 90% percentile-t bootstrap interval is $[.958, 1.030]$.



The percentile-t bootstrap makes the implicit approximation that the bootstrap quantile functions (such as those displayed in Figure 1) are constant functions, at least over the relevant range of the potential confidence interval, which is false in the AR model. This non-constancy persists in large sample if we cast the leading coefficient as local-to-unity. Setting $\alpha = 1 + c/n$ and holding c fixed as $n \to \infty$, then the t-statistic for α has the asymptotic

distribution

$$t(\alpha) \Rightarrow \frac{\int_0^1 W_c dW}{\left(\int_0^1 W_c^2\right)^{1/2}} \tag{4}$$

where W_c is a detrended diffusion process. The asymptotic distribution of (4) depends on α through the reparameterization c and hence the t-statistic is non-pivotal. As c is not consistently estimated by least squares, it follows that the percentile-t interval has incorrect first-order asymptotic coverage. That is, in the local-to-unity framework, the interval does not properly control Type I error. This difficulty was pointed out by Basawa, et. al. (1991). This means that even in large samples, the coverage probability of the percentile-t interval is quite poor for true values of α near the unit root, and the Type I error is not controlled globally in the parameter space. It follows that the percentile-t interval (even asymptotically) has the incorrect size[2] over the stationary region $\alpha \in (-1, 1)$.

As the conventional bootstrap fails to provide an asymptotically correct confidence interval, it would appear that there is no hope to construct one with proper coverage. Fortunately, this is not the case. Indeed, we can easily read what I call the grid bootstrap confidence interval from Figure 1, and this interval has excellent coverage properties. The endpoints are found from the intersections of the quantile functions with the t-statistic line, and are marked by the open circles. These endpoints are projected onto the x-axis and marked by the solid arrowheads. This 90% interval is $[.956, 1.034]$. The left endpoint is almost identical to that from the percentile-t bootstrap, but the right endpoint is somewhat higher.

Unlike the percentile-t interval, the grid bootstrap interval is asymptotically justified even in the local-to-unity setting. We are able to show that the grid bootstrap confidence interval has first-order correct asymptotic coverage for both stationary and local-to-unity autoregressive models. Thus the grid bootstrap asymptotically controls type I error globally in the parameter space.

Our methods are a natural extension of several earlier papers on confidence interval construction in autoregressive models. Stock (1991) showed how to construct asymptotically valid confidence intervals for the largest autoregressive root in local-to-unity AR models. Andrews (1993) showed how to construct exact confidence intervals in the AR(1) model with Gaussian errors[3]. Andrews and Chen (1994) suggested an approximation to extend this method to higher-order AR models. Nakervis and Savin (1996) used bootstrap methods

[2]Size is the maximum probability of Type I error over the parameter space.

[3]If we impose use the assumption that the errors are iid Gaussian, our grid-α interval corresponds to Andrews' interval.

for hypothesis testing in the AR(1) model. While our paper borrows ideas and methods from these earlier papers, ours is the first to show how to construct correct bootstrap confidence intervals in autoregressive models.

The grid bootstrap intervals presented in this paper are similar to the test inversion bootstrap (TIB) intervals and studentized test inversion bootstrap (STIB) intervals of Carpenter (1999). Similar intervals appear in DiCiccio and Romano (1990), Garthwaite and Buckland (1992) and Kabaila (1993), although there are differences between these papers regarding the choice of studentization and the estimation of nuisance parameters. Carpenter (1999) presents a number of applications of this method, including the AR(1) model, but does not provide a theory or justification relevant to the local to unity model.

An alternative to the methods presented here might be the subsampling confidence intervals of Romano and Wolf (1998), who discuss confidence interval construction for the AR(1) model, allowing for a possible unit or explosive root, using reestimates based on subsamples. This method is quite distinct from those presented here.

The organization of the paper is as follows. Section 2 describes the sampling framework, introduces the grid bootstrap, and gives conditions under which the grid bootstrap provides first-order asymptotic coverage, and discusses computation. Section 3 investigates the autoregressive model. Formal proofs of bootstrap consistency are provided for both stationary and near-integrated cases. Section 4 investigates the bootstrap methods using Monte Carlo methods. Both AR(1) and AR(2) models are examined. Section 5 provides several applications to the Nelson-Plosser dataset. The velocity and real per capita GNP series are carefully explored as examples of the AR(1) and AR(2) models, respectively. Grid-t bootstrap intervals are provided for the leading AR coefficient for all 13 series. Proofs of the theorems are presented in the appendix.

2 The Grid Bootstrap

2.1 Framework

A sample X_n is generated from a distribution $G_n(x \mid \alpha, \eta) = P(X_n \leq x \mid \alpha, \eta)$ which depends on a parameter of interest $\alpha \in R$ and a nuisance parameter $\eta \in \Xi$, where n denotes sample size. The nuisance parameter space Ξ may be infinite dimensional (i.e., the space of error distributions), and is endowed with a metric $d(\eta, \eta')$. Let $\hat{\alpha}$ denote an estimate of α and $s(\hat{\alpha})$ its standard error. We assume that for each α there is some estimator $\hat{\eta}(\alpha) \in \Xi$ of the nuisance parameter η, which may (but need not) be a function of α.

Let $S_n(\alpha)$ be a non-degenerate test statistic of the hypothesis $H_0 : \alpha_0 = \alpha$ which is a monotonic function of α. For example, two obvious choices include the non-studentized estimate $b(\alpha) = \hat{\alpha} - \alpha$ and the t-statistic $t(\alpha) = (\hat{\alpha} - \alpha)/s(\hat{\alpha})$. The statistic $S_n(\alpha)$ has a sampling distribution which possibly depends on (α, η). Let

$$F_n(x \mid \alpha, \eta) = P(S_n(\alpha) \leq x \mid \alpha, \eta)$$

denote its distribution function. For fixed (α, η), let $q_n(\theta \mid \alpha, \eta)$ be the inverse of F_n (making the approximation that F_n is continuous in x). This function satisfies

$$F_n(q_n(\theta \mid \alpha, \eta) \mid \alpha, \eta) = \theta.$$

Since $q_n(\theta \mid \alpha, \eta)$ is the θ quantile of the distribution of $S_n(\alpha)$, we will refer to $q_n(\theta \mid \alpha, \eta)$ as the quantile function of the distribution. Note that for fixed (α, η) the quantile function $q_n(\theta \mid \alpha, \eta)$ is real-valued and increasing in θ.

We define the *bootstrap quantile function* $q_n^*(\theta \mid \alpha) = q_n(\theta \mid \alpha, \hat{\eta}(\alpha))$. We call this a bootstrap function since it is evaluated at the estimate $\hat{\eta}(\alpha)$, and is thus random. We define the $\beta-$level *grid bootstrap confidence region* for α as the set

$$C_g = \{\alpha \in R : q_n^*(\theta_1 \mid \alpha) \leq S_n(\alpha) \leq q_n^*(\theta_2 \mid \alpha)\} \tag{5}$$

where $\theta_1 = 1-(1-\beta)/2$ and $\theta_2 = (1-\beta)/2$, so $\beta = \theta_2 - \theta_1$. This confidence region is defined as the set of parameter values for which the test statistic does not fall out of a $100\beta\%$ likelihood set, where the latter is defined assuming that α is the true value. If $S_n(\alpha) = \hat{\alpha} - \alpha$, we will call C_g the grid-α interval, if $S_n(\alpha) = t(\alpha)$, we will call C_g the grid-t interval. This confidence region is designed to be central or equal-tailed, i.e. it is the intersection of two one-sided confidence intervals, and is designed to have the probability of error equally likely in either direction.

When there is no nuisance parameter η (or the sampling distribution does not depend on η) the quantile functions $q_n(\theta \mid \alpha)$ are non-random and the confidence region C_g has exact coverage. In this context the region C_g is traditional confidence region, described in most standard statistics texts. See, e.g. Cramér (1946, Chapter 34).

The confidence region C_g has several properties which are shared by confidence regions constructed from criterion functions[4]: that C_g may be disjoint, empty, or may contain the entire parameter space. This is not necessarily undesirable; indeed Dufour (1997) has shown

[4]For example, the Anderson-Rubin (1949) confidence interval for parameters in a linear simultaneous equations model.

that in certain contexts it is necessary for properly-sized confidence intervals to be unbounded with positive probability. (In such cases the natural conclusion is that the sample provides no information concerning the parameter.)

In most cases C_g will be a simple interval, with the left endpoint α_L given by the intersection of $S_n(\alpha)$ and $q_n^*(\theta_2 \mid \alpha)$, and the right endpoint α_U given by the intersection of $S_n(\alpha)$ and $q_n^*(\theta_1 \mid \alpha)$. When C_g is disjoint, a conservative confidence interval $\overline{C}_g = [\alpha_L, \alpha_U]$ for α can be defined as the convex hull of C_g.

It is helpful at this point to contrast the definition of C_g with that for a conventional bootstrap confidence region, which can be defined by the formula

$$C_b = \{\alpha \in R : q_n^*(\theta_1 \mid \hat{\alpha}) \leq S_n(\alpha) \leq q_n^*(\theta_2 \mid \hat{\alpha})\}.$$

The difference between the grid bootstrap region C_g and the conventional bootstrap region C_b is that the former allows the bootstrap quantile function $q_n^*(\theta_1 \mid \alpha)$ to be a free function of α, while the conventional bootstrap approximates this function by evaluating it at the point estimate $\hat{\alpha}$.

2.2 Bootstrap Consistency

It is helpful to first state the conditions for first-order accuracy of the conventional bootstrap.

Proposition 1 *Suppose that for a sequence of estimators $\hat{\alpha}$ and $\hat{\eta} \in \Xi$, $|\hat{\alpha} - \alpha| \to_p 0$ and $d(\hat{\eta}, \eta) \to_p 0$. Suppose that for all sequences α_n and $\eta_n \in \Xi$ such that $|\alpha_n - \alpha| \to 0$ and $d(\eta_n, \eta) \to 0$ then $F_n(x \mid \alpha_n, \eta_n)$ converges weakly to a continuous distribution function $F(x \mid \alpha, \eta)$. Then $P(\alpha \in C_b) \to \beta$ as $n \to \infty$.*

Proposition 1 states that the conventional bootstrap confidence interval has correct first-order coverage if the parameters are consistently estimated and $S_n(\alpha)$ has an asymptotic distribution, where the convergence to the asymptotic distribution is locally uniform in the parameter space. We can give an analogous condition for the grid bootstrap.

Proposition 2 *Suppose that for a sequence of estimators $\hat{\eta}(\alpha) \in \Xi$, $d(\hat{\eta}(\alpha), \eta) \to_p 0$. Suppose that for all sequences η_n such that $d(\eta_n, \eta) \to 0$ then $F_n(x \mid \alpha, \eta_n)$ converges weakly to a continuous distribution function $F(x \mid \alpha, \eta)$. Then $P(\alpha \in C_g) \to \beta$ as $n \to \infty$.*

Proposition 2 gives conditions under which the grid bootstrap confidence interval is first-order accurate. The requirement is that the nuisance parameter are consistently estimated,

while no restriction is made concerning the estimate of the parameter of interest. The conditions of Proposition 2 are strictly less restrictive than those of Proposition 1, suggesting that the grid bootstrap applies more generally than the conventional bootstrap, in the sense of first-order asymptotic coverage.

Carpenter (1999) examines an analog of our grid bootstrap intervals, where the nuisance parameter estimate $\hat{\eta}(\alpha) = \hat{\eta}$ does not depend on α. He shows that if the model satisfies the properties of Hall's (1988) smooth function model then the one-sided[5] grid-t interval (his STIB interval) has coverage error of order n^{-1}, and the one-sided grid-α interval (his TIB interval) has coverage error of order $n^{-1/2}$. These are identical to the rates obtained by the percentile-t interval.

It is possible that the grid-t interval has coverage error of order n^{-1} under weaker conditions than the conventional percentile-t (in analogy to the lesser conditions of Proposition 2 relative to Proposition 1), but this is unknown and left to future research.

2.3 Computation

To calculate the grid bootstrap confidence interval C_g, we need the bootstrap quantile function $q_n^*(\theta \mid \alpha)$. These are generally unknown, but for a given α these quantiles may be estimated using simulation methods. This technique is frequently mislabeled "bootstrapping" but it is more properly labelled a simulation estimate of a bootstrap quantile. The steps of this calculation are as follows. For a given α, let $G_n^*(x \mid \alpha) = G_n\left(x \mid \alpha, \hat{\eta}(\alpha)\right)$ be the bootstrap distribution of the sample. We assume that it is possible by simulation to generate random samples X_n^* from this distribution. Generate B such samples. For each sample X_n^*, calculate the test statistic $S_n^*(\alpha)$. Sort the B simulated test statistics $S_n^*(\alpha)$. The $100\theta\%$ order statistic $\hat{q}_n^*(\theta \mid \alpha)$ is the simulation estimate of $q_n^*(\theta \mid \alpha)$. The accuracy of the estimate is increasing in B. Setting $B = 999$ or $B = 1999$ are common choices[6] which result in fairly accurate estimates of relevant quantiles.

We need to calculate the bootstrap quantile function $q_n^*(\theta \mid \alpha)$ as a function of α, not just for a single value. A numerically intensive (but feasible) solution is to select a fine grid $A_G = [\alpha_1, \alpha_2, ..., \alpha_G]$ and calculate $\hat{q}_n^*(\theta \mid \alpha)$ at each $\alpha \in A_G$. A more efficient solution is to recognize that $\hat{q}_n^*(\theta \mid \alpha)$ is likely to be a smooth functions of α, so non-parametric methods are appropriate. A computationally straightforward method is to apply kernel regression.

[5]He does not examine symmetric two-sided intervals.

[6]It is convenient to pick B so that $(B+1)\theta = i$ is an integer, for then the i'th ordered value of $S_n^*(\alpha)$ is the quantile estimate $\hat{q}_n^*(\theta \mid \alpha)$.

The technique works as follows. First, pick a grid $A_G = [\alpha_1, \alpha_2, ..., \alpha_G]$ and calculate $\hat{q}_n^*(\theta \mid \alpha)$ by simulation at each $\alpha \in A_G$. Second, smooth the estimated function $\hat{q}_n^*(\theta \mid \alpha)$ using kernel regression. For given α, the kernel estimate is

$$\tilde{q}_n^*(\theta \mid \alpha) = \frac{\sum_{j=1}^{G} K\left(\frac{\alpha - \alpha_j}{h}\right) \hat{q}_n^*(\theta \mid \alpha_j)}{\sum_{j=1}^{G} K\left(\frac{\alpha - \alpha_j}{h}\right)},$$

where $K(u)$ is a kernel function and h is a bandwidth. In my simulations and applications I use the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)\,1(|u| \leq 1)$ and pick the bandwidth by least-squares cross-validation. In practice, graphical displays of the unsmoothed estimates $\hat{q}_n^*(\theta \mid \alpha)$ and the smoothed estimates $\tilde{q}_n^*(\theta \mid \alpha)$ helps to assess estimation accuracy and bandwidth choice.

The estimator $\tilde{q}_n^*(\theta \mid \alpha)$ is not arbitrary. Chamberlain (1994) shows that a parametric regression function fit to the estimates $\hat{q}_n^*(\theta \mid \alpha)$ is a near-efficient GMM estimator of the unknown quantile functions $q_n^*(\theta \mid \alpha)$. Since the form of the quantile functions is unknown and α is scalar it is appropriate to use a non-parametric estimator. We select kernel regression since it is particularly flexible and easy to implement.

The confidence set C_g is defined as the set of points α for which $S_n(\alpha)$ lies between $q_n^*(\theta_1 \mid \alpha)$ and $q_n^*(\theta_2 \mid \alpha)$. Our estimate $\hat{C}_g$ is defined analogously:

$$\hat{C}_g = \{\alpha \in R : \tilde{q}_n^*(\theta_1 \mid \alpha) \leq S_n(\alpha) \leq \tilde{q}_n^*(\theta_2 \mid \alpha)\}.$$

This may be found graphically by displaying graphs of $\tilde{q}_n^*(\theta_1 \mid \alpha)$, $S_n(\alpha)$, and $\tilde{q}_n^*(\theta_2 \mid \alpha)$. In most cases $\hat{C}_g$ will be a simple interval, with the left endpoint $\hat{\alpha}_L$ given by the intersection of $S_n(\alpha)$ and $\tilde{q}_n^*(\theta_2 \mid \alpha)$, and the right endpoint $\hat{\alpha}_U$ given by the intersection of $S_n(\alpha)$ and $\tilde{q}_n^*(\theta_1 \mid \alpha)$.

An important issue for implementation is the choice of B and G. Consistent estimation of the bootstrap quantile function $q_n^*(\theta \mid \alpha)$ requires that both B and G diverge to infinity, which suggests that both need to be fairly large in applications, but it is unclear how to determine their values optimally in practice. While the broader goal is to calculate confidence intervals with the desired coverage accuracy, the more narrow practical goal is to accurately estimate C_g as defined in (5) which reduces to accurate estimation of the endpoints α_L and α_U determined by the intersections of $S_n(\alpha)$ with $q_n^*(\theta_1 \mid \alpha)$ and $q_n^*(\theta_2 \mid \alpha)$, respectively. The accuracy of the estimates $\hat{\alpha}_L$ and $\hat{\alpha}_U$ can be assessed as follows. Consider $\hat{\alpha}_L$. Let $D_q =$

$\frac{d}{d\alpha}q_n^*(\theta_2 \mid \alpha)$ and $D_s = \frac{d}{d\alpha}S_n(\alpha)$ and assume $D_q > D_s$. If $|\tilde{q}_n^*(\theta_2 \mid \alpha) - q_n^*(\theta_2 \mid \alpha)| \leq \varepsilon$, then based on a linear approximation, $|\hat{\alpha}_L - \alpha_L| \leq \varepsilon / (D_q - D_s)$. This shows that the accuracy of $\hat{\alpha}_L$ as an estimate of α_L depends on the accuracy of $\tilde{q}_n^*(\theta_2 \mid \alpha)$ (which can be improved by increasing B), but also on the relative slopes $D_q - D_s$. α_L will be more accurately estimated when $D_q - D_s$ is large (for example, when $q_n^*(\theta_2 \mid \alpha)$ is increasing in α). The estimate is particularly inaccurate if $D_q \approx D_s$, which occurs when the functions $q_n^*(\theta_2 \mid \alpha)$ and $S_n(\alpha)$ have similar slopes at α_L. We show later (Figre 3) that this is an empirically relevant concern. When it is determined that $D_q \approx D_s$, the number of bootstrap replications B will have to be increased to ensure accurate endpoint estimation.

Some experimentation suggested that choices are low as $G = 25$ and $B = 399$ can produce reasonable results in many cases. (These setting are used in the simulations of section 4.2). For the empirical results reported in section 5, we use higher resolution, setting $G = 200$ and $B = 1999$ in most cases. In all cases, the computation requirements are quite modest.

3 Autoregressive Models

3.1 Bootstrap Intervals

The observed data is $(Y_{-k+1}, ..., Y_1, ..., Y_n)$. The AR(k) model with trend is

$$\begin{aligned} Y_t &= \mu_0 + \mu_1 t + y_t \quad (6) \\ y_t &= a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_k y_{t-k} + e_t, \quad (7) \end{aligned}$$

$t = 1, ..., n$, with e_t independent and identically distributed (iid) with unknown distribution function $P(\cdot)$, $Ee_t = 0$, and $Ee_t^2 < \infty$. The initial condition $y_0^* = (y_{-k+1}, \ldots, y_0)$ is fixed.

The single-equation representation is

$$Y_t = \mu_0' + \mu_1' t + a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_k Y_{t-k} + e_t. \quad (8)$$

An alternative representation is the so-called ADF reparameterization

$$Y_t = \mu_0' + \mu_1' t + \rho_1 Y_{t-1} + \rho_2 \Delta Y_{t-1} + \cdots + \rho_k \Delta Y_{t-k+1} + e_t. \quad (9)$$

We focus on (9). Let $\rho = (\rho_1, ..., \rho_k)$ be the autoregressive parameters from (9). Our goal is to construct confidence intervals for ρ.

The model is estimated by least-squares. Let $\hat{\rho}$ denote the estimate of ρ from OLS estimation of (9). Let $\hat{e}_t$ denote the OLS residuals and $(s_1, ..., s_k)$ the least-squares standard errors for $\hat{\rho}$. Let $t_j = (\hat{\rho}_j - \rho_j) / s_j$ denote the t-statistic for ρ_j.

The model is described by the parameters (ρ, P, μ_0, μ_1). The sampling distribution of the t-statistic t_j and estimate $\hat{\rho}_j$ are invariant to the values of μ_0 and μ_1, so to evaluate these distributions we can describe the model by the set $(\rho, P) \in \left(R^k \times \Xi\right)$, where Ξ is the space of mean-zero random variables. P is estimated by $\hat{P}$, the empirical distribution of the residuals $\hat{e}_t$. (Since an intercept is included in the regression, note that $\hat{P} \in \Xi$.)

The conventional percentile-t confidence interval for ρ_j, denoted $C_b(j)$, is obtained by evaluating the sampling distribution of the t-statistic t_j at the point estimate $(\hat{\rho}, \hat{P})$. This may be numerically implemented by simulating time-series from the equation (7) using the autoregressive coefficients $\hat{\rho}$ and drawing the errors independently from the OLS residuals $(\hat{e}_1, ..., \hat{e}_n)$.

The grid bootstrap confidence interval for ρ_j, denoted $C_g(j)$, is obtained by taking that parameter as the parameter of interest (α in the notation of section 2) and the remaining parameters as nuisance parameters (η in the notation of section 2). For example, the grid bootstrap confidence interval for ρ_1 is constructed by setting $\alpha = \rho_1$ and $\eta = (\rho_2, ..., \rho_k, P)$. Our constrained estimate of η is $\hat{\eta}(\alpha) = (\hat{\rho}_2(\alpha), ..., \hat{\rho}_k(\alpha), \hat{P})$ where $\hat{P}$ is described above, and $(\hat{\rho}_2(\alpha), ..., \hat{\rho}_k(\alpha))$ are the constrained OLS estimates of (9), imposing the constraint $\rho_1 = \alpha$. These can be computed from OLS regression of $Y_t - \alpha Y_{t-1}$ on $(1, t, \Delta Y_{t-1}, ..., \Delta Y_{t-k+1})$. Let $\hat{\rho}(\alpha) = (\alpha, \hat{\rho}_2(\alpha), ..., \hat{\rho}_k(\alpha))$. The grid bootstrap evaluates the sampling distribution of the test statistic at the estimates $(\alpha, \hat{\eta}(\alpha)) = \left(\hat{\rho}(\alpha), \hat{P}\right)$. This may be numerically implemented by simulating time-series from the equation (7) using the autoregressive coefficients $\hat{\rho}(\alpha)$ and drawing the errors independently (with replacement) from $(\hat{e}_1, ..., \hat{e}_n)$.

3.2 Bootstrap Consistency

Bose (1988) has shown that for stationary autoregressions with finite 8th moments,

$$\sup_x \left| F_n(x \mid \phi) - F_n(x \mid \hat{\phi}) \right| = o(n^{-1/2})$$

(in the notation of section 2.2) which is a stronger result than Proposition 1. Under weaker conditions, we now show that the conventional and grid bootstrap methods achieve first-order asymptotic coverage. Let $a(z) = \left(1 - a_1 Z - a_2 Z^2 - \cdots - a_k Z^k\right)$ be the lag polynomial corresponding to model (7).

Theorem 1 *In model (6)-(7), if all roots of $a(z)$ lie outside the unit circle and $Ee_t^{2r} < \infty$ for some $r > 1$ then the conditions of Proposition 1 are satisfied. Hence for all $1 \leq j \leq k$, $P\left(\rho_j \in C_b(j)\right) \to \beta$ and $P\left(\rho_j \in C_g(j)\right) \to \beta$ as $n \to \infty$.*

Theorem 1 shows that if the autoregression is stationary, then both the conventional bootstrap and the grid bootstrap achieve first-order correct asymptotic coverage. Since the limiting distributions of the t-statistics are standard normal, we expect that both methods achieve asymptotic refinements (along the lines of Carpenter (1999)), but do not provide a proof here.

We now consider near-non-stationary autoregressions and focus on confidence intervals for ρ_1. Suppose that

$$\begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + \cdots + \rho_k \Delta y_{t-k+1} + e_t \\ \rho_1 &= 1 + C/n \end{aligned}$$

with C fixed as $n \to \infty$. Assume that the roots of $\overline{\rho}(z) = 1 - \rho_2 z - \cdots - \rho_k z^{k-1}$ lie outside the unit circle. This is known as a near unit root model, nesting a pure unit root ($C = 0$), large stationary roots ($C < 0$) and mildly explosive roots ($C > 0$). Technically, we should let ρ_1 and y_t depend on n (array notation) but we will not do so here to preserve continuity in notation.

In this model, it is known that the asymptotic distributions of the OLS estimate $\hat{\rho}_1$ and its t-statistic are non-standard, indeed, letting $\overline{\rho} = \overline{\rho}(1)$,

$$n(\hat{\rho}_1 - \rho_1) \to_d \overline{\rho} \frac{\int_0^1 W_c dW}{\int_0^1 W_c^2}, \tag{10}$$

$$t_1 = \frac{\hat{\rho}_1 - \rho_1}{s_1} \to_d \frac{\int_0^1 W_c dW}{\left(\int_0^1 W_c^2\right)^{1/2}} \tag{11}$$

where W is a standard Brownian motion and W_c is a detrended Gaussian diffusion process with parameter $c = C/\overline{\rho}$. Since the asymptotic distributions depend on C which is not consistently estimable (observe that $\hat{C} = n(\hat{\rho}_1 - 1)$ has an asymptotic distribution given by (10)), it follows that the conventional bootstrap cannot achieve first-order asymptotic consistency. This was observed by Basawa, et. al. (1991). The grid bootstrap, in contrast, is able to produce consistent coverage.

Theorem 2 *In model (6)-(7), if $\rho_1 = 1 + C/n$, all roots of $\overline{\rho}(z)$ lie outside the unit circle, and $Ee_t^{2r} < \infty$ for some $r > 1$, then $P(\rho_1 \in C_g(1)) \to \beta$ as $n \to \infty$.*

Theorems 1 and 2 together show that the grid bootstrap achieves bootstrap consistency for ρ_1 throughout the parameter space, including $\rho_1 = 1$, unlike the conventional percentile-t bootstrap. This global property is good news for the grid bootstrap, for it suggests that this

procedure should be less sensitive to model parameters than the conventional bootstrap, and thus has better size[7].

While it would be desirable to establish an asymptotic refinement for the grid bootstrap, it appears impossible given present knowledge because it is unknown whether there exists an Edgeworth expansion for this model. While Abadir (1993) and Knight and Satchell (1993) provide Edgeworth expansions for the random walk model with Gaussian errors, their arguments do not carry over to the non-Gaussian case. Furthermore, in AR(k) models other than the AR(1), it is quite likely that an asymptotic refinement is impossible, since the limiting distributions (10) and (11) depend on the nuisance parameters $(\rho_2, ..., \rho_k)$ through c.

The asymptotic distributions of the t-statistics for ρ_2 through ρ_k, however, are standard normal, so the both the conventional percentile-t bootstrap and the grid bootstrap are expected to achieve first-order asymptotic consistency for the coefficients $(\rho_2, ..., \rho_k)$. We do not provide a proof of this proposition, however, as it appears to involve a rather delicate argument. As the nuisance parameter C is not consistently estimated, the conditions of Proposition 1 and 2 are not satisfied, implying that an alternative proof method is necessary. Furthermore, since the asymptotic distribution of the t-statistics are standard normal, it is possible that the bootstrap procedures will achieve an asymptotic refinement. This is not obvious, however, since the nuisance parameter c is not consistently estimated. Whether or not the bootstrap methods achieve an asymptotic refinement may depend on whether the second term in the Edgeworth expansion depends on c (and this is currently unknown).

4 Monte Carlo Simulations

To assess the performance of the grid bootstrap in practice, we report two Monte Carlo experiments, for the AR(1) and AR(2) models, respectively. We consider the problem of constructing central confidence intervals with $100\beta = 90\%$ coverage. (Equivalently, we are assessing the properties of one-sided 95% confidence intervals.) A correctly constructed central confidence interval will have the property that in 5% of the samples the true value will lie to the left of the confidence interval, and in 5% of the samples the true value will lie to the right of the confidence interval.

The bootstrap methods require a treatment of the initial condition. We treat the initial condition $y_0^* = (y_{-k+1}, ... , y_0)$ as fixed, and estimate its value by the first k values from least

[7]The *size* of a test is the maxmimum probability of the Type I error over the values of the parameters.

squares detrending of the time-series. These fixed initial values are then fixed for each bootstrap evaluation. Since the bootstrap distributions are quite sensitive to the initial conditions when the model is explosive we set the initial condition to zero when the autoregressive parameters used in the bootstrap evaluation have an explosive or unit root. The treatment of the initial conditions was symmetric across the various bootstrap methods compared.

All experiments are based on 5000 replications, and consider samples of size $n = 60$, 120, and 240.

4.1 AR(1) Model

Samples were generated from the AR(1) model $y_t = \alpha y_{t-1} + e_t$ setting $\alpha = .6$, .9, 1.0 and 1.02 using Gaussian errors. The initial value y_0 was drawn from the unconditional distribution when $\alpha < 1$, and was set to zero for $\alpha \geq 1$.

Seven confidence interval methods were compared[8]. All methods are based on OLS estimation of the AR(1) with trend

$$Y_t = \mu_0 + \mu_1 t + \rho Y_{t-1} + e_t.$$

1. Conventional asymptotic confidence interval $\hat{\alpha} \pm 1.645 s(\hat{\alpha})$.
2. Stock's (1991) local-to-unity asymptotic confidence interval.
3. Percentile bootstrap (e.g. Efron and Tibshirani, 1993, chapter 13).
4. Percentile-t bootstrap (e.g. Hall, 1992).
5. Biased-corrected percentile bootstrap (Kilian, 1998).
6. Grid-α.
7. Grid-t.

The conventional bootstrap methods used $B = 999$ simulated samples for each Monte Carlo replication. For the grid bootstrap methods, we A_G to be $G = 50$ evenly spaced points on $[\hat{\alpha} \pm 6s(\hat{\alpha})]$ and generated $B = 399$ simulated samples at each grid point. The quantiles were smoothed using an Epanechnikov kernel with the bandwidth selected by least-squares cross-validation. When the confidence intervals C_g are disjoint, we report the convexified region $\overline{C}_g$.

[8]To be fair, most of the existing methods (other than Stock's interval) were designed and motivated for stationary autoregressions, so there is no reason to be surprised if they fail to provide good coverage for near non-stationary cases.

The results are summarized in Table 1. For each method of computing confidence intervals, the percentage of samples in which the true value of α lay to the left of the estimated confidence interval is reported in the row labeled P_L. Similarly, the percentage of samples in which the true value of α lay to the right of the estimated confidence interval is reported in the row labeled P_R. Ideally, these percentages should be close to 0.05. (The standard errors for the percentages are 0.003).

Table 1: Actual AR(1) Confidence Interval Type I Error

	$\alpha = .6$		$\alpha = .9$		$\alpha = 1.0$		$\alpha = 1.02$	
	P_L	P_R	P_L	P_R	P_L	P_R	P_L	P_R
$n = 60$								
Asymptotic	.01	.14	.00	.34	.00	.76	.00	.89
Stock	.03	.17	.05	.06	.06	.05	.05	.06
Percentile	.00	.35	.00	.87	.00	1.00	.00	1.00
Percentile-t	.06	.07	.05	.13	.02	.31	.01	.43
Kilian	.03	.09	.00	.13	.00	.33	.00	.48
Grid-α	**.05**	**.05**	**.04**	**.04**	**.05**	**.05**	**.03**	**.07**
Grid-t	**.05**	**.05**	**.05**	**.04**	**.05**	**.04**	**.03**	**.07**
$n = 120$								
Asymptotic	.01	.12	.00	.23	.00	.77	.00	.64
Stock	.00	.50	.05	.06	.05	.05	.04	.05
Percentile	.00	.24	.00	.61	.00	1.00	.00	.96
Percentile-t	.05	.06	.07	.09	.02	.30	.15	.28
Kilian	.04	.07	.00	.09	.00	.29	.00	.73
Grid-α	**.04**	**.04**	**.05**	**.05**	**.05**	**.05**	**.07**	**.07**
Grid-t	**.05**	**.04**	**.05**	**.05**	**.05**	**.05**	**.06**	**.07**
$n = 240$								
Asymptotic	.02	.08	.01	.16	.00	.77	.05	.11
Stock	.00	1.00	.04	.07	.05	.05	.04	.05
Percentile	.01	.15	.00	.41	.00	1.00	.01	.19
Percentile-t	.05	.05	.06	.07	.02	.31	.05	.03
Kilian	.04	.05	.03	.07	.00	.27	.01	.19
Grid-α	**.05**	**.05**	**.05**	**.05**	**.05**	**.05**	**.04**	**.04**
Grid-t	**.05**	**.05**	**.05**	**.05**	**.05**	**.05**	**.05**	**.05**

The asymptotic interval is biased downwards, which means that the true value is too rarely to the left of the confidence interval, and is much too often to the right of the confidence interval. The percentile method is even more biased than the asymptotic interval, with considerably worse sampling performance. The percentile-t method works well in large samples with stationary roots, but is quite biased at or near the unit root. Kilian's bias-corrected

percentile method performs somewhat similar to the percentile-t method, with poor coverage probabilities in small samples or near the unit root. The only non-grid method which provides good coverage is Stock's confidence interval, which provides very accurate coverage rates for α equal or near unity (including the explosive case), but is quite poor for $\alpha = .6$, and the error increases with sample size.

In contrast, the grid bootstrap methods provide near-exact or conservative coverage for all experiments. The grid-α and grid–t are near-exact central confidence intervals for $\alpha \leq 1$, but are slightly biased for $\alpha > 1$. In addition to the coverage rates, the median length of the confidence intervals was also computed (but not reported). The grid-α and grid-t intervals had nearly identical lengths.

This simulation verifies the predictions of the theory: the grid bootstrap methods are the only confidence interval techniques which work well globally in the parameter space. Other methods work well in subsets of the parameter space, but not globally, and hence have poor size.

4.2 AR(2) Model

Samples were generated from the AR(2) model

$$y_t = \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t.$$

setting $\rho_1 = .6$, .9, and 1.0, $\rho_2 = -.4$ and .4, and using Gaussian errors. The initial values (y_{-1}, y_0) were drawn from the unconditional distribution when $\rho_1 < 1$, and were set to zero for $\rho_1 = 1$.

Three confidence interval methods for ρ_1 and ρ_2 were compared. All methods are based on OLS estimation of the AR(2) with trend

$$Y_t = \mu + \beta t + \rho_1 Y_{t-1} + \rho_2 \Delta Y_{t-1} + e_t.$$

We compared the performance of the conventional asymptotic 90% confidence interval, the percentile-t interval, and the grid-t interval. (In AR(k) models, the Stock procedure provides confidence intervals for the largest autoregressive root, not for the individual autoregressive parameters, so could not be included in this comparison).

To reduce the computation burden, the grid-t intervals were calculated using $G = 25$ evenly spaced points on $[\hat{\alpha} \pm 5s(\hat{\alpha})]$. Otherwise the methods are identical as described in the previous section.

The results are summarized in Tables 2 and 3 for ρ_1 and ρ_2, respectively. The tables report the probabilities of the two errors $P_L = P(\rho < C)$ and $P_R = P(\rho > C)$, which should be close to 5% if the methods are working properly.

For nearly all the cases considered, the grid-t interval has near correct coverage. In contrast, the asymptotic intervals for ρ_1 are quite poor in all cases considered, and the asymptotic intervals for ρ_2 are poor when $\rho_2 = -.4$. The percentile-t intervals for ρ_1 perform similarly to the results found for the AR(1) model, with good performance in large samples and stationary parameters, but poor performance when ρ_1 is near unity. The percentile-t intervals for ρ_2, on the other hand, are generally quite well behaved, except when $n = 60$ and $\rho_1 = 1$.

In summary, the simulation results confirm the theoretical predictions. The asymptotic and percentile-t intervals for ρ_1 are highly inaccurate when ρ_1 is large and this inaccuracy can be eliminated through the use of the grid-t bootstrap. For inference on the parameter ρ_2, both the conventional percentile-t interval and the grid-t interval are quite accurate.

Table 2: AR(2) Confidence Intervals for ρ_1

		$n = 60$			$n = 120$			$n = 240$		
	ρ_1	.6	.9	1.0	.6	.9	1.0	.6	.9	1.0
$\rho_2 = -.4$										
Asymptotic	P_L	.00	.00	.00	.01	.00	.00	.02	.01	.00
	P_R	.20	.42	.77	.14	.28	.77	.10	.20	.77
Percentile-t	P_L	.07	.05	.02	.07	.06	.02	.06	.07	.02
	P_R	.09	.15	.31	.06	.10	.31	.06	.08	.31
Grid-t	P_L	.05	.06	.04	.06	.05	.05	.05	.06	.05
	P_R	.04	.04	.06	.04	.04	.05	.05	.05	.06
$\rho_2 = .4$										
Asymptotic	P_L	.01	.00	.00	.02	.01	.00	.02	.01	.00
	P_R	.14	.28	.76	.10	.18	.77	.08	.13	.78
Percentile-t	P_L	.06	.07	.03	.06	.06	.02	.05	.06	.02
	P_R	.07	.11	.31	.06	.08	.31	.05	.06	.31
Grid-t	P_L	.05	.06	.05	.05	.05	.05	.05	.05	.05
	P_R	.05	.05	.06	.05	.05	.06	.05	.04	.06

Table 3: AR(2) Confidence Intervals for ρ_2

		$n = 60$			$n = 120$			$n = 240$		
	ρ_1	.6	.9	1.0	.6	.9	1.0	.6	.9	1.0
$\rho_2 = -.4$										
Asymptotic	P_L	.16	.15	.20	.12	.12	.14	.09	.10	.11
	P_R	.01	.01	.01	.02	.02	.01	.02	.02	.02
Percentile-t	P_L	.07	.08	.10	.06	.06	.07	.06	.05	.06
	P_R	.06	.06	.05	.06	.05	.04	.05	.05	.05
Grid-t	P_L	.05	.05	.07	.05	.05	.06	.05	.05	.06
	P_R	.05	.05	.04	.05	.04	.04	.05	.05	.04
$\rho_2 = .4$										
Asymptotic	P_L	.07	.07	.05	.06	.06	.05	.06	.06	.05
	P_R	.03	.04	.06	.04	.04	.05	.04	.04	.05
Percentile-t	P_L	.06	.06	.04	.05	.06	.04	.06	.05	.05
	P_R	.06	.06	.09	.06	.06	.07	.05	.05	.06
Grid-t	P_L	.05	.05	.04	.05	.05	.04	.05	.05	.04
	P_R	.05	.05	.07	.05	.06	.06	.05	.05	.05

5 Empirical Illustrations

5.1 Velocity

We now illustrate the application of the grid bootstrap to an AR(1) using a widely-studied data series. We use the annual log velocity series of Nelson and Plosser (1982) extended by Schotman and van Dijk (1991) to cover the years 1869-1988. The AR(1) specification was used by Nelson and Plosser in their analysis so we adopt the same specification here. Figure 1 shows the construction of the percentile-t and grid-t confidence intervals. The solid downward sloping line is the t-statistic function $t(\alpha) = (\hat{\alpha} - \alpha)/s(\hat{\alpha})$. The dashed lines are the 5% and 95% bootstrap quantile functions, which are clearly quite non-linear in α. The open circles denote the intersection points, and the black arrows indicate the endpoints of the grid-t interval. The percentile-t intervals can be read using the dotted lines, which show how the percentile-t approximates the bootstrap quantile functions by flat lines at the OLS estimate $\hat{\alpha}$. The open arrows denote the percentile-t endpoints.

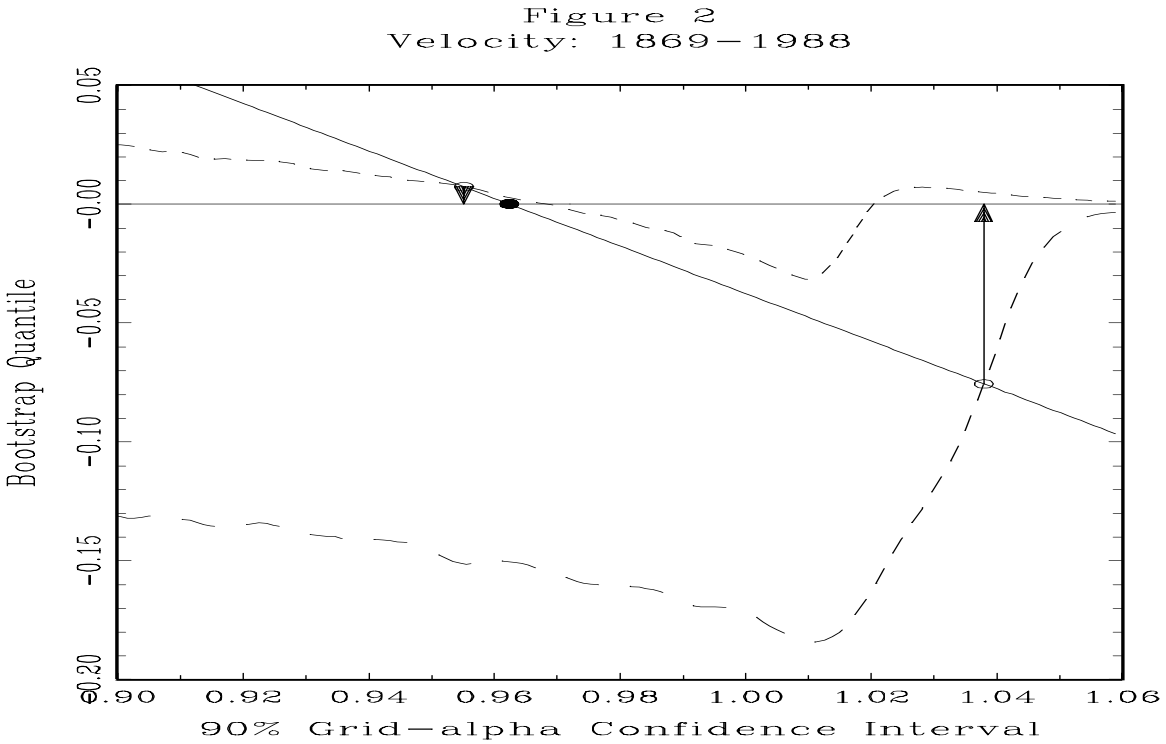


Figure 2 shows a similar construction for the grid-α interval. The solid line denotes the normalized estimate function $b(\alpha) = \hat{\alpha} - \alpha$, and the dotted lines the 5% and 95% bootstrap quantile functions. The intersections denote the endpoints of this grid bootstrap confidence interval.

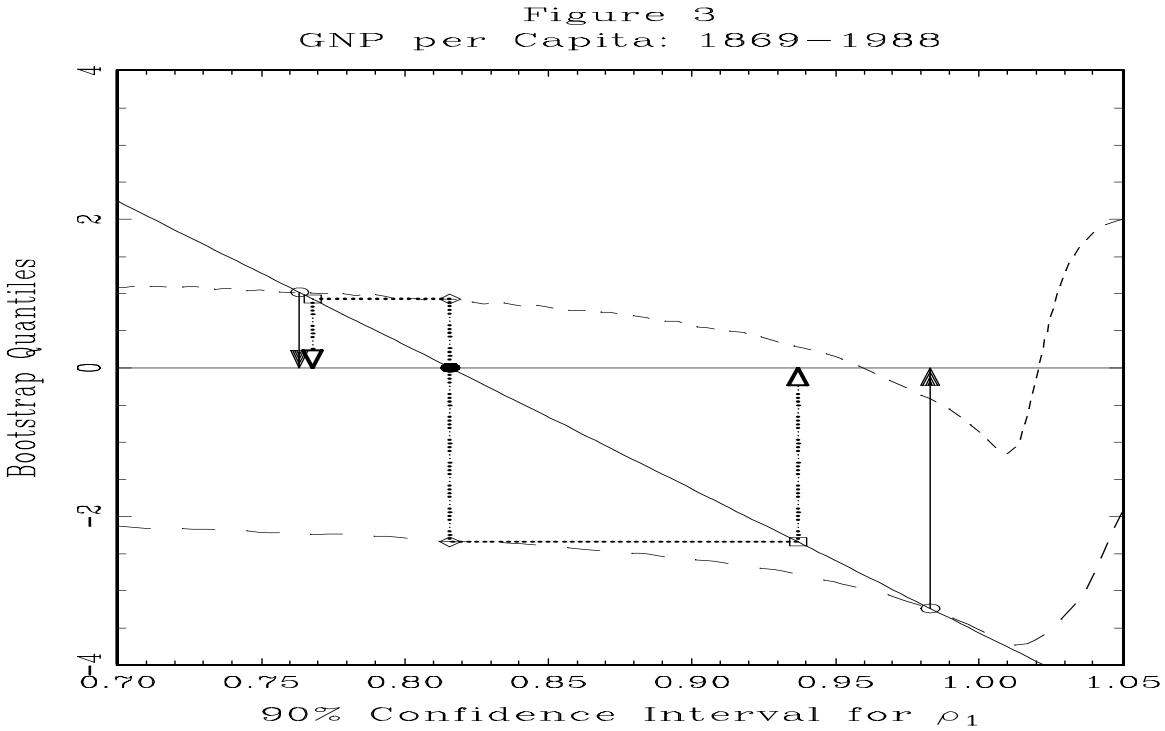
Table 4: Extended Nelson-Plosser Velocity, 1869-1988, AR(1)

	α_L	α_U
Asymptotic	.924	1.0011
Stock	.944	1.036
Percentile	.813	.968
Percentile-t	.958	1.030
Kilian	.862	1.015
Grid-α	.955	1.038
Grid-t	.956	1.034

Table 2 reports the confidence interval endpoints constructed by the various methods described in section 4.1. For these calculations, the conventional bootstrap methods used 1999 bootstrap replications, and the grid bootstrap method used 1999 at each of 200 gridpoints.

5.2 Real Per Capita GNP

Considerable attention has been devoted to the persistence properties of U.S. output. We now explore this series (real per capita GNP from 1909 through 1988) in greater detail. Following Nelson and Plosser (1982), we use an AR(2) with trend. 90% confidence intervals were calculated for the two AR coefficients both by the percentile-t method and the grid-t bootstrap using 9999 bootstrap replications and a grid with 200 gridpoints. The results are reported in Table 5. Figures 3 and 4 plot the bootstrap grid-t quantile functions and 90% confidence intervals for the two parameters.



The reason for the choice of an extremely large number of bootstrap replications was

because of the difficulty in identifying the right endpoint of the grid-t interval for ρ_1. As can be seen in Figure 3, the right endpoint lies in a downward sloping region of the function $q_n^*(.05 \mid \alpha)$, and it appears that the slope is quite similar to that of the t-statistic function. As a result, the exact point of intersection is difficult to determine unless $q_n^*(.05 \mid \alpha)$ is precisely estimated, as discussed in section 2.3. It is for this reason that the number of bootstrap replications was set high. Observe that this contrasts with the situation in Figures 1 and 2, where the right endpoint lies in the region where the quantile functions are upward sloping, implying greater estimation precision.

Figure 4
GNP per Capita: 1869—1988

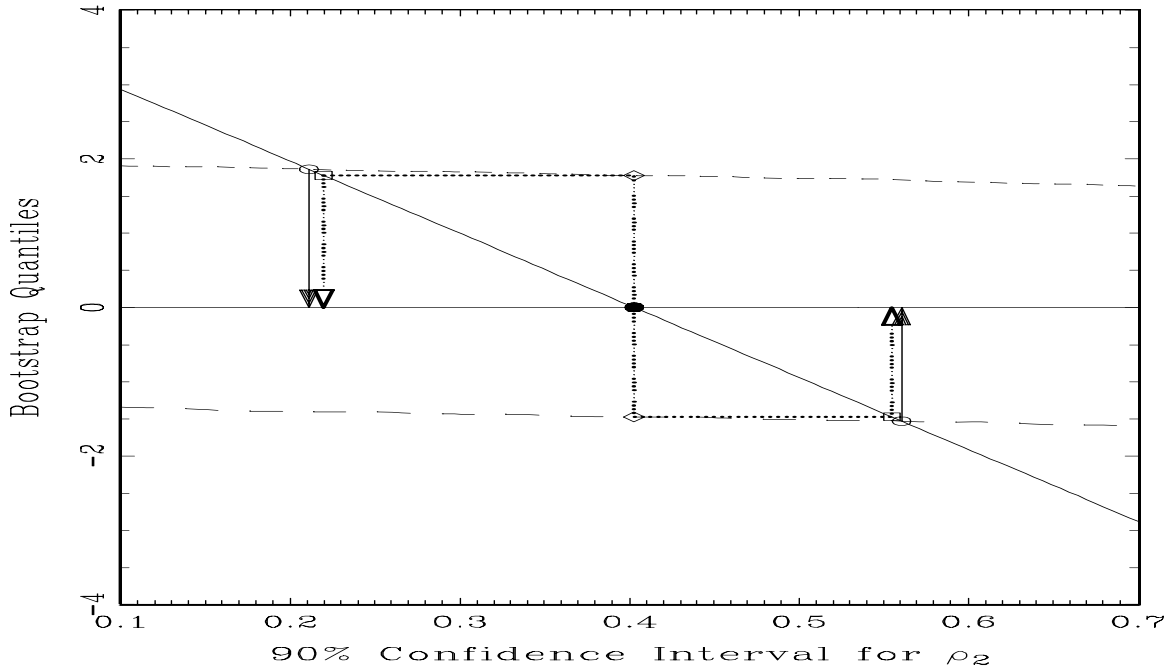


Table 5: Real per Capita GNP, 1909-1988
90% Bootstrap Confidence Intervals, AR(2)

Parameter	$\hat{\rho}$	$s(\hat{\rho})$	Percentile-t interval	Grid-t Bootstrap interval
ρ_1	.816	.052	(.768, .937)	(.763, .983)
ρ_2	.402	.103	(.220, .554)	(.211, .560)

The grid-t bootstrap yields strictly larger confidence intervals than the percentile-t bootstrap for both parameters. Most notably, the right endpoint of the confidence interval for ρ_1 is much larger (.983) using the grid bootstrap than the percentile-t (.937). The reason can be seen from Figure 3. The lower bootstrap quantile function falls as ρ_1 increases towards unity (reflecting the increased bias in OLS estimation), and this drop-off is entirely missed by the percentile-t bootstrap. The left endpoint is less affected because the upper bootstrap quantile function is relatively flat in that region.

Figure 4 shows that the distribution of the t-statistic for ρ_2 is relatively insensitive to the value of ρ_2, so that the grid-t and percentile-t intervals will be quite similar. Never-the-less, the grid-t bootstrap produces a slightly larger confidence interval for ρ_2 since the bootstrap quantile functions (lower and upper) are decreasing as ρ_2 moves towards unity.

5.3 Nelson-Plosser Revisited

Nelson and Plosser (1982) made a major impact on econometrics by making the forceful case that the hypothesis of an autoregressive unit root could not be rejected for most major long-term macroeconomic time series. To make this case, these authors used the Augmented Dickey-Fuller test with asymptotic critical values. Stock (1991) considered the problem of constructing asymptotic confidence intervals based on a local-to-unity approximation for the largest autoregressive root. Schotman and van Dijk (1991) extended the original Nelson-Plosser data (which ran to 1970) through the year 1988. We apply the grid bootstrap to construct confidence intervals for ρ_1 for both data sets.

We use the same transformations as Nelson-Plosser (natural logs for all series with the exception of the bond yield) and run the tests with a constant and time trend included in the regressions. We use the same autoregressive order k as Nelson-Plosser. Table 6 reports the 13 series used, the sample period, sample size, autoregressive order, leading coefficient $\hat{\rho}_1$ and standard error $s(\hat{\rho}_1)$, and 90% confidence intervals for ρ_1 constructed by the grid-t bootstrap. These bootstrap intervals were constructed using 1999 bootstrap replications at each of 200 gridpoints.

The results for the original Nelson-Plosser data are quite consistent with those reported by Nelson and Plosser. Our 90% confidence intervals contain unity for every series except the unemployment rate. The confidence intervals are quite different from those which would be constructed by an asymptotic approximation. Most notably, the left endpoints are quite close to the point estimate.

The extension of the sample to 1988 provides some important new information. While the point estimates are for the most part not meaningfully changed, most of the confidence intervals are shorter. In fact, for two of the series (Real per Capita GNP and the Unemployment Rate), the bootstrap confidence intervals do not include unity, suggesting that these series are trend-stationary. This result is consistent with the findings of Diebold and Senhadji (1996).

Table 6: Grid-t Bootstrap Confidence Intervals for ρ_1

Series	Period	n	k	$\hat{\rho}_1$	$s(\hat{\rho}_1)$	90% interval
Nelson-Plosser Data Set						
Real GNP	1909-1970	62	2	.825	.058	(.775, 1.030)
Nominal GNP	1909-1970	62	2	.907	.029	(.880, 1.014)
Real per Capita GNP	1909-1970	62	2	.818	.059	(.762, 1.031)
Industrial Production	1860-1970	111	6	.835	.063	(.783, 1.048)
Employment	1890-1970	81	3	.861	.051	(.816, 1.034)
Unemployment Rate	1890-1970	81	4	.706	.081	(.620, .954)
GNP Deflator	1889-1970	82	2	.915	.033	(.890, 1.025)
Consumer Prices	1860-1970	111	4	.968	.016	(.961, 1.019)
Wages	1900-1970	71	3	.910	.039	(.884, 1.035)
Real Wages	1900-1970	71	2	.831	.055	(.782, 1.029)
Velocity	1869-1970	102	1	.941	.035	(.929, 1.043)
Bond Yield	1900-1970	71	3	1.032	.046	(1.020, 1.078)
S&P 500	1871-1970	100	4	.908	.043	(.878, 1.040)
Extended Nelson-Plosser Data Set						
Real GNP	1909-1988	80	2	.824	.050	(.773, 1.013)
Nominal GNP	1989-1988	100	2	.936	.022	(.916, 1.012)
Real per Capita GNP	1909-1988	80	2	.816	.052	(.763, .983)
Industrial Production	1860-1988	129	6	.841	.058	(.789, 1.040)
Employment	1890-1988	99	3	.864	.046	(.822, 1.024)
Unemployment Rate	1890-1988	99	4	.715	.071	(.634, .909)
GNP Deflator	1889-1988	100	2	.968	.020	(.962, 1.025)
Consumer Prices	1860-1988	129	4	.987	.010	(.989, 1.018)
Wages	1900-1988	89	3	.939	.029	(.922, 1.028)
Real Wages	1900-1988	89	2	.929	.041	(.906, 1.042)
Velocity	1869-1988	120	1	.962	.023	(.956, 1.034)
Bond Yield	1900-1988	89	3	.953	.034	(.958, 1.051)
S&P 500	1871-1988	188	4	.932	.035	(.911, 1.036)

6 Conclusion

Conventional bootstrap methods work well in regular statistical problems. The autoregressive model is non-regular in that the asymptotic distribution of the t-statistic changes discontinuously at the unit circle. This is a context where it is known that conventional inference can fail. To remedy this problem, we propose basing confidence intervals on test statistic inversion, and call this method the grid bootstrap. This method is first-order consistent under strictly broader conditions than for the conventional bootstrap, and these conditions are satisfied for the AR(k) model with a near unit root.

We do not explore the high-order properties of our grid bootstrap. We expect the grid bootstrap to achieve an asymptotic refinement under the same conditions as the conventional percentile-t bootstrap, and possibly under more general conditions. We leave this demonstration to future research.

7 Appendix

Proof of Proposition 1: Let $\phi = (\alpha, \eta)$ and $\hat{\phi} = (\hat{\alpha}, \hat{\eta})$. Observe that

$$\begin{aligned} C_b &= \left\{\alpha \in R : F_n\left(q_n\left(\theta_1 \mid \hat{\phi}\right) \mid \hat{\phi}\right) \leq F_n\left(S_n(\alpha) \mid \hat{\phi}\right) \leq F_n\left(q_n\left(\theta_2 \mid \hat{\phi}\right) \mid \hat{\phi}\right)\right\} \\ &= \left\{\alpha \in R : \theta_1 \leq F_n\left(S_n(\alpha) \mid \hat{\phi}\right) \leq \theta_2\right\}, \end{aligned}$$

so

$$P\left(\alpha \in C_b\right) = P\left(\theta_1 \leq F_n\left(S_n(\alpha) \mid \hat{\phi}\right) \leq \theta_2\right). \tag{12}$$

As discussed in section 2.3 of Beran (1987), the assumptions (a re-statement of Beran's "Condition 1") imply that $F_n\left(x \mid \hat{\phi}\right) \to_p F\left(x \mid \phi\right)$ uniformly in x. They also imply that $F_n\left(x \mid \phi\right) \to F\left(x \mid \phi\right)$, where $F_n\left(x \mid \phi\right)$ is the distribution of $S_n(\alpha)$. Thus $F_n\left(S_n(\alpha) \mid \hat{\phi}\right) \to_d U[0,1]$ and from (12) we conclude $P\left(\alpha \in C_b\right) \to \theta_2 - \theta_1 = \beta$. □

Proof of Proposition 2: Let $F_n^*(x \mid \alpha) = F_n\left(x \mid \alpha, \hat{\eta}(\alpha)\right)$. Observe that $F_n^*(q_n^*\left(\theta \mid \alpha\right)) \mid \alpha) = \theta$. Thus

$$\begin{aligned} C_g &= \left\{\alpha \in R : F_n^*\left(q_n^*\left(\theta_1 \mid \alpha\right) \mid \alpha\right) \leq F_n^*\left(S_n(\alpha) \mid \alpha\right) \leq F_n^*\left(q_n^*\left(\theta_2 \mid \alpha\right) \mid \alpha\right)\right\} \\ &= \left\{\alpha \in R : \theta_1 \leq F_n^*\left(S_n(\alpha) \mid \alpha\right) \leq \theta_2\right\}, \end{aligned}$$

so

$$P\left(\alpha \in C_g\right) = P\left(\theta_1 \leq F_n^*\left(S_n(\alpha) \mid \alpha\right) \leq \theta_2\right).$$

The assumptions imply that $F_n^*(x \mid \alpha) = F_n(x \mid \alpha, \hat{\eta}(\alpha)) \to_p F(x \mid \alpha, \eta)$ uniformly in x, and that $F_n(x \mid \alpha, \eta) \to F(x \mid \alpha, \eta)$, where $S_n(\alpha)$ has distribution $F_n(x \mid \alpha, \eta)$. It follows that $F_n^*(S_n(\alpha) \mid \alpha) \to_d U[0,1]$, and thus $P(\alpha \in C_g) \to \theta_2 - \theta_1 = \beta$. $\square$

Proof of Theorem 1: We show that the conditions of Proposition 1 hold, and thus the conclusions of Proposition 1 and 2 follow. Let $\phi = (\rho, P)$ and $\hat{\phi} = (\hat{\rho}, \hat{P})$. Define Mallow's distance (Mallows (1972))

$$d_{2r}(H, G) = \inf_{X,Y} \|X - Y\|_{2r}$$

where the infimum is over all possible joint distributions of (X, Y) whose marginal distributions are H and G, respectively. Then set

$$d(\phi, \phi') = |\rho - \rho'| + d_{2r}(P, P').$$

Under the assumptions, it is known that $d(\phi, \hat{\phi}) \to_p 0$. (For a discussion of convergence with respect to d_{2r} see Shao and Tu (1995, section 3.1.2). It remains to show that for all sequences ϕ_n such that $P_n \in \Xi$ and $d(\phi_n, \phi) \to 0$, then $F_n(x \mid \phi_n)$ converges weakly to a continuous distribution function $F(x \mid \phi)$.

Let $x_t = \left(1 \;\; \frac{t}{n} \;\; y_{t-1} \;\; y_{t-2} \;\; \cdots \;\; y_{t-k}\right)$ and set

$$X_n = \left(\frac{1}{n}\sum_{t=1}^{n} e_t^2, \quad \frac{1}{n}\sum_{t=1}^{n} x_t x_t', \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t e_t \right).$$

It is sufficient to study X_n as the statistics $\hat{\rho}_j - \rho_j$ and $t_j = (\hat{\rho}_j - \rho_j)/s_j$ are continuous functions of X_n. Under the assumptions, it is well known that

$$X_n \to_d X = \left(\sigma^2, \quad M, \quad Z \right), \tag{13}$$

where $Z \sim N(0, \sigma^2 M)$ and

$$M = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} E x_t x_t'.$$

Now for any $\phi_n = (\rho_n, P_n)$ where $d(\phi_n, \phi) \to 0$ and $P_n \in \Xi$, let y_{nt} be determined by ϕ_n. That is, y_{nt} follows the process

$$y_{nt} = \rho_{n1} y_{nt-1} + \rho_{n2} \Delta y_{nt-1} + \cdots + \rho_{nk} \Delta y_{nt-k+1} + e_{nt} \tag{14}$$

where e_{nt} are iid mean-zero draws form P_n and $\rho_n = (\rho_{n1}, ..., \rho_{nk})$. This process is non-stationary (unless the initial condition is drawn from the unconditional distribution). As long as the initial condition is bounded it does not affect the distribution theory presented

here, so we make the simplifying assumption that (14) has been generated over an infinite time horizon.

Let $x_{nt} = \left(1 \ \frac{t}{n} \ y_{nt-1} \ y_{nt-2} \ \cdots \ y_{nt-k}\right)$ and

$$X_n^* = \left(\frac{1}{n}\sum_{t=1}^{n} e_{nt}^2, \ \ \frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}', \ \ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{nt}e_{nt} \right).$$

The proof of the theorem is completed by showing that $X_n^* \to_d X$ as $n \to \infty$. (Since the limit is the same as in (13), and the statistics are continuous functions of X_n and X_n^*, their limiting distributions are the same.)

Let R satisfy $Ee_t^{2r} < R < \infty$. Since $d_{2r}(P, P_n) \to 0$ then $Ee_{nt}^{2r} \to Ee_t^{2r} < R$ (Bickel and Freedman (1981)). Thus for $\overline{n}$ sufficiently large, for all $n \geq \overline{n}$, $Ee_{nt}^{2r} \leq R$. Thus e_{nt}^2 is uniformly integrable. Hence as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n} e_{nt}^2 = \frac{1}{n}\sum_{t=1}^{n} \left(e_{nt}^2 - Ee_{nt}^2\right) + Ee_{nt}^2 \to_p \sigma^2 \tag{15}$$

by the WLLN for independent uniformly integrable random arrays and the fact that $d_{2r}(P, P_n) \to 0$ implies $Ee_{nt}^2 \to Ee_t^2 = \sigma^2$.

We next show that

$$\frac{1}{n}\sum_{t=1}^{n} y_{nt}^2 \to_p \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Ey_t^2. \tag{16}$$

Extended arguments of this form show that $\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}' \to_p M$.

Since $a(L)y_t = e_t$ and the roots of $a(L)$ lie outside the unit circle, then $y_t = b(L)e_t$ where $b(L) = a(L)^{-1} = 1 + b_1 L + b_2 L^2 + \cdots$. Since $a(L)$ is finite-order, it follows that the b_j coefficients decline exponentially, so $\sum_{j=1}^{\infty} j\,|b_j| < \infty$. For any ρ_n there is a corresponding polynomials $a_n(L)$. Since $|\rho_n - \rho| \to 0$, there is an $\overline{n}$ sufficiently large so that for all $n \geq \overline{n}$, $b_n(L) = a_n(L)^{-1} = 1 + b_{n1}L + b_{n2}L^2 + \cdots$ exists and $\sum_{j=1}^{\infty} j\,|b_{nj}| \leq K < \infty$. For the remainder of the argument assume that $n \geq \overline{n}$. Thus $y_{nt} = b_n(L)e_{nt}$. For $j = 0, 1, ...$, let

$$f_{nj} = \sum_{k=0}^{\infty} b_{nk}b_{nk+j}.$$

Following the decomposition for squared linear processes introduced by Phillips and Solo (1992),

$$y_{nt}^2 = f_{n0}e_{nt}^2 + 2e_{nt}\tilde{e}_{nt-1} - (z_{nt} - z_{nt-1})$$

where

$$\tilde{e}_{nt-1} = \sum_{j=1}^{\infty} f_{nj}e_{nt-j}$$

and

$$z_{nt} = \sum_{k=0}^{\infty} \sum_{s=k+1}^{\infty} b_{ns}^2 e_{nt-k}^2 + 2 \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \sum_{s=k+1}^{\infty} b_{ns} b_{ns+j} e_{nt-k} e_{nt-k-j}.$$

Hence

$$\frac{1}{n} \sum_{t=1}^{n} y_{nt}^2 = f_{n0} \frac{1}{n} \sum_{t=1}^{n} e_{nt}^2 + 2 \frac{1}{n} \sum_{t=1}^{n} e_{nt} \tilde{e}_{nt-1} + \frac{z_{n0} - z_{nn}}{n}. \tag{17}$$

We now examine the terms on the right-hand-side (17). First, notice that $e_{nt}\tilde{e}_{nt-1}$ is a MDS, and is uniformly integrable since

$$\|e_{nt}\tilde{e}_{nt-1}\|_r \leq \sum_{j=1}^{\infty} |f_{nj}| \, \|e_{nt}e_{nt-1}\|_r \leq \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} |b_{nk}b_{nk+j}| \, R^{2/r} \leq \left(\sum_{k=0}^{\infty} |b_{nk}| \right)^2 R^{2/r} \leq K^2 R^{2/r} < \infty.$$

Hence $\frac{1}{n} \sum_{t=1}^{n} e_{nt}\tilde{e}_{nt-1} \to_p 0$. Second, $z_{nt} = O_p(1)$ since

$$\begin{aligned} \|z_{nt}\|_r &\leq \sum_{k=0}^{\infty} \sum_{s=k+1}^{\infty} b_{ns}^2 \left\| e_{nt-k}^2 \right\|_r + 2 \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \sum_{s=k+1}^{\infty} |b_{ns}| \, |b_{ns+j}| \, \|e_{nt-k}e_{nt-k-j}\|_r \\ &\leq \left(K + 2K^2 \right) R^{1/2r} < \infty. \end{aligned}$$

Thus $\dfrac{z_{n0} - z_{nn}}{n} \to_p 0$. Finally,

$$\frac{1}{n} \sum_{t=1}^{n} y_{nt}^2 = f_{n0} \frac{1}{n} \sum_{t=1}^{n} e_{nt}^2 + o_p(1) \to_p f_0 \sigma^2$$

by (15), where $f_0 = \sum_{k=0}^{\infty} b_k^2$. This yields (16) as needed since

$$f_0 \sigma^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E y_t^2.$$

Finally, since $x_{nt}e_{nt}$ is a martingale difference array, $n^{-1/2} \sum_{t=1}^{n} x_{nt}e_{nt} \to_d N(0, \sigma^2 M)$ if

$$\frac{1}{n} \sum_{t=1}^{n} x_{nt} x_{nt}' e_{nt}^2 \to_p \sigma^2 M \tag{18}$$

and

$$\frac{1}{\sqrt{n}} \max_{t \leq n} |x_{nt}e_{nt}| \to_p 0. \tag{19}$$

(See Davidson (1994), Theorem 24.3).

Consider (18). It consists of elements of the form

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} y_{nt-1}^2 e_{nt}^2 &= f_{n0} \frac{1}{n} \sum_{t=1}^{n} e_{nt-1}^2 e_{nt}^2 + 2 \frac{1}{n} \sum_{t=1}^{n} e_{nt-1} \tilde{e}_{nt-2} e_{nt}^2 + \frac{1}{n} \sum_{t=1}^{n} \Delta z_{nt-1} e_{nt}^2 \\ &\to {}_p f_0 \sigma^2 \sigma^2 \end{aligned}$$

by calculations similar to those above, and the facts that e_{nt}^2 is independent of e_{nt-1}^2, $e_{nt-1}\tilde{e}_{nt-2}$, and Δz_{nt-1}, that $E(e_{nt-1}\tilde{e}_{nt-2}) = 0$ and $E(\Delta z_{nt-1}) = 0$. This establishes (18).

To show (19), observe that $\|y_{nt}\|_r \leq \sum_{k=0}^{\infty} |b_{nk}| \, \|e_{nt-k}\|_r \leq KR^{1/2r}$, thus $E|x_{nt}|^{2r} \leq kK^r R$ and hence $E|x_{nt}e_{nt}|^{2r} = E|x_{nt}|^{2r} E|e_{nt}|^{2r} \leq kK^r R^2$. Thus for any $\delta > 0$,

$$P\left(\frac{1}{\sqrt{n}} \max_{t \leq n} |x_{nt}e_{nt}| > \delta\right) \leq \sum_{t=1}^{n} P\left(\frac{1}{\sqrt{n}} |x_{nt}e_{nt}| > \delta\right) \leq \sum_{t=1}^{n} \frac{E|x_{nt}e_{nt}|^{2r}}{n^r \delta^{2r}} \leq \frac{kK^r R^2}{n^{r-1}\delta^{2r}} \to 0.$$

This completes the demonstration of $X_n^* \to_d X$ as $n \to \infty$, which completes the proof. □

Proof of Theorem 2: For simplicity of exposition, we will discuss the proof for the case where there is no included constant and time trend. In this case, under our assumptions, the asymptotic distributions of $n(\hat{\rho}_1 - \rho_1)$ and t_1 are given by (10) and (11), respectively, where $W_c(r)$ is the Gaussian diffusion process which solves $dW_c(r) = cW_c(r) + dW(r)$.

Let $\alpha = \rho_1 = 1 + C/n$, $\eta = (\rho_2, ..., \rho_k, P) \in \left(R^{k-1} \times \Xi\right)$ and set

$$d(\eta, \eta') = \sum_{j=2}^{k} \left|\rho_j - \rho_j'\right| + d_{2r}(P, P').$$

It is clear that $d(\eta, \hat{\eta}(\alpha)) \to_p 0$.

Take any sequence $\eta_n \in \left(R^{k-1} \times \Xi\right)$ such that $d(\eta, \eta_n) \to 0$. Let y_{nt} be determined by (α, η_n). That is, y_{nt} follows the process

$$y_{nt} = (1 + C/n)\, y_{nt-1} + \rho_{n2}\Delta y_{nt-1} + \cdots + \rho_{nk}\Delta y_{nt-k+1} + e_{nt}$$

where e_{nt} are iid mean-zero draws form P_n. Let $\hat{\rho}_{n1}$ denote the OLS estimate of ρ_1 from this sample and t_{n1} its t-statistic. To complete the proof, we need to show that the asymptotic distributions of $\hat{\rho}_{n1}$ and t_{n1} are (10) and (11), respectively.

As discussed in the previous section, for n sufficiently large, $Ee_{nt}^{2r} \leq R$, and e_{nt}^2 is uniformly integrable. Since $\frac{1}{n}\sum_{t=1}^{[nr]} Ee_{nt}^2 \to r\sigma^2$ for all r, $\frac{1}{n}\sum_{t=1}^{n} e_{nt}^2 \to_p \sigma^2$, and $n^{-1/2} \max_{t \leq n} |e_{nt}| \to_p 0$, (by arguments similar to (18) and (19)) it follows by Donsker's theorem for martingale difference arrays (Davidson, 1994, Theorem 27.14) that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_{nt} \Rightarrow \sigma W(r) \tag{20}$$

as $n \to \infty$. Let $S_{nt} = (1 + c/n)S_{nt-1} + e_{nt}$. Then (20) and Hansen (1992, Theorem 3.1) imply

$$\begin{aligned} \frac{1}{\sqrt{n}} S_{n[nr]} &= \exp\left(-\frac{[nr]}{n}c\right) \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \exp\left(\left(\frac{t}{n}\right)c\right) e_{nt} + o_p(1) \\ &\Rightarrow \exp(-rc) \int_0^r \exp(\lambda c)\, \sigma dW(\lambda) \\ &= \sigma W_c(r). \end{aligned}$$

Let $\overline{\rho}_n(z) = 1 - \rho_{n2}z - \cdots - \rho_{nk}z^{k-1}$. For sufficiently large n, $\overline{\rho}_n(z) = \overline{\rho}_n + \rho_n^*(z)(1-z)$ where $\overline{\rho}_n = \overline{\rho}_n(1)$ and the coefficients of $\rho_n^*(z)$ have exponential decay. Let $\xi_{nt} = \rho_n^*(L)\Delta y_{nt}$ and $S_{nt}^* = \overline{\rho}_n(L)y_{nt}$. Observe that

$$\Delta S_{nt}^* = \frac{c}{n}S_{nt-1}^* + \frac{c}{n}\xi_{nt-1} + e_{nt}.$$

Since $n^{-1/2}\sup_{t\leq n}|\xi_{nt}| \to_p 0$, then $n^{-1/2}S_{n[nr]}^* = n^{-1/2}S_{n[nr]} + o_p(1) \Rightarrow \sigma W_c(r)$. Now let $\overline{b}_n(z) = \overline{\rho}_n(z)^{-1}$ so $\overline{b} = 1/\overline{\rho}$. Then

$$\frac{1}{\sqrt{n}}y_{n[nr]} = \overline{b}_n(L)S_{n[nr]}^* = \overline{b}_n S_{n[nr]}^* + o_p(1) \Rightarrow \overline{b}\sigma W_c(r) = \frac{\sigma}{\overline{\rho}}W_c(r).$$

By the continuous mapping theorem we conclude that

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{nt-1}^2 \Rightarrow \frac{\sigma^2}{\overline{\rho}^2}\int_0^1 W_c^2 \tag{21}$$

and by Hansen (1992, Theorem 3.1)

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{nt-1}e_{nt} \Rightarrow \frac{\sigma^2}{\overline{\rho}}\int_0^1 W_c dW. \tag{22}$$

Results (21) and (22) are the fundamental determinants of the asymptotic distributions of $\hat{\rho}_{n1}$ and t_{n1} (10) and (11). The only addition equations needed to establish is (15), and that for $j = 1, ..., k$,

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} y_{nt-1}\Delta y_{nt-j} \to_p 0,$$

and both follow from standard arguments. We conclude that the asymptotic distributions of $\hat{\rho}_{n1}$ and t_{n1} are (10) and (11), respectively, which completes the proof. □

References

[1] Abadir, K.M. (1993): "The limiting distribution of the autocorrelation coefficient under a unit root," *The Annals of Statistics*, 21, 1058–1070.

[2] Anderson, T.W. and H. Rubin (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *The Annals of Mathematical Statistics*, 20, 46-63.

[3] Andrews, D.W.K. (1993): "Exactly median-unbiased estimation of first order autoregressive/unit root models," *Econometrica,* 61, 139-166

[4] Andrews, D.W.K. and H.-Y. Chen (1994): "Approximately median-unbiased estimation of autoregressive models," *Journal of Business and Economic Statistics*, 12, 187-204.

[5] Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991): "Bootstrapping unstable first-order autoregressive processes," *Annals of Statistics*, 19, 1098-1101.

[6] Beran, R. (1987): "Prepivoting to reduce level error of confidence sets," *Biometrika*, 74, 457-468.

[7] Bickel, P.J. and D.A. Freedman (1981): "Some asymptotic theory for the bootstrap," *Annals of Statistics*, 9, 1196-1217.

[8] Bose, A. (1988): "Edgeworth correction by bootstrap in autoregressions," *Annals of Statistics*, 16, 1709-1722.

[9] Carpenter, J. (1999): "Test inversion bootstrap confidence intervals," *Journal of the Royal Statistical Society, Series B*, 61, 159-172.

[10] Chamberlain, G. (1994): "Quantile regression, censoring, and the structure of wages," in *Advances in Econometrics: Sixth World Congress*, ed. C. Sims, 171-209.

[11] Cramér, H. (1946): *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

[12] Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.

[13] DiCiccio, T.J. and J.P. Romano (1990): "Nonparametric confidence limits by resampling methods and least favorable families," *International Statistical Review*, 58, 59-76.

[14] Diebold, F.X. and A.S. Senhadji (1996): "The uncertain unit root in real GNP: Comment," *American Economic Review*, 86, 1291-1298.

[15] Dufour, J.M. (1997): "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica*, 65, 1365-1387.

[16] Efron, B. and R.J. Tibshirani (1993): *An Introduction to the Bootstrap*, New York: Chapman-Hall.

[17] Fuller, W. A. (1996): *Introduction to Statistical Time Series*, 2nd Edition. New York: Wiley.

[18] Garthwaite, P.H. and S.T. Buckland (1992): "Generating Monte Carlo confidence intervals by the Robbins-Monro process," *Applied Statistician*, 41, 159-171.

[19] Hall, P. (1988): "Theoretical comparison of bootstrap confidence intervals" (with discussion), *Annals of Statistics*, 16, 927-985.

[20] Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

[21] Hansen, B.E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes," Econometric Theory, 7, 489-500.

[22] Kabaila, P. (1993): "Some properties of profile bootstrap confidence intervals," *Australian Journal of Statistics*, 35, 205-214.

[23] Kilian, L. (1998): "Small-sample confidence intervals for impulse response functions," *Review of Economics and Statistics*, 80, 218-230.

[24] Knight, J.L. and S.E. Satchell (1993): "Asymptotic expansions for random walks with normal errors," *Econometric Theory,* 10, 363-376.

[25] Mallows, C.L. (1972): "A note on asymptotic joint normality," *Annals of Mathematical Statistics*, 39, 755-771.

[26] Nankervis, J.C. and N.E. Savin (1996): "The level and power of the bootstrap t test in the AR(1) model with trend," *Journal of Business and Economic Statistics*, 14, 161-168.

[27] Nelson, C.R. and C.I. Plosser (1982): "Trends and random walks in macroeconomic time series," *Journal of Monetary Economics*, 10, 139-162.

[28] Phillips, P.C.B. (1977): "Approximations to some finite sample distributions associated with a first-order stochastic difference equation," *Econometrica*, 45, 463-485.

[29] Phillips, P.C.B. and V. Solo (1992): "Asymptotics for linear processes," *The Annals of Statistics*, 20, 971-1001.

[30] Romano, J.P. and M. Wolf (1998): "Subsampling confidence intervals for the autoregressive root," unpulished manuscript, Department of Statistics, Stanford University.

[31] Runkle, D.E. (1985): "Vector autoregressions and reality," *Journal of Business and Economic Statistics*, 5, 437-454.

[32] Schotman, P.C. and H.K. van Dijk (1991): "On Bayesian routes to unit roots," *Journal of Applied Econometrics*, 6, 387-401.

[33] Shao, J. and D. Tu (1995): *The Jackknife and Bootstrap*. NY: Springer.

[34] Stock, J.H. (1991): "Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series," *Journal of Monetary Economics,* 28, 435-460.