Bootstrapping Autoregressive Processes with Possible Unit Roots

Atsushi Inoue*
North Carolina State University

Lutz Kilian†
University of Michigan
and CEPR

November 12, 1999

Abstract

An important question in applied work is how to bootstrap autoregressive processes involving highly persistent time series of unknown order of integration. In this paper, we show that in many cases of interest in applied work the standard bootstrap algorithm for unrestricted autoregressions remains valid for processes with exact unit roots; no pre-tests are required, at least asymptotically, and applied researchers may proceed as in the stationary case. Specifically, we prove the first-order asymptotic validity of bootstrapping any linear combination of the slope parameters in autoregressive models with drift. We also establish the bootstrap validity for the marginal distribution of slope parameters and for most linear combinations of slope parameters in higher-order autoregressions without drift. The latter result is in sharp contrast to the well-known bootstrap invalidity result for the random walk without drift. A simulation study examines the finite-sample accuracy of the bootstrap approximation both for integrated and for near-integrated processes. We find that in many, but not all circumstances, the bootstrap distribution closely approximates the exact finite-sample distribution.

KEYWORDS: Autoregressions, bootstrap asymptotics, integrated processes, resampling.

1 Introduction

An important question in applied work is how to bootstrap autoregressive processes involving highly persistent time series of unknown order of integration. Applied researchers often are interested in conducting inference on autoregressive slope parameters (or smooth functions thereof such as impulse responses or half-lives). The asymptotic validity of the standard residual-based bootstrap algorithm for autoregressive slope parameters in stationary models follows from Bose (1988). Datta (1995) extends the validity of this bootstrap algorithm to univariate AR(p) models with some explosive roots (also see Basawa et al. (1989) for a discussion of the explosive AR(1) model). However, he explicitly rules out the knife-edge case of an exact unit root. There are no theoretical results to our knowledge that would justify the use of the standard bootstrap algorithm for processes with exact unit roots.

The main contribution of this paper is to show that in many cases of interest in applied work the standard bootstrap algorithm for unrestricted autoregressions remains valid for processes with exact unit roots. A direct implication of our results is that the standard bootstrap algorithm will often provide an asymptotically valid approximation to the true distribution of the autoregressive slope parameters, whether the population process is integrated or not. This fact is important for applied work because it alleviates, at least asymptotically, the need for unit root pre-tests.

The possibility of exact unit roots has received considerable attention in time series econometrics. While any univariate autoregressive process with an exact unit root can be converted into a stationary process by first differencing, the existence of a unit root is rarely known from economic theory. In practice, applied researchers often decide to difference the data based on the outcome of unit root pre-tests. There is an extensive literature documenting the potential pitfalls in basing econometric inference on the outcome of such tests (see Cochrane 1991; Blough 1992; Faust 1996; also see Elliott 1998). This fact calls into question bootstrap inference based on unit-root pre-tests. Our results show that in many cases applied users of the bootstrap may proceed as though the process were known to be level-stationary. There is no need to be concerned about the possible existence of a unit root.

Our approach builds on and extends existing theoretical results on bootstrapping unrestricted autoregressions with exact unit roots. In a well-known paper, Basawa et al. (1991a) establish the asymptotic invalidity for the random walk model without drift of the residual-based bootstrap algorithm for autoregressive slope parameters based on unrestricted autoregressions. Their original result is based on a very special parametric resampling scheme which one would like to apply only when the error distribution of the autoregressive process is known to be standard normal. Datta (1996) generalizes Basawa et al.'s result and proves the invalidity of the standard nonparametric bootstrap

algorithm of Bose (1988) for the same model. The theoretical validity of the standard bootstrap algorithm for other autoregressions with exact unit roots (such as random walks with drift or higher-order autoregressions with or without drift) has not been examined to date.

In this paper, we show that the Basawa et al. (1991a) bootstrap invalidity result is a very special case only. We establish that for many other autoregressive processes the standard bootstrap algorithm is asymptotically valid even in the presence of a unit root. Specifically, we prove the first-order asymptotic validity of bootstrapping any linear combination of slope parameters in autoregressive models with drift. Our results include the random walk model with drift as a special case. In addition, we establish the first-order bootstrap validity for the marginal distribution of slope parameters and for most linear combinations of slope parameters in higher-order autoregressions without drift. Our results apply not only when bootstrapping models with known zero intercept, but also in the practically more important case of unknown and possibly nonzero intercept.

The cases for which we establish the validity of the bootstrap are cases for which the asymptotic normality of the OLS estimator in the local-to-unity model follows from Jeganathan (1991) or for which asymptotic normality may be established by similar arguments. Jeganathan's local-to-unity model implies that the limiting distribution of the OLS estimator of the autoregressive model without drift depends on nuisance parameters. Similar arguments may be applied to the bootstrap estimator with the added complication that in the bootstrap world the nuisance parameters will be random. This means that the standard arguments for the asymptotic validity of the bootstrap cannot be applied.

The main technical innovation of this paper is to derive the limit distribution of the bootstrap estimator in these circumstances. We show that in models without drift the bootstrap estimator of the sum of the autoregressive coefficients converges to a random limiting distribution. In contrast, in models with drift the limit distribution of the bootstrap estimator of this sum is shown to be Gaussian. In fact, the bootstrap estimator converges to the same normal limit distribution as the usual OLS estimator. We also show that, both for models with and without drift, the bootstrap estimator of coefficients on lagged differences in the augmented Dickey-Fuller representation will converge to the same Gaussian limit distribution as the OLS estimator.

Our results imply that the standard bootstrap is asymptotically valid for the slope parameter in the random walk model with drift. In contrast, in the random walk model without drift, the limit distribution of the OLS estimator is nonstandard and the bootstrap estimator converges to a random distribution. Paraphrasing Basawa et al. (1991b, p.1016): "The main cause of the failure of the standard bootstrap estimator at the unit circle in the AR(1) model is the fact that the limit distributions of the slope parameter are drastically different for the stationary, explosive and unit root case; the

unit root serving as a threshold. This introduces instability in the bootstrap samples, which in turn leads to the eventual invalidity of the standard bootstrap estimator at the unit circle." This quote suggests that the bootstrap validity in the random walk model with drift is not obvious.

It is well known of course that the OLS estimator of the slope parameter of the random walk model with drift has a Gaussian limit distribution (see e.g., West 1988). Asymptotic normality, however, is neither necessary nor has it been shown to be sufficient for establishing the asymptotic validity of the bootstrap. Note that the asymptotic distribution of the slope parameter in the random walk model with drift exhibits a discontinuity at the unit circle not unlike that referred to by Basawa et al. This discontinuity occurs because the slope parameter in the AR(1) model with nonzero intercept has a different variance and even different rate of convergence depending on whether there is a unit root or not. The resulting discontinuity of the asymptotic distribution at the unit circle may seem to suggest that the bootstrap will fail just as in the random walk model without drift. We prove that this is not the case.

We also study higher-order autoregressive models with and without drift. Sims, Stock and Watson (1990) establish the asymptotic normality of the OLS estimator of the slope parameters of the level representation of these models. We show that the bootstrap estimator of these parameters converges to the same normal limit distribution as the usual OLS estimator. Thus, the standard bootstrap approximation remains asymptotically valid even in the presence of a unit root.

This result is nontrivial because the bootstrap estimator in integrated processes depends on nuisance parameters. We show that in higher-order processes these nuisance parameters do not affect the asymptotic distribution of the bootstrap estimator of the level slope parameters. The intuition for this result is as follows. Recall that the bootstrap estimator of the slope parameters of higher-order autoregressions can be expressed in terms of the sum of the autoregressive coefficients on the one hand and of coefficients on lagged differences on the other. We show that the bootstrap estimator of the latter coefficients is always $\sqrt{T}$ consistent and Gaussian. Although in models without drift the bootstrap estimator of the sum of the autoregressive coefficients converges to a random distribution, it converges faster than the remaining terms and is asymptotically negligible. This fact ensures the $\sqrt{T}$ consistency and asymptotic normality of the bootstrap estimator of the autoregressive slope parameters. Similarly, in higher-order models with drift, the bootstrap validity follows from the fact that the sum of the autoregressive coefficients converges at rate $T^{3/2}$ compared with $T^{1/2}$ for the remaining terms.

Our results do not only apply to slope parameters, but they also provide the basis for inference on linear and smooth nonlinear functions of slope parameters (such as impulse responses), provided the limit distribution is nondegenerate. Like Sims, Stock

and Watson (1990), however, we find that for certain "forbidden" linear combinations the validity of the normal approximation and of the bootstrap approximation breaks down. For the latter case, several modified bootstrap algorithm of varying degrees of generality and finite-sample accuracy have been proposed by Basawa et al. (1991b), Datta (1996), Heimann and Kreiss (1996), Datta and Sriram (1997) and Romano and Wolf (1998). Related work also includes Hansen (1999).

Our theoretical results imply that for inference on autoregressive slope parameters in many cases no pre-tests are required, at least asymptotically, and applied researchers may proceed as in the stationary case. This result is important for applied work because in most empirical studies interest centers on smooth functions of autoregressive slope parameters, and the existence of a unit root is of no direct interest. An important question is how relevant these asymptotic results are for sample sizes of interest in practice. In a preliminary Monte Carlo study, we examine the finite-sample accuracy of the bootstrap estimator of the autoregressive slope parameters both for integrated and for near-integrated processes. As in our theoretical discussion, we carefully distinguish regression models with and without intercept and population models with and without drift. We find that in many, but not all circumstances, the bootstrap distribution closely approximates the exact finite sample distribution. Thus, from a practical point of view the question of whether there is a unit root or not becomes much less important for conducting bootstrap inference.

The remainder of the paper is organized as follows. Section 2 contains the main results. The Monte Carlo simulation evidence is discussed in section 3. The concluding remarks are in Section 4. The proofs are relegated in the appendix.

2   Bootstrap Asymptotic Theory

The properties of the autoregressive bootstrap estimator depend critically on the properties of the population model as well as on the regression model. We begin with a review of the standard residual-based bootstrap procedure for autoregressions. Let $(\hat{\phi}_1, ..., \hat{\phi}_p)'$ denote the ordinary least squares (OLS) estimator of $(\phi_1, ..., \phi_p)'$ in the scalar AR(p) process $\phi(L)y_t = \delta + \varepsilon_t$ where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$. Let $(\hat{\phi}_1^*, ..., \hat{\phi}_p^*)'$ denote the corresponding bootstrap OLS estimator of $(\phi_1, ..., \phi_p)'$. We distinguish two closely related bootstrap algorithms referred to as Algorithm 1 and Algorithm 2.

Algorithm 1: Algorithm 1 is designed for the autoregressive model with known intercept $\delta = 0$. Let $\hat{\varepsilon}_t = \hat{\phi}(L)y_t$ denote the residuals of the OLS estimate of process (1) given $\{y_t\}_{t=1-p}^T$. Define the centered residuals $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t$. Let $\tilde{F}_T$ denote the empirical distribution function of $\tilde{\varepsilon}_t$. $\tilde{F}_T$ associates probability mass T^{-1}

4

with $\tilde{\varepsilon}_t$, $t = 1, ..., T$. Treating $\tilde{F}_T$ as the bootstrap population distribution, random samples $\{\varepsilon_t^*\}_{t=1}^T$ may be drawn from $\tilde{F}_T$. Thus, conditional on the data, the random variable ε_t^* is iid with distribution function $\tilde{F}_T$. Now construct the bootstrap sample $\{y_t^*\}_{t=1-p}^T$ recursively from $\hat{\phi}(L)y_t^* = \varepsilon_t^*$, given initial values $y_0^* = y_0, ..., y_{1-p}^* = y_{1-p}$. The bootstrap estimator $(\hat{\phi}_1^*, ..., \hat{\phi}_p^*)$ then may be obtained by OLS from $\{y_t^*\}_{t=1-p}^T$.

Algorithm 2: Algorithm 2 is designed for models with unknown and possibly nonzero intercept δ. In that case, $\hat{\varepsilon}_t = \hat{\phi}(L)y_t - \hat{\delta}$ where $\hat{\delta}$ is the OLS estimate of δ and $\hat{\varepsilon}_t$ is mean zero by construction. Thus, conditional on the data, the random variable ε_t^* is iid with distribution function $\hat{F}_T$ and bootstrap data may be generated recursively from $\hat{\phi}(L)y_t^* = \hat{\delta} + \varepsilon_t^*$ as described previously.

In our theoretical analysis, we distinguish population models with drift and without drift and regression models with and without intercept. In Theorem 1, the population model is a higher-order autoregression without drift, and the regression model does not include an intercept. The case of higher-order models with known zero drift is primarily of theoretical interest because it extends the Basawa et al./Datta analysis of the AR(1) process. The regression model with intercept is dealt with in Theorems 2 and 3. Theorems 2 and 3 make different assumptions about the population model. Theorem 2 covers both first-order and higher-order autoregressive models with drift. This theorem applies to trending data such as consumption or output. In Theorem 3, we analyze autoregressive models without drift. The latter assumption is most relevant for nontrending data such as interest rates, inflation rates or real exchange rates, for which the possible existence of a unit root is often a practical concern.

2.1 Higher-Order Autoregressions without Drift when the Regression Model does not Include an Intercept

Before stating the main results, we establish some notation. Let $\overset{L}{\to}$ denote convergence in law and $\overset{d}{\to}$ denote convergence in distribution. Let ω-a.s. denote weak convergence almost surely conditional on the sample (see Giné and Zinn, 1990). $0_{q \times 1}$ denotes a q-dimensional vector of zeros. I_p denotes the p-dimensional identity matrix.

Consider the following AR(p) process

$$\phi(L)y_t = \varepsilon_t, \tag{1}$$

where $p > 1$, $y_0 = y_1 = \cdots = y_{1-p} = 0$ and $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$ with $E|\varepsilon_t|^{2+\kappa} < \infty$ for some $\kappa > 0$ and F satisfying a Lipschitz condition, i.e., there is $M > 0$ such that $|F(x') - F(x)| < M|x' - x|$ for all x and x'. Suppose that $\phi(z) = (1 - L)\psi(z)$ with $|\psi(z)| \neq 0$ for all $|z| \leq 1$ such that the process is integrated of order 1.

The process y_t can be written as

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \\
&= \beta' x_t + \varepsilon_t,
\end{aligned}
\tag{2}
$$

where

$$
\begin{aligned}
\rho &= \phi_1 + \phi_2 + \cdots + \phi_p = 1 \\
\zeta_j &= -[\phi_{j+1} + \phi_{j+2} + \cdots + \phi_p] \ \text{for } j = 1, 2, ..., p-1 \\
\beta &= (\rho, \zeta_1, \zeta_2, ..., \zeta_{p-1})' \\
x_t &= (y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-p+1})'.
\end{aligned}
$$

Let $\hat{\beta}$ denote the OLS estimator of β. We know from Hasza and Fuller (1979) that under the conditions stated

$$
\Upsilon_T(\hat{\beta} - \beta) \xrightarrow{L} \tilde{\gamma} = \left[\begin{array}{c} \dfrac{\psi(1) \int_0^1 B(r) dB(r)}{\int_0^1 B(r)^2 dr} \\ \sigma \Gamma^{-\frac{1}{2}} W_{p-1} \end{array} \right],
\tag{3}
$$

where $\Upsilon_T = \text{diag}(T, T^{1/2}, ..., T^{1/2})$, B is a Brownian motion, W_{p-1} is a $(p-1)$-dimensional standard normal random vector, B and W_{p-1} are independent,

$$
\Gamma = \left[\begin{array}{cccc} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{array} \right],
$$

and γ_j is the jth autocovariance of Δy_t. Let $\hat{\beta}^*$ denote the bootstrap OLS estimator of $\hat{\beta}$ based on Algorithm 1.

As noted earlier, the standard bootstrap results for autoregressions such as Bose (1988) and Datta (1995) do not cover this model. Theorem 1 uses local asymptotics to prove that the standard bootstrap is valid for the non-unit-root parameters ζ_j, $j = 1, ..., p-1$, in the ADF representation (2). We also show that the standard bootstrap fails for ρ. Corollary 1 establishes that the bootstrap invalidity for ρ is irrelevant if we are interested in inference on smooth functions of both ρ and ζ_j, $j = 1, ..., p-1$, such as slope parameters of the level representation (1) or smooth functions thereof. Note that we do not claim to have solved the unit root problem for ρ. Rather we show that the bootstrap invalidity for ρ does not matter for many statistics of interest in applied work.

Theorem 1. (Asymptotic Properties of the Bootstrap in Integrated AR(p) Processes without Drift)

Under the stated assumptions about process (1),

$$\Upsilon_T(\hat{\beta}^* - \hat{\beta}) \overset{L}{\to} \left[\begin{array}{c} \dfrac{\psi(1)\int_0^1 S(r,\tilde{\gamma}_0)dB(r)}{\int_0^1 S(r,\tilde{\gamma}_0)^2 dr} \\[1em] \sigma\Gamma^{-\frac{1}{2}}W_{p-1} \end{array} \right] \quad \omega\text{-a.s.} \tag{4}$$

where

$$S(r,\gamma_0) = B(r) + \psi(1)\gamma_0 \int_0^r \exp\left[(r-s)\psi(1)\gamma_0\right]B(s)ds, \tag{5}$$

and $\tilde{\gamma}_0$ denotes the first element of the random vector $\tilde{\gamma}$ on the RHS of (3).

Theorem 1 generalizes the results of Basawa *et al.* (1991a) and Datta (1996) to higher-order autoregressions without drift. Note that the bootstrap fails to mimic the limiting distribution of the OLS estimator for ρ, but it recovers the limiting distribution of the OLS estimator for the non-unit-root parameters $(\zeta_1, \zeta_2, ..., \zeta_{p-1})'$. The latter fact has important implications for the distribution of the slope parameters in the level representation of the autoregressive process. The slope parameters of the level representation (1) can be expressed as linear combinations of ρ and of ζ_i, $i = 1, 2..., p-1$. Specifically, $\phi_1 = \rho + \xi_1$, $\phi_j = \zeta_j - \zeta_{j-1}$ for $j = 2, ..., p-1$ and $\phi_p = -\zeta_{p-1}$. Although the bootstrap estimator of ρ converges to a random limit distribution, it does so at a rate so fast, that any linear combination of bootstrap estimators involving coefficients on lagged differences will be $\sqrt{T}$ consistent and will converge to the usual Gaussian limit distribution. Hence, the bootstrap provides an asymptotically valid approximation to the marginal distribution of the autoregressive slope parameters, even in the presence of a unit root. Note that the variance of the bootstrap estimator is continuous in ρ and thus there is no discontinuity at the unit circle, unlike in the random walk model.

This result is of considerable importance for applied work. Applied researchers typically are not interested in inference about ρ or about ζ_i, $i = 1, 2, ..., p-1$. Rather they are interested in inference on linear and smooth nonlinear functions of the autoregressive slope parameters (such as impulse responses and half-lives). For this class of problems, it is the distribution of the autoregressive slope parameters in the level representation (1) that matters rather than the distribution of the parameters of the augmented Dickey-Fuller representation (2). Hence, the invalidity of the bootstrap estimator of ρ is irrelevant. Theorem 1 implies that the standard bootstrap approximation to the marginal distribution of the autoregressive slope parameters in (1) will be as sound asymptotically as the usual Gaussian approximation. Corollary 1 considers the implications of Theorem 1 for the distribution of linear combinations of the slope parameters in the level representation of the autoregressive process.

Corollary 1. (Validity of the Bootstrap for Slope Parameters in Integrated

AR(p) Processes without Drift)

Consider a linear combination of slope parameters $c'\phi$ where $c = (c_1, c_2, ..., c_p)' \neq (\lambda, \lambda, ..., \lambda)'$ for all λ and $\phi = (\phi_1, \phi_2, ..., \phi_p)'$. Let $\hat{\phi}$ and $\hat{\phi}^*$ denote the OLS estimator and the bootstrap estimator, respectively. Suppose that the assumptions of Theorem 1 hold. Then

$$T^{1/2}c'(\hat{\phi} - \phi) \xrightarrow{d} N(0, c'\Omega c) \tag{6}$$

$$T^{1/2}c'(\hat{\phi}^* - \hat{\phi}) \xrightarrow{d} N(0, c'\Omega c) \quad \omega\text{-a.s.} \tag{7}$$

where

$$\Omega = D \begin{bmatrix} 0 & 0_{1 \times (p-1)} \\ 0_{(p-1) \times 1} & \sigma^2 \Gamma^{-1} \end{bmatrix} D',$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

The assumption $p > 1$ is crucial for Corollary 1. For $p = 1$, $\phi_1 = \rho$ will have a nonstandard limiting distribution and the bootstrap will not be valid (see Datta, 1996). Specifically, $T(\hat{\rho}^* - \hat{\rho})$ will converge to a random distribution.

Theorem 1 and Corollary 1 validate the application of the standard nonparametric bootstrap to individual slope parameters, ϕ_j, for $j = 1, 2, ..., p$, and to linear combinations of slope parameters except to those proportional to $\rho = \phi_1 + \cdots + \phi_p$. When $c = (\lambda, \lambda, ..., \lambda)'$ for some $\lambda \neq 0$, $c'\phi = \lambda\rho$ and thus the bootstrap will be invalid. This result parallels the results for the usual OLS estimator by Sims, Stock and Watson (1990). Our results also provide the basis for bootstrap inference on smooth nonlinear functions of $(\rho, \zeta_1, ..., \zeta_{p-1})'$ such as impulse responses, provided the limiting distribution is nondegenerate.

It is important to stress that our result pertains to the case of integrated processes. Note that a different limiting distribution is obtained in the stationary case. Nevertheless, the bootstrap is valid in both cases. This feature suggests that applied users need not be concerned about the possible existence of a unit root, at least asymptotically.

2.2 Autoregressions with Drift when the Regression Model Includes an Intercept

Having established the conditions for the asymptotic validity of the bootstrap in the integrated $AR(p)$ model without drift, we now turn to models with drift. The regression model is assumed to include an intercept. This is not the only regression model one may wish to consider for this process. If the true order of integration is unknown, a researcher may choose to include a linear time trend in the regression as well. A discussion of the regression model with both intercept and linear time trend is deferred to the end of section 2.3.

Consider the following $AR(p)$ process

$$\phi(L)y_t = \delta + \varepsilon_t, \tag{8}$$

where $p \geq 1$, $y_0 = y_1 = \cdots = y_{1-p} = 0$, and $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$ with $E|\varepsilon_t|^{2+\kappa} < \infty$ for some $\kappa > 0$ and F satisfying a Lipschitz condition. Suppose that $\phi(z) = (1 - L)\psi(z)$ with $|\psi(z)| \neq 0$ for all $|z| \leq 1$. Write (8) as

$$
\begin{aligned}
y_t &= \frac{\delta}{\psi(1)} + \rho y_{t-1} + \zeta_1 \left(\Delta y_{t-1} - \frac{\delta}{\psi(1)} \right) + \cdots + \zeta_{p-1} \left(\Delta y_{t-p+1} - \frac{\delta}{\psi(1)} \right) + \varepsilon_t \\
&= \bar{\beta}' \bar{x}_t + \varepsilon_t,
\end{aligned}
$$

where

$$
\begin{aligned}
\rho &= \phi_1 + \phi_2 + \cdots + \phi_p, \\
\zeta_j &= -[\phi_{j+1} + \phi_{j+2} + \cdots + \phi_p] \text{ for } j = 1, 2, ..., p-1, \\
\bar{\beta} &= \begin{cases} (\delta, \rho)' & \text{if } p = 1 \\ (\delta/\psi(1), \rho, \zeta_1, \zeta_2, ..., \zeta_{p-1})' & \text{if } p > 1, \end{cases} \\
\bar{x}_t &= \begin{cases} (1, y_{t-1})' & \text{if } p = 1 \\ (1, y_{t-1}, \Delta y_{t-1} - \delta/\psi(1), ..., \Delta y_{t-p+1} - \delta/\psi(1))' & \text{if } p > 1. \end{cases}
\end{aligned}
$$

Let $\hat{\bar{\beta}}$ denote the OLS estimator of $\bar{\beta}$. We know from Dickey and Fuller (1981) that for $p = 1$

$$\bar{\Upsilon}_T(\hat{\bar{\beta}} - \bar{\beta}) \overset{d}{\to} N \left(0_{2 \times 1}, \sigma^2 \begin{bmatrix} 1 & \frac{\delta}{2} \\ \frac{\delta}{2} & \frac{\delta^2}{3} \end{bmatrix}^{-1} \right) \tag{9}$$

and that for $p > 1$

$$\bar{\Upsilon}_T(\hat{\bar{\beta}} - \bar{\beta}) \overset{d}{\to} N \left(0_{(p+1) \times 1}, \sigma^2 \begin{bmatrix} 1 & \frac{\delta}{2\psi(1)} & 0_{1 \times (p-1)} \\ \frac{\delta}{2\psi(1)} & \frac{\delta^2}{3\psi(1)^2} & 0_{1 \times (p-1)} \\ 0_{(p-1) \times 1} & 0_{(p-1) \times 1} & \Gamma \end{bmatrix}^{-1} \right), \tag{10}$$

9

where

$$
\bar{\Upsilon}_T = \begin{cases} \operatorname{diag}(T^{1/2}, T^{3/2}) & \text{if } p = 1, \\ \operatorname{diag}(T^{1/2}, T^{3/2}, T^{1/2}, ..., T^{1/2}) & \text{if } p > 1, \end{cases}
$$

$$
\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix},
$$

and γ_j is the jth autocovariance of $\Delta y_t - \delta/\psi(1)$. Let $\hat{\bar{\beta}}^*$ denote the bootstrap OLS estimator of $\hat{\bar{\beta}}$ based on Algorithm 2.

Theorem 2. (Asymptotic Properties of the Bootstrap in Integrated AR(p) Processes with Drift)

Under the stated assumptions about process (8) for $p = 1$

$$
\bar{\Upsilon}_T(\hat{\bar{\beta}}^* - \hat{\bar{\beta}}) \xrightarrow{d} N\left(0_{2\times1}, \sigma^2 \begin{bmatrix} 1 & \frac{\delta}{2} \\ \frac{\delta}{2} & \frac{\delta^2}{3} \end{bmatrix}^{-1}\right) \quad \omega\text{-a.s.} \tag{11}
$$

and for $p > 1$

$$
\bar{\Upsilon}_T(\hat{\bar{\beta}}^* - \hat{\bar{\beta}}) \xrightarrow{d} N\left(0_{(p+1)\times1}, \sigma^2 \begin{bmatrix} 1 & \frac{\delta}{2\psi(1)} & 0_{1\times(p-1)} \\ \frac{\delta}{2\psi(1)} & \frac{\delta^2}{3\psi(1)^2} & 0_{1\times(p-1)} \\ 0_{(p-1)\times1} & 0_{(p-1)\times1} & \Gamma \end{bmatrix}^{-1}\right) \quad \omega\text{-a.s.} \tag{12}
$$

Note that the results for the model with drift are stronger than for the model without drift. Unlike in Theorem 1, the bootstrap is valid not only for $(\zeta_1, ..., \zeta_{p-1})'$ but also for ρ.

Corollary 2. (Validity of the Bootstrap for the Slope Parameters in Integrated AR(p) Processes with Drift)

Suppose that the assumptions of Theorem 2 hold. Consider a linear combination of slope parameters $c'\phi$ where $\phi = (\phi_1, \phi_2, ..., \phi_p)'$, and $\bar{c} = (\bar{c}_1, \bar{c}_2, ..., \bar{c}_p)' \neq 0_{p\times1}$. Let $\hat{\phi}$ and $\hat{\phi}^*$ denote the OLS estimator and the bootstrap OLS estimator, respectively.

1. If $(\bar{c}_1, \bar{c}_2, ..., \bar{c}_p)' = (\lambda, \lambda, ..., \lambda)'$ for some $\lambda \neq 0$, then

$$
T^{3/2}\bar{c}'(\hat{\phi} - \phi) \xrightarrow{d} N\left(0, \frac{12\lambda^2\sigma^2}{\delta^2}\right)
$$

10

$$T^{3/2}\bar{c}'(\hat{\phi}^* - \hat{\phi}) \quad \overset{d}{\to} \quad N\left(0, \frac{12\lambda^2\sigma^2}{\delta^2}\right) \quad \omega\text{-a.s.}$$

2. If $(\bar{c}_1, \bar{c}_2, ..., \bar{c}_p)' \neq (\lambda, \lambda, ..., \lambda)'$ for all λ, then

$$T^{1/2}\bar{c}'(\hat{\phi} - \phi) \quad \overset{d}{\to} \quad N(0, \sigma^2\bar{c}'\bar{\Omega}\bar{c}) \tag{13}$$

$$T^{1/2}\bar{c}'(\hat{\phi}^* - \hat{\phi}) \quad \overset{d}{\to} \quad N(0, \sigma^2\bar{c}'\bar{\Omega}\bar{c}) \quad \omega\text{-a.s.} \tag{14}$$

where

$$\bar{\Omega} \;=\; \sigma^2 \bar{D}\Gamma^{-1}\bar{D}',$$

$$\bar{D} \;=\; \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}.$$

There are two main differences between Corollary 1 and Corollary 2. First, while Corollary 1 requires $p > 1$, Corollary 2 does not. Thus, the bootstrap is valid for the random walk model with drift. This result is in marked contrast to the case of the random walk model without drift analyzed by Datta (1996). Second, unlike Corollary 1, Corollary 2 establishes the asymptotic validity of the bootstrap for any linear combination of $(\phi_1, \phi_2, ..., \phi_p)'$.

2.3   Autoregressions without Drift when the Regression Model Includes an Intercept

Lastly, consider the case in which an $AR(p)$ model with intercept is fitted to data generated from the integrated $AR(p)$ process without drift described by (1). Let

$$\bar{\beta} \;=\; \begin{cases} (c, \rho)' & \text{if } p = 1 \\ (c, \rho, \zeta_1, \zeta_2, ..., \zeta_{p-1})' & \text{if } p > 1, \end{cases}$$

$$\bar{x}_t \;=\; \begin{cases} (1, y_{t-1})' & \text{if } p = 1 \\ (1, y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-p+1})' & \text{if } p > 1. \end{cases}$$

Let $\hat{\bar{\beta}}$ denote the OLS estimator of $\bar{\beta}$. We know from Dickey and Fuller (1981) that for $p = 1$

$$\bar{\Upsilon}_T(\hat{\bar{\beta}} - \bar{\beta}) \overset{L}{\to} \begin{bmatrix} \sigma & 0 \\ 0 & \psi(1) \end{bmatrix} \begin{bmatrix} 1 & \int_0^1 B(r)dr \\ \int_0^1 B(r)dr & \int_0^1 B(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} B(1) \\ \int_0^1 B(r)dB(r) \end{bmatrix}, \tag{15}$$

11

and that for $p > 1$

$$\bar{\Upsilon}_T(\hat{\bar{\beta}} - \bar{\beta}) \xrightarrow{L} \left[\begin{array}{c} \left[\begin{array}{cc} \sigma & 0 \\ 0 & \psi(1) \end{array} \right] \left[\begin{array}{cc} 1 & \int_0^1 B(r)dr \\ \int_0^1 B(r)dr & \int_0^1 B(r)^2 dr \end{array} \right]^{-1} \left[\begin{array}{c} B(1) \\ \int_0^1 B(r)dB(r) \end{array} \right] \\ \sigma \Gamma^{-\frac{1}{2}} W_{p-1} \end{array} \right]. \quad (16)$$

Let $\hat{\bar{\beta}}^*$ denote the bootstrap OLS estimator of $\hat{\bar{\beta}}$ based on Algorithm 2.

Theorem 3. (Asymptotic Properties of the Bootstrap in Integrated AR(p) Processes without Drift)

Suppose that the assumptions of Theorem 1 hold with $p \geq 1$. For $p = 1$,

$$\bar{\Upsilon}_T(\hat{\bar{\beta}}^* - \hat{\bar{\beta}}) \xrightarrow{L} \left[\begin{array}{cc} \sigma & 0 \\ 0 & \psi(1) \end{array} \right] \left[\begin{array}{cc} 1 & \int_0^1 S(r,\tilde{\gamma}_0)dr \\ \int_0^1 S(r,\tilde{\gamma}_0)dr & \int_0^1 S(r,\tilde{\gamma}_0)^2 dr \end{array} \right]^{-1} \left[\begin{array}{c} S(1,\tilde{\gamma}_0) \\ \int_0^1 S(r,\tilde{\gamma}_0)dB(r) \end{array} \right] \quad \omega\text{-a.s.}$$
$$(17)$$

and for $p > 1$

$$\bar{\Upsilon}_T(\hat{\bar{\beta}}^* - \hat{\bar{\beta}}) \xrightarrow{L} \left[\begin{array}{c} \left[\begin{array}{cc} \sigma & 0 \\ 0 & \psi(1) \end{array} \right] \left[\begin{array}{cc} 1 & \int_0^1 S(r,\tilde{\gamma}_0)dr \\ \int_0^1 S(r,\tilde{\gamma}_0)dr & \int_0^1 S(r,\tilde{\gamma}_0)^2 dr \end{array} \right]^{-1} \left[\begin{array}{c} S(1,\tilde{\gamma}_0) \\ \int_0^1 S(r,\tilde{\gamma}_0)dB(r) \end{array} \right] \\ \sigma \Gamma^{-\frac{1}{2}} W_{p-1} \end{array} \right] \quad \omega\text{-a.s.}$$
$$(18)$$

Note that the results of Theorem 3 for the slope parameters in higher-order autoregressions parallel those in Theorem 1. Moreover, the bootstrap invalidity result of Datta (1996) for the random walk model without drift is preserved when the regression model includes an intercept.

Corollary 3. (Validity of the Bootstrap for Slope Parameters in Integrated AR(p) Processes without Drift)

Consider a linear combination of slope parameters $c'\phi$ where $c = (c_1, c_2, ..., c_p)' \neq (\lambda, \lambda, ..., \lambda)'$ for all λ and $\phi = (\phi_1, \phi_2, ..., \phi_p)'$. Let $\hat{\phi}$ and $\hat{\phi}^*$ denote the OLS estimator and the bootstrap estimator in Theorem 3, respectively. Suppose that the assumptions of Theorem 1 hold. Then

$$T^{1/2}c'(\hat{\phi} - \phi) \xrightarrow{d} N(0, c'\Omega c) \quad (19)$$

$$T^{1/2}c'(\hat{\phi}^* - \hat{\phi}) \xrightarrow{d} N(0, c'\Omega c) \quad \omega\text{-a.s.} \quad (20)$$

where Ω is defined in Corollary 1.

Corollary 3 shows that the inclusion of an intercept term does not affect the conclusions obtained earlier for the autoregressive model without drift under the assumption

12

of a known drift of zero.

So far we have deliberately postponed discussion of the population model with drift, when the regression model includes both an intercept and a linear time trend. In that case, the drift no longer dominates the unit root, and one would expect the asymptotic results for the slope parameters to be identical to those in Corollary 3. A formal proof of this conjecture is likely to be conceptually straightforward, but tedious. This extension is beyond the scope of this paper, however, because it would require the generalization of the Jeganathan (1991) results to models with deterministic time trends.

Our proofs assume that the lag order is known. Alternatively, the lag order could be estimated based on information-based criteria (see Paulsen 1984; Tsay 1984). In that case, the bootstrap algorithm must be modified slightly to account for the lag-order uncertainty (see Kilian (1998) for further discussion).

3   Monte Carlo Evidence

The theoretical results of section 2 imply that in many cases the standard bootstrap algorithm will provide an asymptotically valid approximation to the true distribution of the autoregressive slope parameters, whether the true process is integrated or not. This result is important because it alleviates, at least asymptotically, the need for unit root pre-tests. In this section, we provide some preliminary Monte Carlo evidence of the accuracy of the proposed bootstrap approximation in finite samples. Further simulation evidence will be needed to corroborate our findings. Our aim here is merely to illustrate the potential usefulness of the proposed bootstrap method in finite samples.

First, consider bootstrap algorithm 1. In Table 1 we study the accuracy of the bootstrap approximation for slope parameters of integrated and near-integrated AR(2) models without drift when the regression model does not include an intercept. While the assumption of a known intercept is perhaps unrealistic, this case is of theoretical interest because it exactly matches the assumptions of Basawa et al. (1991a) and Datta (1996) except for the additional autoregressive lag. The table shows the percentiles of the true finite-sample distribution of $T^{1/2}(\hat{\phi}_i - \phi_i)$ $i = 1, 2$, and the average of the corresponding bootstrap percentiles of $T^{1/2}(\hat{\phi}_i^* - \hat{\phi}_i)$ obtained in 100 trials. The bootstrap percentiles for each trial are based on 20,000 replications. The results are not sensitive to increasing the number of trials. We consider sample sizes of 100, 300, and 500 observations, corresponding to small, moderately large and large sample sizes. Given the asymptotic validity result in Theorem 1, one would expect the bootstrap approximation to be close to the exact finite-sample distribution for sufficiently large sample sizes. Panel (a) suggests that the bootstrap approximation is reasonably accurate even for small and moderately large samples and is excellent for large samples.

The results in panel (a) support the use of bootstrap algorithm 1 when the population process has an exact unit root. In practice, however, it is rarely known with certainty whether the true process contains a unit root or not. We therefore investigate how well the standard bootstrap approximation continues to work for near-integrated processes without drift. Panels (b) and (c) in Table 1 focus on similar AR(2) population processes with dominant roots of approximately 0.99 and 0.98, respectively. We find that the standard bootstrap approximation works equally well for near-integrated higher-order autoregressive processes. Thus, the simulation results in Table 1 support the view that under suitable conditions valid bootstrap inference is possible using standard resampling algorithms without taking a stand on the unit root question.

We now turn to more general models. In most applied work the intercept is unknown and the fitted model will include an intercept as described in bootstrap algorithm 2. This bootstrap estimator has different properties than the estimator underlying Table 1. As in Theorems 2 and 3 we distinguish population models with drift (Tables 2 and 3) and without drift (Table 4). The results in Table 2 are based on the same AR(2) data generating process as the results in Table 1, except that the population model includes an additional drift term of the same magnitude as the standard deviation of the innovation. Applying bootstrap algorithm 2 produces results not unlike those in Table 1. Panel (a) shows that the bootstrap approximation to the distribution of the slope parameters is adequate for moderately large samples and highly accurate for large sample sizes. Similar results hold for the near-integrated processes in panels (b) and (c).

Table 3 shows the corresponding results for the random walk model with drift. Again, the bootstrap approximation works well for all roots considered and mimics the changes in the finite-sample distributions of the slope parameters as the root approaches unity. This result is in sharp contrast to the deterioration of the accuracy of the bootstrap approximation near the unit circle in the near-random walk model without drift.

Next, we re-examine the case of the population model without drift. The results in Table 4 differ from the results in Table 1 in that we use bootstrap algorithm 2 rather than bootstrap algorithm 1. Not imposing the zero intercept results in a sharp deterioration of the accuracy of the bootstrap approximation in small samples. The apparent reason is the additional small-sample bias of the OLS estimator in the model with intercept, which is further propagated by the bootstrap estimator. Nevertheless, there is clear evidence of gradual convergence of the bootstrap percentiles to the exact finite-sample percentiles, as the sample size is increased. In sharp contrast to the exact unit root case in panel (a), for the corresponding near-integrated processes without drift in panels (b) and (c) the bootstrap approximation continues to work well for moderately large and large samples.

How important for applied work is the slow convergence of the bootstrap distribution

for exact unit root processes without drift? Most processes of interest in empirical research are trending like the processes considered in Table 2. On the other hand, there are some highly persistent processes (such as interest rates, inflation rates or real exchange rates) that may not be trending. Those processes are unlikely to be exact unit root processes without drift, however. More likely they have roots very close to unity and means different from zero. For such processes the results in Table 2 suggest that the bootstrap approximation may be reasonably accurate after all. The evidence in Table 2 shows that a nonzero drift in population tends to eliminate the small-sample bias of the OLS estimator, resulting in much higher bootstrap accuracy. A similar effect can be expected for near-integrated processes with roots very close to unity. In Table 5 we explore this possibility for a process with a root of approximately 0.998. Table 5 shows that even for roots very close to unity a small nonzero intercept (relative to the standard deviation of the innovation) drastically improves the finite-sample accuracy of the bootstrap approximation. Thus, unless we have strong a priori reason to believe that the true process is indeed integrated with zero drift, the bootstrap approximation is likely to be adequate even for time series such as nominal or real interest rates or for real exchange rates. A detailed investigation of this conjecture more appropriately would be the subject of a separate and more extensive simulation study.

4   Concluding Remarks

We studied the asymptotic validity of the standard bootstrap algorithm for unrestricted autoregressions in models with exact unit roots. The main contribution of this paper has been two-fold. First, we derived the limit distribution of the bootstrap estimator for the augmented Dickey-Fuller representation of an integrated univariate autoregressive process. Second, on the basis of this theoretical result, we showed that, in many cases of interest in applied work, asymptotically valid bootstrap inference may be conducted without conditioning on the outcome of unit root pre-tests.

There are two natural extensions of this result. One extension involves generalizing the results of this paper to possibly integrated and/or cointegrated vector valued processes. This extension will require a generalization of the results in Jeganathan (1991) to vector valued processes. The other extension is a study of the conditions under which the bootstrap approximation described in this paper provides asymptotic refinements for the studentized estimator of the slope parameter (or smooth functions thereof). Our results in this paper imply that the bootstrap approximation is as sound asymptotically as the usual normal approximation, but they do not establish that the bootstrap approximation provides higher-order refinements. We conjecture that such refinements will be possible in the model with drift, but not in the model without drift, owing to

the rate of convergence of the estimator of the sum of the autoregressive coefficients. Both extensions are nontrivial and more appropriately dealt with in a separate paper.

Appendix

Proof of Theorem 1. The proof is organized as follows: First, we consider the population process (2) with ρ and ζ_j replaced by deterministic sequences ρ_T and $\zeta_{T,j}$ such that $\rho_T = 1 + \gamma_0/T + o(T^{-1})$ and $\zeta_{T,j} = \zeta_j + \gamma_j/T^{1/2} + o(T^{-1/2})$ for $j = 1, 2, ..., p-1$ where $\gamma = (\gamma_0, \gamma_1, ..., \gamma_{p-1})'$ is fixed. Next, we shall derive the limiting distribution of the bootstrap OLS estimator when $\rho, \zeta_1, ..., \zeta_{p-1}$ in (2) are replaced with their OLS estimates.

Consider the following triangular array:

$$
\begin{aligned}
y_{T,t} &= \phi_{T,1} y_{T,t-1} + \phi_{T,2} y_{T,t-2} + \cdots + \phi_{T,p} y_{T,t-p} + \varepsilon_t \\
&= \rho_T y_{T,t-1} + \zeta_{T,1}(1 - \rho_T L)y_{T,t-1} + \cdots + \zeta_{T,p-1}(1 - \rho_T L)y_{T,t-p+1} + \varepsilon_t \\
&= \beta_T' x_{T,t} + \varepsilon_t,
\end{aligned}
$$

where $y_{T,0} = y_{T,1} = \cdots = y_{T,1-p} = 0$, $\varepsilon_t \sim \mathrm{iid}(0, \sigma^2)$,

$$
\begin{aligned}
\rho_T &= \phi_{T,1} + \phi_{T,2} + \cdots + \phi_{T,p} \\
\zeta_{T,j} &= -[\phi_{T,j+1} + \phi_{T,j+2} + \cdots + \phi_{T,p}] \text{ for } j = 1, 2, ..., p-1 \\
\beta_T &= (\rho_T, \zeta_{T,1}, \zeta_{T,2}, ..., \zeta_{T,p-1})' \\
&= \beta + \Upsilon_T^{-1} \gamma \\
x_{T,t} &= (y_{T,t-1}, (1 - \rho_T L)y_{T,t-1}, (1 - \rho_T L)y_{T,t-2}, ..., (1 - \rho_T L)y_{T,t-p+1})' \\
\Upsilon_T &= \mathrm{diag}\,(T, T^{1/2}, ..., T^{1/2}).
\end{aligned}
$$

Let $z_{T,t} = (\psi_T(L)\varepsilon_t, (1 - \rho_T L)y_{T,t-1}, (1 - \rho_T L)y_{T,t-2}, ..., (1 - \rho_T L)y_{T,t-p+1})'$ where $\psi_T(L) = \phi_T(L)/(1 - \rho_T L)$. Then by Propositions 1, 4 and 5 of Jeganathan (1991),

$$
\Upsilon_T^{-1} \sum_{t=1}^{T} z_{T,t} z_{T,t}' \Upsilon_T^{-1} \xrightarrow{L} \begin{bmatrix} \int_0^1 S(r, \gamma_0)^2 dr & 0_{1 \times (p-1)} \\ 0_{(p-1) \times 1} & \Gamma \end{bmatrix} \tag{21}
$$

$$
\Upsilon_T^{-1} \sum_{t=1}^{T} z_{T,t} \varepsilon_t \xrightarrow{L} \begin{bmatrix} \int_0^1 S(r, \gamma_0) dB(r) \\ \sigma \Gamma W_{p-1} \end{bmatrix} \tag{22}
$$

where B is a Brownian motion, W_{p-1} is a $(p-1)$-dimensional standard normal random vector and B and W_{p-1} are independent. In other words, by Theorem 1 of Jeganathan (1991),

$$
\lim_{T \to \infty} P\left(\Upsilon_T \left(\sum_{t=1}^{T} z_{T,t} z_{T,t}' \right)^{-1} \sum_{t=1}^{T} z_{T,t} \varepsilon_t \leq x \right) = H(\gamma, x) \tag{23}
$$

17

where $H(\gamma, x)$ is the joint distribution function of the p-dimensional random vector

$$
\begin{bmatrix}
\frac{\psi(1)\int_0^1 S(r,\gamma_0)dB(r)}{\int_0^1 S(r,\gamma_0)^2 dr} \\
\sigma\Gamma^{-\frac{1}{2}}W_{p-1}
\end{bmatrix}.
\tag{24}
$$

Next, we shall derive the limiting distribution of the bootstrap OLS estimator when $\rho, \zeta_1, ..., \zeta_{p-1}$ in (2) are replaced with their OLS estimates. Let η_T denote a random measure on $\Re^p$ defined by

$$
\eta_T(S) = \int_S H_T(dx)
$$

where S is a Borel set in $\Re^p$,

$$
H_T(x) = P^* \left(\Upsilon_T \left(\sum_{t=1}^T z_t^* z_t^{*\prime} \right)^{-1} \sum_{t=1}^T z_t^* \varepsilon_t^* \le x \right).
$$

P^* is the probability measure induced by the bootstrap conditional on the original data. In other words, $\eta_T(S) = P^* \left(\Upsilon_T \left(\sum_{t=1}^T z_t^* z_t^{*\prime} \right)^{-1} \sum_{t=1}^T z_t^* \varepsilon_t^* \in S \right)$. Noting that $H(\gamma, x)$ is continuous in γ for each fixed x, we define a random measure η on $\Re^p$ by

$$
\eta(S) = \int_S H(\tilde{\gamma}, dx)
$$

where $\tilde{\gamma}$ is the p-dimensional random vector given by the RHS of (3). We want to show that, as $T \to \infty$

$$
\eta_T \Rightarrow \eta \quad \omega\text{-a.s.}
\tag{25}
$$

As Datta (1996, Remark 2.2) points out, η_T is not a function of $\hat{\beta}$ alone, but also a function of $\tilde{F}_T$, and thus it is difficult to prove (25) directly. In order to circumvent this difficulty, we follow a strategy similar to that adopted by Datta (1996) for the random walk model. Recall that in population $\phi(L)y_t = \varepsilon_t$ where $\varepsilon_t \overset{iid}{\sim} F$. First, consider a thought experiment in which we bootstrap the slope parameters under the counterfactual assumption that the true error distribution is known, i.e.,

$$
\hat{\phi}(L)\tilde{y}_t^* = \tilde{\varepsilon}_t^*, \quad t = 1, 2, ..., T,
\tag{26}
$$

where $\tilde{\varepsilon}_t^* \overset{iid}{\sim} F$ and $\tilde{y}_0^* = \cdots = \tilde{y}_{1-p}^* = 0$. The bootstrap analogue of η_T under this resampling scheme is given by

$$
\tilde{\eta}_T(S) = \int_S \tilde{H}_T(dx)
$$

18

where $\tilde{H}_T(x) = P^*\left(\Upsilon_T\left(\sum_{t=1}^{T}\tilde{z}_t^*\tilde{z}_t^{*'}\right)^{-1}\sum_{t=1}^{T}\tilde{z}_t^*\tilde{\varepsilon}_t^* \leq x\right)$, $\tilde{z}_t^* = (\tilde{u}_t^*, \Delta\tilde{y}_{-1}^*, ..., \Delta\tilde{y}_{t-p}^*)'$ and $\tilde{u}_t^* = \hat{\psi}(L)\tilde{y}_t^*$. Note that the assumption underlying Basawa et al.'s (1991a) proof is a special case of this algorithm with F replaced by the standard normal distribution.

Next, consider the standard bootstrap resampling scheme

$$\hat{\phi}(L)y_t^* = \varepsilon_t^*, \quad t = 1, 2, ..., T,$$

where $\varepsilon_t^* \overset{iid}{\sim} \tilde{F}$ and $y_0^* = \cdots = y_{1-p}^* = 0$. Let $z_t^* = (u_t^*, \Delta y_{-1}^*, ..., \Delta y_{t-p}^*)'$ where $u_t^* = \hat{\psi}(L)y_t^*$. Keeping the difference in notation between these two resampling schemes in mind, let $\tilde{U}_T^* = \Upsilon_T^{-1}\sum_{t=1}^{T}\tilde{z}_{t-1}^*\tilde{\varepsilon}_t^*$, $U_T^* = \Upsilon_T^{-1}\sum_{t=1}^{T}z_{t-1}^*\varepsilon_t^*$, $\tilde{V}_T^* = \Upsilon_T^{-1}\sum_{t=1}^{T}\tilde{z}_t^*\tilde{z}_t^{*'}\Upsilon_T^{-1}$ and $V_T^* = \Upsilon_T^{-1}\sum_{t=1}^{T}z_t^*z_t^{*'}\Upsilon_T^{-1}$.

We will proceed as follows. First, in Lemma 1 we will provide bounds for the difference between $\tilde{U}_T^*$ and U_T^* and between $\tilde{V}_T^*$ and V_T^*. We show that this difference vanishes asymptotically for $\hat{\rho}$ close to unity. This fact allows us to substitute U_T^* and V_T^* for $\tilde{U}_T^*$ and $\tilde{V}_T^*$, respectively, and to proceed as though F were known. Second, in Lemma 2, we will prove that the distance between the bootstrap innovation and the true innovation vanishes asymptotically. This result will be used in proving Lemma 3. Third, in Lemma 3, we will derive the limit distribution of U_T^* and V_T^* and show that it is of a known form, corresponding to the numerator and the denominator of (4). Finally, we will apply the continuous mapping theorem to show that the ratio of U_T^* and V_T^* indeed converges to the limit distribution of the bootstrap estimator in (4).

Let $\mathcal{M}(\Re^s)$ denote the space of probability measures on $\Re^s$ topologized by weak convergence and let $\Rightarrow$ denote weak convergence of probability measures. Given an $n \times m$ matrix $A = \{a_{ij}\}$, let $\|A\| = \sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|$. For $\mu, \nu \in \Gamma_r = \{\gamma \in \mathcal{M}(\Re^q) : \int\|x\|^r\gamma(dx) < \infty\}$, let $d_r(\mu, \nu)$ be the infimum of $(E\|X - Y\|^r)^{1/r}$, over all possible joint distributions of two s-dimensional random vectors X and Y, whose marginal distributions are μ and ν, respectively. The notation $L^{X,Y}$ stands for the joint distribution of X and Y. Define a metric on $\mathcal{M}(\Re^s)$ by

$$d(\mu, \nu) = \sum_{n=1}^{\infty}2^{-n}\left(\left|\int f_n d\mu - \int f_n d\nu\right| \wedge 1\right)$$

for $\mu, \nu \in \mathcal{M}(\Re^s)$, where f_n is some bounded uniformly continuous real-valued function on $\Re^s$ and $a \wedge b$ stands for the minimum of a and b.

Lemma 1.

$$d_1(\tilde{V}_T^*, V_T^*) \leq \left\{1 + 2(p-1)\left(\sum_{j=0}^{\infty}\hat{\theta}_j^2\right)^{\frac{1}{2}}\left[1 + (1 \vee |\hat{\rho}|^{T-1})(T-1)|\hat{\rho} - 1|\right]\right.$$

19

$$+(p-1)^2 \left(\sum_{j=0}^{\infty} \hat{\theta}_j^2\right) \left[1 + (1 \vee |\hat{\rho}|^{T-1})(T-1)|\hat{\rho}-1|\right]^2 /(1 \vee |\hat{\rho}|^{T-1})\Bigg\}$$

$$\times (1 \vee |\hat{\rho}|^{T-1})(\sigma + \hat{\sigma})d_2(\varepsilon_1, \varepsilon_1^*), \tag{27}$$

$$d_1(\tilde{U}_T^*, U_T^*) = \Bigg\{ (1 \vee |\hat{\rho}|^{T-1})T|\hat{\rho}-1|(\sigma+\hat{\sigma}) \left[1 + \left(\sum_{j=0}^{\infty} \hat{\theta}_j^2\right)^{\frac{1}{2}}\right]$$

$$+(\sigma + \hat{\sigma})\left(\sum_{j=0}^{\infty} \hat{\theta}_j^2\right)^{\frac{1}{2}} + (1 \vee |\hat{\rho}|^{T-1})T^{-\frac{1}{2}} + d_2(\varepsilon_1, \varepsilon_1^*)\Bigg\} d_2(\varepsilon_1, \varepsilon_1^*),$$

$$\tag{28}$$

$$d_1((\tilde{U}_T^*, \tilde{V}_T^*), (U_T^*, V_T^*)) \leq \text{ RHS of (27) + RHS of (28)}, \tag{29}$$

where $a \vee b$ stands for the maximum of a and b.

Proof of Lemma 1.

Note that

$$E^*|\tilde{u}_{t-1}^{*2} - u_{t-1}^{*2}| \leq E^*|\tilde{u}_{t-1}^* + u_{t-1}^*|E^*|\tilde{u}_{t-1}^* - u_{t-1}^*|$$
$$\leq (E^*|\tilde{u}_{t-1}| + E^*|u_{t-1}|)E^*|\tilde{u}_{t-1}^* - u_{t-1}^*|. \tag{30}$$

Since $u_t^* = \sum_{s=1}^t \hat{\rho}^{t-s}\varepsilon_s$ and $\tilde{u}_t^* = \sum_{s=1}^t \hat{\rho}^{t-s}\tilde{\varepsilon}_s$, it follows that

$$E^*|\tilde{u}_{t-1}^* - u_{t-1}^*| \leq \left(\sum_{j=1}^{t-1} \hat{\rho}^{2j}\right)^{\frac{1}{2}} d_2(\varepsilon_1, \varepsilon_1^*) \leq (1 \vee |\hat{\rho}|^{t-1})d_2(\varepsilon_1, \varepsilon_1^*)(t-1)^{\frac{1}{2}}. \tag{31}$$

Since $u_{t-1}^* = \sum_{s=1}^{t-1} \hat{\rho}^{t-s}\varepsilon_s^*$ and $\tilde{u}_t^* = \sum_{s=1}^{t-1} \hat{\rho}^{t-s}\tilde{\varepsilon}_s^*$, it follows that

$$E^*|u_{t-1}| \leq (1 \vee |\hat{\rho}|^{t-1})\hat{\sigma}(t-1)^{\frac{1}{2}} \tag{32}$$

$$E^*|\tilde{u}_{t-1}| \leq (1 \vee |\hat{\rho}|^{t-1})\sigma(t-1)^{\frac{1}{2}} \tag{33}$$

Thus, it follows from (30), (31), (32) and (33) that

$$E^*|\tilde{u}_{t-1}^{*2} - u_{t-1}^{*2}| \leq (1 \vee |\hat{\rho}|^{t-1})(\sigma + \hat{\sigma})d_2(\varepsilon_1, \varepsilon_1^*)(t-1). \tag{34}$$

Observe that

$$E|\tilde{u}_{t-1}^* \Delta \tilde{y}_{t-j}^* - u_{t-1}^* \Delta y_{t-j}^*|$$
$$\leq E|(\tilde{u}_{t-1}^* - u_{t-1}^*)\Delta \tilde{y}_{t-j}^*| + E|(\Delta \tilde{y}_{t-1}^* - \Delta y_{t-1}^*)u_{t-j}^*|$$
$$\leq E|\tilde{u}_{t-1}^* - u_{t-1}^*|E|\Delta \tilde{y}_{t-j}^*| + E|\Delta \tilde{y}_{t-1}^* - \Delta y_{t-1}^*|E|u_{t-j}^*|. \tag{35}$$

Since

$$
\begin{aligned}
\Delta y_{t-j}^* &= \frac{1-L}{1-\hat{\rho}L}\hat{\psi}^{-1}(L)\varepsilon_{t-j}^* \\
&= \left[1+(\hat{\rho}-1)\sum_{i=1}^{\infty}\hat{\rho}^{i-1}L^i\right]\sum_{k=0}^{\infty}\hat{\theta}_k\varepsilon_{t-j-k}^*, \\
&= \left[1+(\hat{\rho}-1)\sum_{i=1}^{t-j}\hat{\rho}^{i-1}L^i\right]\sum_{k=0}^{t-j}\hat{\theta}_k\varepsilon_{t-j-k}^*, \\
\Delta \tilde{y}_{t-j}^* &= \left[1+(\hat{\rho}-1)\sum_{i=1}^{t-j}\hat{\rho}^{i-1}L^i\right]\sum_{k=0}^{t-j}\hat{\theta}_k\tilde{\varepsilon}_{t-j-k}^*,
\end{aligned}
$$

where $\hat{\theta}_j$ is the jth moving average coefficient of the moving average representation of $\hat{\psi}^{-1}(L)$, it follows that

$$
E|\Delta\tilde{y}_{t-j}^*| \leq \left[1+(1\vee|\hat{\rho}|^{t-j})|\hat{\rho}-1|(t-j)\right]\left(\sum_{k=0}^{\infty}\hat{\theta}_k^2\right)^{\frac{1}{2}}\sigma, \tag{36}
$$

$$
E|\Delta\tilde{y}_{t-j}^*-\Delta y_{t-j}^*| \leq \left[1+(1\vee|\hat{\rho}|^{t-j})|\hat{\rho}-1|(t-j)\right]\left(\sum_{k=0}^{\infty}\hat{\theta}_k^2\right)^{\frac{1}{2}}d_2(\varepsilon_1,\varepsilon_1^*). \tag{37}
$$

It follows from (32), (34), (35), (36) and (37) that

$$
E\left|\tilde{u}_{t-1}^*\Delta\tilde{y}_{t-j}^*-u_{t-1}^*\Delta y_{t-j}^*\right|
$$
$$
\leq \left[1+(1\vee|\hat{\rho}|^{t-j})|\hat{\rho}-1|(t-j)\right]\left(\sum_{k=0}^{\infty}\hat{\theta}_k^2\right)^{\frac{1}{2}}(1\vee\hat{\rho}^{t-1})(\sigma+\hat{\sigma})d_2(\varepsilon_1,\varepsilon_1^*)(t-1)^{\frac{1}{2}}. \tag{38}
$$

Similarly, we obtain

$$
E^*|\Delta\tilde{y}_{t-i}^*\Delta\tilde{y}_{t-j}^*-\Delta y_{t-i}^*\Delta y_{t-j}^*|
$$
$$
\leq (\sigma+\hat{\sigma})\sum_{k=0}^{\infty}\hat{\theta}_k^2\left[1+(1\vee|\hat{\rho}|^{t-i})|\hat{\rho}-1|(t-i)\right]
$$
$$
\times\left[1+(1\vee|\hat{\rho}|^{t-j})|\hat{\rho}-1|(t-j)\right]d_2(\varepsilon_1,\varepsilon_1^*). \tag{39}
$$

21

Since

$$d_1(\tilde{V}_T^*, V_T^*) = \sum_{i=1}^{p} \sum_{j=1}^{p} \left| \left[\Upsilon_T^{-1} \sum_{t=1}^{T} \left(\tilde{z}_t^* \tilde{z}_t^{*'} - z_t^* z_t^{*'} \right) \Upsilon_T \right]_{i,j} \right|$$

where $[A]_{i,j}$ denote the (i,j)th element of A, (27) follows from (34), (38) and (39).

Since

$$\begin{aligned} u_t^* &= \hat{\rho} u_{t-1}^* + \varepsilon_t^* \\ u_t^{*2} &= \hat{\rho}^2 u_{t-1}^{*2} + 2\hat{\rho} u_{t-1}^* \varepsilon_t^* + \varepsilon_t^{*2} \\ \sum_{t=1}^{T} u_t^{*2} &= \hat{\rho}^2 \sum_{t=1}^{T} u_{t-1}^{*2} + 2\hat{\rho} \sum_{t=1}^{T} u_{t-1}^* \varepsilon_t^* + \sum_{t=1}^{T} \varepsilon_t^{*2}, \end{aligned}$$

it follows that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} u_t^* \varepsilon_t^* &= \frac{1}{2\hat{\rho}T} \left(\sum_{t=1}^{T} u_t^{*2} - \hat{\rho}^2 \sum_{t=1}^{T} u_{t-1}^{*2} - \sum_{t=1}^{T} \varepsilon_t^{*2} \right), \\ &= \frac{u_T^*}{2\hat{\rho}T} + \frac{T(1-\hat{\rho}^2)}{2\hat{\rho}} \frac{1}{T^2} \sum_{t=1}^{T} u_t^{*2} - \frac{1}{2\hat{\rho}T} \sum_{t=1}^{T} \varepsilon_t^{*2}. \end{aligned} \tag{40}$$

Similarly,

$$\frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t^* \tilde{\varepsilon}_t^* = \frac{\tilde{u}_T^*}{2\hat{\rho}T} + \frac{T(1-\hat{\rho}^2)}{2\hat{\rho}} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{u}_t^{*2} - \frac{1}{2\hat{\rho}T} \sum_{t=1}^{T} \tilde{\varepsilon}_t^{*2}. \tag{41}$$

It follows from (34), (40) and (41) that

$$E^* \left| \frac{1}{T} \sum_{t=1}^{T} u_t^* \varepsilon_t^* - \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t^* \tilde{\varepsilon}_t^* \right|$$

$$\leq \left\{ (1 \vee |\hat{\rho}|^{T-1}) \left[T^{-\frac{1}{2}} + T(1-\hat{\rho}^2)(\sigma + \hat{\sigma}) \right] + d_2(\varepsilon_1, \varepsilon_1^*) \right\} \frac{d(\varepsilon_1, \varepsilon_1^*)}{2\hat{\rho}}. \tag{42}$$

Since

$$E^* \left| \Delta y_{t-j}^* \varepsilon_t^* - \Delta \tilde{y}_{t-j} \tilde{\varepsilon}_t \right| \leq E \left| \Delta y_{t-j}^* - \Delta \tilde{y}_{t-j} \right| \hat{\sigma} + E |\Delta \tilde{y}_{t-j}| d_2(\varepsilon_1, \varepsilon_1^*),$$

it follows from (36) and (37) that

$$E^* \left| \Delta y_{t-j}^* \varepsilon_t^* - \Delta \tilde{y}_{t-j} \tilde{\varepsilon}_t \right|$$

$$\leq \left[1 + (1 \vee |\hat{\rho}|^{T-1})(T-1)|\hat{\rho} - 1| \right] (\sigma + \hat{\sigma}) \left(\sum_{j=0}^{\infty} \hat{\theta}_j^2 \right)^{\frac{1}{2}} d_2(\varepsilon_1, \varepsilon_1^*). \tag{43}$$

22

(28) follows from (42) and (43).

Then (29) follows immediately from (27) and (28). ∎

In the rest of the proof, we treat $L^{\tilde{U}_T^*, \tilde{V}_T^*}$, $L^{U_T^*, V_T^*}$ and $L^{U,V}$ as $\mathcal{M}(\Re^p \times M_{p,p})$-valued random elements, where $M_{p,p}$ is a set of $p \times p$ real matrices. Lemma 2 shows that the distance between the bootstrap innovation and the population innovation vanishes asymptotically. This result will be used in the proof of Lemma 3.

Lemma 2.
$$d_2(\varepsilon_t^*, \varepsilon_t) \to 0 \quad \omega\text{-a.s.} \tag{44}$$

for all t.

Proof of Lemma 2. By Lemma 8.3 of Bickel and Freedman (1981), it is sufficient to show that $\tilde{F}_T \Rightarrow F$ ω-a.s. and $E^*(\tilde{\varepsilon}_t^2) \to \sigma^2$ ω-a.s.

Since $\{\varepsilon_t\}$ is iid, $F_T \Rightarrow F$ ω-a.s. where F_T is the empirical distribution of $\{\varepsilon_t\}$. By the stochastic equicontinuity of $\{F_T(\cdot)\}$,

$$\sup_{x,x':|x-x'|\le\delta_T} |F_T(x') - F(x') - F_T(x) + F(x)| = o_{as}(1) \tag{45}$$

for $\delta_T \to 0$ as $T \to \infty$. By applications of Theorem 3 and Lemma 1 of Zheng (1992),

$$\hat{\rho} - \rho = O_{as}(T^{-1}(\log\log T)^2), \tag{46}$$
$$\hat{\zeta}_j - \zeta_j = O_{as}(T^{-1/2}(\log\log T)^{1/2}). \tag{47}$$

By the functional law of iterated logarithm,

$$\sum_{t=1}^{[Tr]} \psi(L)\varepsilon_t = O_{as}((T\log\log T)^{1/2}) \tag{48}$$

for $r \in [0,1]$. Because the distribution of ε_t, F, satisfies the Lipschitz condition, there is $M > 0$ such that

$$|F(x') - F(x)| \le M|x' - x| \tag{49}$$

By the standard arguments about convergence of empirical processes, it follows from (45)-(49) that

$$\hat{F}_T \Rightarrow F \quad \omega\text{-a.s.} \tag{50}$$

where $\hat{F}_T$ is the empirical distribution of $\hat{\varepsilon}_t$. It follows from (46)-(48) that

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t = o_{as}(1). \tag{51}$$

23

By the stochastic equicontinuity of $\hat{F}_T$ and from (49)-(51), it follows that

$$\tilde{F}_T \Rightarrow F \quad \omega\text{-a.s.} \tag{52}$$

Arguments analogous to the proof of Lemma 4.1 of Datta (1996) lead to

$$E^*(\varepsilon_t^{*2}) = \hat{\sigma}^2 \overset{a.s.}{\rightarrow} \sigma^2 \tag{53}$$

An application of Lemma 8.3 of Bickel and Freedman (1981) to (52) and (53) completes the proof of Lemma 2.

Lemma 3. Let

$$
\begin{aligned}
U &= \left[\begin{array}{c} \int_0^1 S(r, \tilde{\gamma}_0) dB(r) \\ \sigma \Gamma^{\frac{1}{2}} W \end{array} \right], \\
V &= \left[\begin{array}{cc} \int_0^1 S(r, \tilde{\gamma}_0)^2 dr & 0_{1 \times (p-1)} \\ 0_{(p-1) \times 1} & \Gamma \end{array} \right],
\end{aligned}
$$

where $\tilde{\gamma}_0$ is the first element of the random variable on the RHS of (3). Then

$$L^{U_T^*, V_T^*} \overset{L}{\rightarrow} L^{U,V} \quad \omega\text{-a.s.} \tag{54}$$

Proof of Lemma 3.

Since $L^{U_T^*, V_T^*}$ and $L^{U,V}$ are $\mathcal{M}(\Re^p \times M_{p,p})$-valued random elements and $\mathcal{M}(\Re^p \times M_{p,p})$ is separable, by Theorems 6.2 and 6.6 of Parthasarathy (1967), it is sufficient to prove

$$E[g(L^{U_T^*, V_T^*})] \rightarrow E[g(L^{U,V})] \tag{55}$$

for every bounded uniformly continuous function g on $\mathcal{M}(\Re^p \times M_{p,p})$.

$$
\begin{aligned}
& \left| E[g(L^{U_T^*, V_T^*})] - E[g(L^{\tilde{U}_T^*, \tilde{V}_T^*})] \right| \\
\leq \quad & E\left[\left| g(L^{\tilde{U}_T^*, \tilde{V}_T^*}) - g(L^{\tilde{U}_T^*, \tilde{V}_T^*}) \right| I(T|\hat{\rho} - 1| \leq K) \right] + M P(T|\hat{\rho} - 1| > K),
\end{aligned} \tag{56}
$$

where M is a bound for g, K is a real number and $I(\cdot)$ is the indicator function. Since $T(\hat{\rho}_T - 1) = O_p(1)$, the second term can be made arbitrarily small by choosing a sufficiently large K. On the set $\{T|\hat{\rho} - 1| \leq K\}$, $d_1(L^{U_T^*, V_T^*}, L^{\tilde{U}_T^*, \tilde{V}_T^*}) \rightarrow 0$ by Lemmas 1 and 2, and thus $d(L^{U_T^*, V_T^*}, L^{\tilde{U}_T^*, \tilde{V}_T^*}) \rightarrow 0$ by Theorem II 6.6 in Parthasarathy (1967). Since g is uniformly continuous and bounded, the first term converges to zero by the dominated convergence theorem. Thus,

$$\left| E[g(L^{U_T^*, V_T^*})] - E[g(L^{\tilde{U}_T^*, \tilde{V}_T^*})] \right| \rightarrow 0. \tag{57}$$

It follows from (3) and the Skorohod representation theorem that there exist $\tilde{\beta}$ and $\tilde{\gamma}$ such that $\Upsilon_T(\tilde{\beta} - \beta) \overset{L}{\to} \tilde{\gamma}$ a.s., $L^{\tilde{\beta}} = L^{\hat{\beta}}$ and $L^{\tilde{\gamma}} = L^{\gamma}$. Let $\tilde{U}_T$ and $\tilde{V}_T$ correspond to $\tilde{U}_T^*$ and $\tilde{V}_T^*$, respectively, with $\hat{\beta}$ replaced by $\tilde{\beta}$. Let $\tilde{U}$ and $\tilde{V}$ correspond to U and V, respectively, with γ replaced by $\tilde{\gamma}$. By Lemma 1,

$$L^{\tilde{U}_T, \tilde{V}_T} \Rightarrow L^{\tilde{U}, \tilde{V}} \quad \omega\text{-a.s.} \tag{58}$$

Since $L^{\tilde{U}_T, \tilde{V}_T} = L^{\tilde{U}_T^*, \tilde{V}_T^*}$ and $L^{\tilde{U}, \tilde{V}} = L^{U,V}$ by the Skorohod representation theorem, it follows from (58) that

$$E[g(L^{\tilde{U}_T^*, \tilde{V}_T^*})] \to E[g(L^{U,V})] \tag{59}$$

by the dominant convergence theorem. ∎

Note that U_T^* and V_T^* correspond to the numerator and the denominator of the limiting expression of the bootstrap estimator in (4). Thus, we can derive the limit distribution of the estimator by analyzing the limit distribution of the ratio $V_T^{*-1} U_T^*$. Our proof relies on an application of the continuous mapping theorem.

It follows from Lemma 3 that

$$\Upsilon_T^{-1} \sum_{t=1}^{T} z_t^* z_t^{*'} \Upsilon_T^{-1} \quad \overset{L}{\to} \quad V \quad \omega\text{-a.s.} \tag{60}$$

$$\Upsilon_T^{-1} \sum_{t=1}^{T} z_t^* \varepsilon_t^* \quad \overset{L}{\to} \quad U \quad \omega\text{-a.s.} \tag{61}$$

Since V is nonsingular with probability one by Lemma 1 and Proposition 4 of Jeganathan (1991), it follows by the continuous mapping theorem (e.g., Theorem 1.3.6 of van der Vaart and Wellner, 1996) that

$$\Upsilon_T \left(\sum_{t=1}^{T} z_t^* z_t^{*'} \right)^{-1} \sum_{t=1}^{T} z_t^* \varepsilon_t^* \quad \overset{L}{\to} \quad V^{-1} U \quad \omega\text{-a.s.}$$

$$\equiv \left(\begin{array}{c} \frac{\int_0^1 S(r, \tilde{\gamma}_0) dB(r)}{\int_0^1 S(r, \tilde{\gamma}_0)^2 dr} \\ \sigma \Gamma^{-\frac{1}{2}} W \end{array} \right) \quad \omega\text{-a.s.} \tag{62}$$

Note that $\Upsilon_T(\hat{\beta}_T^* - \hat{\beta})$ can be expressed in terms of z_t^* as

$$\Upsilon_T(\hat{\beta}_T^* - \hat{\beta}) = M_T \left(\sum_{t=1}^{T} z_t^* z_t^{*'} \right)^{-1} \sum_{t=1}^{T} z_t^* \varepsilon_t^*, \tag{63}$$

25

where

$$
M_T = \begin{pmatrix} \hat{\psi}(1)T & 0 & \cdots & 0 \\ \sum_{j=1}^{p-1} \hat{\psi}_j T^{1/2} & & & \\ \vdots & & I_{p-1}T^{1/2} & \\ \hat{\psi}_{p-1}T^{1/2} & & & \end{pmatrix},
$$

from which we obtain (4). ∎

Proof of Corollary 1.

Since

$$
\begin{aligned}
\phi_1 &= \rho + \zeta_1, \\
\phi_j &= \zeta_j - \zeta_{j-1} \text{ for } j = 2, ..., p-1, \\
\phi_p &= -\zeta_{p-1},
\end{aligned}
$$

(6) and (7) follow from (3) and Theorem 1, respectively. ∎

Proof of Theorem 2.

The proof of Theorem 2 is analogous to the proof of Theorem 1 except that Step 1 of the proof of Theorem 1 and Lemma 1 are replaced by Lemma 4 and Lemma 6, respectively. Although the bootstrap estimator of ρ converges to a random limit distribution, its rate of convergence is T compared with $\sqrt{T}$ for the coefficients of the lagged differences. Thus, any linear combination of bootstrap estimators involving coefficients on lagged differences will be $\sqrt{T}$ consistent and will converge to the same Gaussian limit distribution as the usual OLS estimator. Hence, the bootstrap provides an asymptotically valid approximation to the distribution of the autoregressive slope parameters, even in the presence of a unit root. ∎

Lemma 4. (Asymptotic Distribution of the Slope Parameters of Integrated AR(p) Processes with Drift)

Consider the following triangular array of an $AR(p)$ process with drift.

$$
\begin{aligned}
y_{T,t} &= \delta_T + \phi_{T,1} y_{T,t-1} + \phi_{T,2} y_{T,t-2} + \cdots + \phi_{T,p} y_{T,t-p} + \varepsilon_t \\
&= \beta_T' x_{T,t} + \varepsilon_t,
\end{aligned}
$$

where $y_{T,0} = y_{T,1} = \cdots = y_{T,1-p} = 0$, $\varepsilon_t \sim \text{iid}(0, \sigma^2)$,

$$
\begin{aligned}
\beta_T &= \begin{cases} (\delta_T/\psi_T(1), \rho_T, \zeta_{T,1}, \zeta_{T,2}, ..., \zeta_{T,p-1})' & \text{if } p > 1 \\ (\delta_T, \rho_T)' & \text{if } p = 1 \end{cases} \\
\phi_T(\rho_T) &= 0,
\end{aligned}
$$

26

$$\zeta_{T,j} = -[\phi_{T,j+1}/\rho_T + \phi_{T,j+2}/\rho_T^2 + \cdots + \phi_{T,p}/\rho_T^{p-j}] \text{ for } j = 1, 2, ..., p-1$$

$$x_{T,t} = \begin{cases} (1, y_{T,t-1}, (1 - \rho_T L)y_{T,t-1} - \frac{\delta_T}{\psi_T(1)}, ..., (1 - \rho_T L)y_{T,t-p+1} - \frac{\delta_T}{\psi_T(1)})' & \text{if } p > 1 \\ (1, y_{T,t-1})' & \text{if } p = 1. \end{cases}$$

Suppose that $T^{3/2}(\rho_{T,j} - 1) \to \gamma_0$, $T^{1/2}(\zeta_{T,j} - \zeta_j) \to \gamma_j$ for $j = 1, 2, ..., p-1$, and $T^{1/2}(\delta_T - \delta) \to \gamma_p$ where $\gamma = (\gamma_0, \gamma_1, ..., \gamma_p)'$ is a $(p+1)$-dimensional vector of constants. Then

$$\bar{\Upsilon}_T(\bar{\beta}_T - \beta_T) \overset{d}{\to} N(0_{(p+1)\times 1}, \sigma^2 \bar{\Gamma}^{-1})$$

where

$$\bar{\Upsilon}_T = \begin{cases} \text{diag} \, (T^{1/2}, T^{3/2}, T^{1/2}, ..., T^{1/2}) & \text{if } p > 1 \\ \text{diag} \, (T^{1/2}, T^{3/2}) & \text{if } p = 1. \end{cases}$$

Proof of Lemma 4. Observe that

$$\bar{\Upsilon}_T(\bar{\beta}_T - \beta_T) = \bar{\Upsilon}_T(\sum_{t=1}^T \bar{x}_{T,t}\bar{x}'_{T,t})^{-1} \sum_{t=1}^T \bar{x}_{T,t}\varepsilon_t$$

$$= \bar{M}_T(\sum_{t=1}^T \bar{z}_{T,t}\bar{z}'_{T,t})^{-1} \sum_{t=1}^T \bar{z}_{T,t}\varepsilon_t$$

where

$$\bar{M}_T = \begin{bmatrix} T^{1/2} & T^{1/2}(\delta_T/\psi_T(1))\sum_{i=1}^{p-1}\sum_{j=i}^{p-1}\psi_{T,j}/\rho_T^{j-i+1} & 0 & \cdots & 0 \\ 0 & T^{3/2}(1 - \sum_{j=1}^{p-1}\psi_{T,j}/\rho_T^j) & 0 & \cdots & 0 \\ 0 & T^{1/2}\sum_{j=1}^{p-1}\psi_{T,j}/\rho_T^j & & & \\ \vdots & \vdots & & T^{1/2}I_{p-1} & \\ 0 & T^{1/2}\psi_{T,p-1}/\rho_T & & & \end{bmatrix},$$

$$\bar{z}_{T,t} = (1, \psi_T(L)\varepsilon_t, (1 - \rho_T L)y_{T,t-1} - \delta_T/\psi_T(1), ..., (1 - \rho_T L)y_{T,t-p+1} - \delta_T/\psi_T(1))'.$$

Arguments analogous to Jeganathan (1991) with his Propositions 1 and 5 replaced by Lemma 5 below lead to

$$\bar{\Upsilon}_T^{-1} \sum_{t=1}^T \bar{z}_{T,t}\bar{z}'_{T,t}\bar{\Upsilon}_T^{-1} \quad \overset{p}{\to} \quad \bar{\Gamma}$$

$$\bar{\Upsilon}_T^{-1} \sum_{t=1}^T \bar{z}_{T,t}\varepsilon_t \quad \overset{d}{\to} \quad N(0_{(p+1)\times 1}, \sigma^2 \bar{\Gamma}).$$

Therefore the desired conclusion follows. ∎

Lemma 5.

$$\left[\begin{array}{cc} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{array}\right] \sum_{t=1}^{T} \left(\begin{array}{c} 1 \\ \bar{u}_{T,t} \end{array}\right) \varepsilon_t \overset{d}{\to} N\left(0_{2\times 1}, \left(\begin{array}{cc} 1 & \frac{\delta}{2} \\ \frac{\delta}{2} & \frac{\delta^2}{3} \end{array}\right)\right), \tag{64}$$

$$\left[\begin{array}{cc} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{array}\right] \sum_{t=1}^{T} \left(\begin{array}{c} 1 \\ \bar{u}_{T,t} \end{array}\right) \left(\begin{array}{c} 1 \\ \bar{u}_{T,t} \end{array}\right)' \left[\begin{array}{cc} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{array}\right]$$

$$\overset{p}{\to} \left(\begin{array}{cc} 1 & \frac{\delta}{2} \\ \frac{\delta}{2} & \frac{\delta^2}{3} \end{array}\right), \tag{65}$$

$$\left[\begin{array}{cc} T^{-1} & 0 \\ 0 & T^{-2} \end{array}\right] \sum_{t=1}^{T} \left(\begin{array}{c} 1 \\ \bar{u}_{T,t} \end{array}\right) (\Delta y_{T,t-j} - \delta_T/\psi_T(1)) = o_p(1), \tag{66}$$

where $\bar{u}_{T,t} = \psi_T(L)y_{T,t}$.

Lemma 4 provides the asymptotics for the OLS estimator of β_T. Note that the usual local-to-unity results, which assume δ_T to be constant, do not apply in our context because δ in the bootstrap world will be a function of the random variable γ_p. Our result shows that the limiting distributions are the same in both cases.

Lemma 6. Let $\tilde{U}_T^* = \Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{z}_{t-1}^* \tilde{\varepsilon}_t^*$, $U_T^* = \Upsilon_T^{-1} \sum_{t=1}^{T} z_{t-1}^* \varepsilon_t^*$, $\tilde{V}_T^* = \Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{z}_t^* \tilde{z}_t^{*\prime} \Upsilon_T^{-1}$ and $V_T^* = \Upsilon_T^{-1} \sum_{t=1}^{T} z_t^* z_t^{*\prime} \Upsilon_T^{-1}$. Then

$$d_1(\tilde{V}_T^*, V_T^*) = O_p\left((1 \vee |\hat{\rho}|^{2T})T^2(\hat{\rho}-1)^2 d_2(\varepsilon_1, \varepsilon_1^*)\right), \tag{67}$$

$$d_1(\tilde{U}_T^*, U_T^*) = O_p\left((1 \vee |\hat{\rho}|^{T})T(\hat{\rho}-1)^2 d_2(\varepsilon_1, \varepsilon_1^*)\right), \tag{68}$$

$$d_1((\tilde{U}_T^*, \tilde{V}_T^*), (U_T^*, V_T^*)) \leq \text{RHS of (67)} + \text{RHS of (68)}. \tag{69}$$

Proof of Lemma 6. The proof of Lemma 6 is analogous to that of Lemma 1 and thus is omitted. ∎

Proof of Corollary 2. The proof of Corollary 2 follows from Theorem 2 by the delta method. ∎

Proof of Theorem 3. The proof of Theorem 3 is analogous to the proof of Theorem 1 with (21) and (22) replaced by

$$\Upsilon_T^{-1} \sum_{t=1}^{T} \bar{z}_{T,t} \bar{z}'_{T,t} \Upsilon_T^{-1} \overset{L}{\to} \left[\begin{array}{cc} 1 & \int_0^1 S(r, \gamma_0)dr \\ \int_0^1 S(r, \tilde{\gamma}_0)dr & \int_0^1 S(r, \gamma_0)^2 dr \end{array}\right] \tag{70}$$

28

$$\Upsilon_T^{-1} \sum_{t=1}^{T} \bar{z}_{T,t} \varepsilon_t \quad \xrightarrow{L} \quad \begin{bmatrix} S(1, \gamma_0) \\ \int_0^1 S(r, \gamma_0) dB(r) \\ \sigma \Gamma^{-\frac{1}{2}} W_{p-1} \end{bmatrix} \tag{71}$$

respectively, where $\bar{z}_{T,t} = (1, \psi_T(L)\varepsilon_t, (1-\rho_T L)y_{T,t-1}, (1-\rho_T L)y_{T,t-2}, ..., (1-\rho_T L)y_{T,t-p+1})'$. Applications of the continuous mapping theorem to equations 35, 36 and 37 of Jeganathan (1991) yield (70) and (71). ■

Proof of Corollary 3. The proof of Corollary 3 is analogous to the proof of Corollary 1 and thus is omitted. ■

References

Basawa, I.V., A.K. Mallik, W.P. McCormick and R.L. Taylor (1989), "Bootstrapping Explosive Autoregressive Processes," *Annals of Statistics,* 17, 1479–1486.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991a), "Bootstrapping Unstable First-Order Autoregressive Processes," *Annals of Statistics,* 19, 1098–1101.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991b), "Bootstrap Test of Significance and Sequential Bootstrap Estimation for Unstable First Order Autoregressive Processes," *Communications in Statistical Theory and Methods,* 20, 1015–1026.

Bickel, P.J. and D.A. Freedman (1981), "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics,* 9, 1196–1217.

Blough, S.R. (1992), "The Relationship between Power and Level for Generic Unit Root Tests in Finite Samples," *Journal of Applied Econometrics,* 7, 295–308.

Bose, A. (1988), "Edgeworth Correction by Bootstrap in Autoregression," *Annals of Statistics,* 16, 1709–1722.

Chan, N.H. and C.Z. Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics,* 15, 1050–1063.

Cochrane, J.H. (1991), "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control,* 15, 275–284.

Datta, S. (1995), "Limit Theory and Bootstrap for Explosive and Partially Explosive Autoregression," *Stochastic Processes and Their Applications,* 57, 285–304.

Datta, S. (1996), "On Asymptotic Properties of Bootstrap for AR(1) Processes," *Journal of Statistical Planning and Inference,* 53, 361–374.

Datta, S. and T.N. Sriram (1997), "A Modified Bootstrap for Autoregressions without Stationarity," *Journal of Statistical Planning and Inference,* 59, 19–30.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica,* 49, 1057–1072.

Elliott, G. (1998), "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica,* 66, 149–158.

Faust, J. (1996), "Near Observational Equivalence and Theoretical Size Problem with Unit Root Tests," *Econometric Theory,* 12, 724–731.

Giné, E. and J. Zinn (1990), "Bootstrapping General Empirical Measures," *Annals of Probability,* 18, 851–869.

Hansen, B.E. (1999), "The Grid Bootstrap and the Autoregressive Model," forth-coming: *Review of Economics and Statistics.*

Hasza, D.P. and W.A. Fuller (1979), "Estimation for Autoregressive Processes With Unit Roots," *Annals of Statistics,* 7, 1106–1120.

Heimann, G. and J-P. Kreiss (1996), "Bootstrapping General First-Order Autore-gression," *Statistics and Probability Letters*, 30, 87–98.

Jeganathan, P. (1991), "On the Asymptotic Behavior of Least-Squares Estimators in AR Time Series with Roots Near the Unit Circle," *Econometric Theory,* 7, 269–306.

Kilian, L. (1998), "Accounting for Lag Order Uncertainty in Autoregressions: The Endogenous Lag Order Bootstrap Algorithm," *Journal of Time Series Analysis,* 19, 531–548.

Lai, T.L. and C.Z. Wei (1983), "Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters," *Journal of Multivariate Analysis,* 13, 1–23.

Parthasarathy, K.R. (1967), *Probability Measures on Metric Spaces*, Academic Press: New York.

Paulsen, J. (1984), "Order Determination of Multivariate Autoregressive Time Series with Unit Roots," *Journal of Time Series Analysis,* 5, 115–127.

Romano, J.P. and M. Wolf (1998), "Subsampling Confidence Intervals for the Autoregressive Root," manuscript, Department of Statistics, Stanford University.

Sims, C.A., J.H. Stock and M.W. Watson (1990), "Inference in Linear Time Series with Some Unit Roots," *Econometrica,* 58, 113–144.

Tsay, R.S. (1984), "Order Selection in Nonstationary Autoregressive Models," *Annals of Statistics,* 12, 1425–1433.

Vaart, A.W. van der and J.A. Wellner (1996), *Weak Convergence and Empirical Processes,* Springer-Verlag: New York.

West, K.D. (1988), "Asymptotic Normality when Regressors Have a Unit Root," *Econometrica,* 56, 1397-1418.

Zheng, H.-M. (1992), "A Log Log Law for Unstable ARMA Models with Applica-tions to Time Series Analysis," *Journal of Multivariate Analysis,* 40, 173–204.

Table 1.
Percentiles of the True Finite-Sample Distribution and Average Bootstrap Percentiles
Regression Model: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$

(a) DGP: $y_t = 0.8 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.71	-1.37	-0.94	-0.11	0.76	1.21	1.58
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.70	-1.35	-0.93	-0.10	0.76	1.21	1.59
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.66	-1.31	-0.88	-0.06	0.78	1.23	1.59
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.30	-0.88	-0.06	0.78	1.22	1.59
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.64	-1.30	-0.87	-0.05	0.79	1.23	1.59
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.64	-1.29	-0.87	-0.05	0.79	1.23	1.59
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.80	-1.42	-0.97	-0.10	0.73	1.16	1.49
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.80	-1.42	-0.97	-0.11	0.72	1.14	1.48
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.71	-1.34	-0.90	-0.06	0.76	1.19	1.54
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.72	-1.35	-0.90	-0.07	0.75	1.17	1.52
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.69	-1.33	-0.89	-0.04	0.78	1.20	1.55
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.69	-1.32	-0.89	-0.05	0.77	1.19	1.54

(b) DGP: $y_t = 0.79 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.71	-1.36	-0.93	-0.10	0.76	1.21	1.59
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.70	-1.35	-0.93	-0.09	0.76	1.22	1.59
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.66	-1.30	-0.88	-0.06	0.79	1.23	1.60
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.30	-0.88	-0.06	0.78	1.23	1.60
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.64	-1.29	-0.87	-0.05	0.79	1.23	1.59
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.64	-1.29	-0.87	-0.04	0.79	1.23	1.60
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.81	-1.43	-0.98	-0.12	0.71	1.14	1.47
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.81	-1.43	-0.98	-0.12	0.70	1.13	1.46
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.73	-1.36	-0.91	-0.08	0.75	1.17	1.52
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.72	-1.36	-0.91	-0.08	0.74	1.16	1.51
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.70	-1.34	-0.90	-0.06	0.77	1.19	1.53
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.70	-1.33	-0.89	-0.06	0.76	1.18	1.53

(c) DGP: $y_t = 0.78 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.70	-1.36	-0.93	-0.10	0.76	1.22	1.59
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.70	-1.35	-0.92	-0.09	0.76	1.22	1.60
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.66	-1.30	-0.88	-0.06	0.79	1.23	1.60
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.30	-0.88	-0.05	0.78	1.23	1.60
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.64	-1.29	-0.87	-0.05	0.79	1.23	1.59
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.64	-1.29	-0.87	-0.04	0.79	1.23	1.60
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.82	-1.44	-0.99	-0.13	0.70	1.13	1.46
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.81	-1.44	-0.98	-0.13	0.70	1.12	1.46
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.73	-1.36	-0.92	-0.08	0.74	1.17	1.52
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.73	-1.36	-0.92	-0.08	0.74	1.16	1.50
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.70	-1.34	-0.90	-0.06	0.76	1.19	1.53
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.70	-1.33	-0.89	-0.06	0.76	1.18	1.53

NOTES: Percentiles of true distribution based on 100000 Monte Carlo trials. Bootstrap percentiles based on average of 100 trials with 20000 bootstrap replications each.

Table 2.
Percentiles of the True Finite-Sample Distribution and Average Bootstrap Percentiles
Regression Model: $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$

(a) DGP: $y_t = 1 + 0.8 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.70	-1.37	-0.96	-0.15	0.69	1.14	1.51
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.67	-1.34	-0.94	-0.14	0.69	1.14	1.51
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.67	-1.32	-0.90	-0.08	0.75	1.19	1.56
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.31	-0.89	-0.08	0.75	1.19	1.56
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.65	-1.31	-0.89	-0.07	0.77	1.21	1.57
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.64	-1.30	-0.88	-0.06	0.77	1.20	1.57
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.52	-1.14	-0.70	0.14	0.95	1.36	1.69
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.52	-1.15	-0.70	0.13	0.93	1.33	1.66
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.56	-1.19	-0.75	0.08	0.90	1.32	1.66
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.56	-1.19	-0.75	0.08	0.89	1.31	1.65
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.57	-1.21	-0.77	0.07	0.89	1.30	1.64
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.57	-1.20	-0.77	0.06	0.88	1.30	1.64

(b) DGP: $y_t = 1 + 0.79 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.71	-1.38	-0.96	-0.15	0.69	1.14	1.51
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.67	-1.35	-0.94	-0.14	0.69	1.14	1.51
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.67	-1.32	-0.90	-0.09	0.75	1.19	1.55
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.31	-0.89	-0.08	0.75	1.19	1.55
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.65	-1.31	-0.89	-0.07	0.77	1.20	1.56
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.30	-0.88	-0.07	0.76	1.20	1.57
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.51	-1.14	-0.70	0.14	0.94	1.35	1.67
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.51	-1.14	-0.70	0.12	0.92	1.32	1.64
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.55	-1.19	-0.75	0.07	0.88	1.30	1.64
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.55	-1.19	-0.75	0.07	0.88	1.29	1.63
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.56	-1.20	-0.77	0.06	0.87	1.29	1.63
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.56	-1.20	-0.77	0.05	0.86	1.28	1.62

(c) DGP: $y_t = 1 + 0.78 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.71	-1.38	-0.97	-0.16	0.69	1.13	1.50
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.68	-1.35	-0.95	-0.15	0.69	1.13	1.51
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.67	-1.33	-0.90	-0.09	0.75	1.19	1.55
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.66	-1.31	-0.90	-0.09	0.74	1.18	1.55
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.65	-1.31	-0.89	-0.07	0.77	1.20	1.56
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.65	-1.31	-0.89	-0.07	0.76	1.20	1.56
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.51	-1.14	-0.71	0.13	0.92	1.33	1.65
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.51	-1.14	-0.71	0.11	0.90	1.30	1.61
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.55	-1.20	-0.76	0.06	0.86	1.28	1.62
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.55	-1.19	-0.76	0.06	0.85	1.26	1.60
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.57	-1.21	-0.79	0.04	0.85	1.26	1.60
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.57	-1.21	-0.78	0.04	0.84	1.26	1.60

NOTES: Percentiles of true distribution based on 100000 Monte Carlo trials. Bootstrap percentiles based on average of 100 trials with 20000 bootstrap replications each.

Table 3.

Percentiles of the True Finite-Sample Distribution and Average Bootstrap Percentiles

Regression Model: $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$

(a) DGP: $y_t = 1 + y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{3/2}(\hat{\phi}_1 - \phi_1)$	-6.461	-5.114	-3.549	-0.599	2.334	3.883	5.196
	$T^{3/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-6.561	-5.205	-3.605	-0.621	2.350	3.927	5.258
$T = 300$	$T^{3/2}(\hat{\phi}_1 - \phi_1)$	-6.079	-4.806	-3.271	-0.346	2.595	4.124	5.373
	$T^{3/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-6.151	-4.856	-3.307	-0.353	2.598	4.156	5.446
$T = 500$	$T^{3/2}(\hat{\phi}_1 - \phi_1)$	-6.019	-4.757	-3.211	-0.274	2.673	4.219	5.516
	$T^{3/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-6.011	-4.731	-3.196	-0.268	2.664	4.208	5.483

(b) DGP: $y_t = 1 + 0.99 y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-0.112	-0.088	-0.061	-0.013	0.032	0.055	0.074
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-0.115	-0.091	-0.063	-0.013	0.033	0.056	0.075
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-0.079	-0.062	-0.042	-0.008	0.023	0.038	0.050
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-0.081	-0.063	-0.043	-0.008	0.023	0.039	0.051
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-0.083	-0.065	-0.044	-0.008	0.024	0.039	0.052
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-0.084	-0.066	-0.045	-0.008	0.024	0.040	0.052

(c) DGP: $y_t = 1 + 0.98 y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-0.182	-0.142	-0.098	-0.024	0.042	0.073	0.099
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-0.188	-0.147	-0.102	-0.025	0.043	0.076	0.102
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-0.183	-0.143	-0.098	-0.023	0.041	0.071	0.094
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-0.188	-0.147	-0.101	-0.024	0.043	0.074	0.098
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-0.206	-0.161	-0.111	-0.025	0.048	0.082	0.108
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-0.209	-0.164	-0.113	-0.026	0.049	0.084	0.111

NOTES: Percentiles of true distribution based on 100000 Monte Carlo trials. Bootstrap percentiles based on average of 100 trials with 20000 bootstrap replications each.

Table 4.
Percentiles of the True Finite-Sample Distribution and Average Bootstrap Percentiles

Regression Model: $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$

(a) DGP: $y_t = 0.8y_{t-1} + 0.2y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-2.07	-1.71	-1.26	-0.41	0.47	0.94	1.32
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.88	-1.52	-1.09	-0.24	0.62	1.08	1.46
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.86	-1.50	-1.07	-0.24	0.61	1.06	1.43
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.76	-1.41	-0.98	-0.15	0.69	1.14	1.51
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.79	-1.44	-1.02	-0.19	0.66	1.10	1.46
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.72	-1.37	-0.95	-0.12	0.72	1.16	1.53
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.93	-1.54	-1.09	-0.23	0.61	1.03	1.37
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.83	-1.46	-1.01	-0.17	0.65	1.06	1.40
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.79	-1.42	-0.97	-0.14	0.69	1.12	1.46
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.75	-1.38	-0.94	-0.11	0.71	1.13	1.47
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.74	-1.38	-0.94	-0.10	0.72	1.15	1.49
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.72	-1.35	-0.91	-0.08	0.74	1.16	1.51

(b) DGP: $y_t = 0.79y_{t-1} + 0.2y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-2.05	-1.69	-1.24	-0.38	0.51	0.97	1.36
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.89	-1.54	-1.10	-0.25	0.62	1.08	1.47
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.82	-1.45	-1.02	-0.19	0.66	1.11	1.48
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.77	-1.42	-0.99	-0.16	0.69	1.14	1.51
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.74	-1.39	-0.97	-0.14	0.71	1.15	1.51
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.73	-1.38	-0.95	-0.12	0.72	1.16	1.53
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.96	-1.59	-1.14	-0.28	0.56	0.98	1.32
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.87	-1.50	-1.05	-0.21	0.61	1.03	1.37
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.83	-1.46	-1.01	-0.18	0.65	1.07	1.42
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.80	-1.43	-0.99	-0.15	0.67	1.09	1.43
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.77	-1.41	-0.97	-0.13	0.69	1.11	1.46
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.76	-1.40	-0.96	-0.12	0.70	1.12	1.47

(c) DGP: $y_t = 0.78y_{t-1} + 0.2y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

		5%	10%	20%	50%	80%	90%	95%
$T = 100$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-2.02	-1.66	-1.22	-0.35	0.55	1.01	1.39
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.91	-1.54	-1.10	-0.25	0.62	1.09	1.47
$T = 300$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.79	-1.43	-1.00	-0.16	0.69	1.13	1.50
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.77	-1.41	-0.98	-0.15	0.70	1.14	1.52
$T = 500$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.72	-1.38	-0.95	-0.12	0.73	1.17	1.53
	$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.72	-1.37	-0.94	-0.11	0.73	1.17	1.54
$T = 100$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.98	-1.60	-1.16	-0.30	0.53	0.95	1.30
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.90	-1.53	-1.08	-0.24	0.59	1.01	1.35
$T = 300$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.82	-1.46	-1.01	-0.18	0.65	1.08	1.43
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.81	-1.44	-1.00	-0.16	0.66	1.08	1.42
$T = 500$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.78	-1.41	-0.97	-0.13	0.69	1.12	1.47
	$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.76	-1.40	-0.96	-0.12	0.70	1.12	1.47

NOTES: Percentiles of true distribution based on 100000 Monte Carlo trials. Bootstrap percentiles based on average of 100 trials with 20000 bootstrap replications each.

Table 5.

Percentiles of the True Finite-Sample Distribution and Average Bootstrap Percentiles

Regression Model: $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$

DGP: $y_t = \alpha + 0.798 y_{t-1} + 0.2 y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0,1)$.

			5%	10%	20%	50%	80%	90%	95%
$T = 500$	$\alpha = 0$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.78	-1.43	-1.00	-0.17	0.68	1.12	1.47
		$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.73	-1.38	-0.95	-0.12	0.71	1.16	1.53
	$\alpha = 0.1$	$T^{1/2}(\hat{\phi}_1 - \phi_1)$	-1.71	-1.36	-0.94	-0.12	0.73	1.16	1.52
		$T^{1/2}(\hat{\phi}_1^* - \hat{\phi}_1)$	-1.70	-1.35	-0.93	-0.11	0.73	1.17	1.53
$T = 500$	$\alpha = 0$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.76	-1.40	-0.96	-0.12	0.70	1.13	1.47
		$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.73	-1.37	-0.93	-0.10	0.72	1.14	1.49
	$\alpha = 0.1$	$T^{1/2}(\hat{\phi}_2 - \phi_2)$	-1.65	-1.30	-0.86	-0.02	0.81	1.23	1.57
		$T^{1/2}(\hat{\phi}_2^* - \hat{\phi}_2)$	-1.66	-1.29	-0.86	-0.03	0.79	1.21	1.56

NOTES: Percentiles of true distribution based on 100000 Monte Carlo trials. Bootstrap percentiles based on average of 100 trials with 20000 bootstrap replications each.