Econometrica, Vol. 70, No. 1 (January, 2002), 377–391

BOOTSTRAPPING AUTOREGRESSIVE PROCESSES WITH POSSIBLE UNIT ROOTS

BY ATSUSHI INOUE AND LUTZ KILIAN[1]

1. INTRODUCTION

SEVERAL AUTHORS HAVE INVESTIGATED the asymptotic properties of the standard residual-based bootstrap method for unrestricted autoregressions in the random walk model (see Basawa et al. (1991a), Datta (1996)). In contrast, there are no theoretical results for the properties of the bootstrap when the true model is a higher-order autoregressive process that is integrated of order one. The main contribution of this paper is to show that in the latter case the bootstrap achieves the correct first-order asymptotic distribution for the non-unit root parameters in the augmented Dickey-Fuller (ADF) representation, but not for the estimated unit root parameter (nor for the deterministic regressors). This result is new because the presence of the estimated unit root parameter invalidates conventional arguments for the asymptotic validity of the bootstrap approach such as the sufficiency conditions presented by Beran and Ducharme (1991, Proposition 1.3).

Our results apply not only for bootstrapping models with known zero intercept, but also in the practically more important case of an unknown intercept and possibly unknown linear time trend. Allowing for deterministic regressors introduces additional complications into the theoretical arguments. It may also affect the properties of the bootstrap estimator. For example, if the population model has a drift, the rate of convergence of the bootstrap estimator of the unit root parameter in the regression model with time trend is $T^{3/2}$ compared to T for the ordinary least-squares (OLS) estimator. This difference is due to the presence of the drift term, which induces a linear trend in the population model, but a nonlinear trend in the bootstrap model. The latter dominates the asymptotic behavior of the bootstrap estimator. This nonstandard result disappears if the drift is zero.

A direct implication of our main result is that the standard residual-based bootstrap algorithm will recover the correct first-order asymptotic distribution of the slope parameters in the level representation of the autoregression, provided that the model includes more than one lag. Sims, Stock, and Watson (1990) establish the asymptotic normality of the OLS estimator of the level slope parameters in higher-order autoregressive models. We show that the bootstrap estimator of the individual level slope parameters converges to the same normal limit distribution as the OLS estimator. Thus, the standard bootstrap approximation remains asymptotically valid for these parameters, even in the presence of a unit root. This fact is important for applied work because it alleviates, at least asymptotically, the need for unit root pre-tests.

The intuition for this result is as follows. Note that the level slope parameters of higher-order autoregressions can be expressed as linear combinations of the unit root parameter and of coefficients on lagged differences. We show that the bootstrap estimator of the latter coefficients is always $\sqrt{T}$ consistent with a Gaussian limit distribution. Although the bootstrap estimator of the unit root parameter under our assumptions converges to a

[1] We thank Mehmet Caner, Jinyong Hahn, Alastair Hall, Bruce Hansen, Shinichi Sakata, the co-editor, and two anonymous referees for helpful comments on earlier drafts of the paper.

random distribution, it converges faster than the coefficients on the lagged differences and its influence on the distribution of the level slope parameters is asymptotically negligible. This fact ensures the $\sqrt{T}$ consistency and asymptotic normality of the bootstrap estimator of the individual autoregressive slope parameters in the level representation of the process.

Our results also imply the first-order asymptotic validity of the bootstrap for the distribution of linear and smooth nonlinear functions of the level slope parameters in higher-order models, provided the limit distribution is nondegenerate. For certain linear and nonlinear functions, however, the asymptotic validity of the bootstrap approximation breaks down. For the latter case, several modified bootstrap algorithms of varying degrees of generality and finite-sample accuracy have been proposed by Basawa et al. (1991b), Datta (1996), Heimann and Kreiss (1996), and Datta and Sriram (1997), among others. Related work also includes Hansen (1999).

The remainder of the paper is organized as follows. Section 2 contains the main results. Section 3 concludes with further discussion and some Monte Carlo evidence. The proofs are in the Appendix.

2. BOOTSTRAP ASYMPTOTIC THEORY

In this section, we analyze the asymptotic properties of the residual-based bootstrap method for higher-order autoregressive processes with an exact unit root in the autoregressive lag-order polynomial. For expository purposes, we focus on the case of autoregressions with drift when the regression model includes an intercept and a linear time trend. Similar results may be obtained for other higher-order models with a unit root. We will briefly summarize these results in Section 3.

We begin with a review of the standard residual-based bootstrap procedure for autoregressions. Let $(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\alpha}, \hat{\delta})'$ denote the ordinary least squares estimator of $(\phi_1, \ldots, \phi_p, \alpha, \delta)'$ in the scalar AR(p) process $\phi(L)y_t = \alpha + \delta t + \varepsilon_t$ where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$. Let $(\hat{\phi}_1^*, \ldots, \hat{\phi}_p^*, \hat{\alpha}^*, \hat{\delta}^*)'$denote the corresponding bootstrap OLS estimator of $(\phi_1, \ldots, \phi_p, \alpha, \delta)'$. We describe the bootstrap algorithm for the regression model with intercept and deterministic time trend. If the regression model does not include a deterministic time trend, the same algorithm applies with $\delta = 0$ imposed throughout.

ALGORITHM: Let $\hat{\varepsilon}_t = \hat{\phi}(L)y_t - \hat{\alpha} - \hat{\delta}t$ denote the residuals of the OLS estimate of the process $\phi(L)y_t = \alpha + \delta t + \varepsilon_t$ given $\{y_t\}_{t=1-p}^T$. $\hat{\varepsilon}_t$ is mean zero by construction. Let $\widehat{F}_T$ denote the empirical distribution function of $\hat{\varepsilon}_t$. $\widehat{F}_T$ associates probability mass T^{-1} with $\hat{\varepsilon}_t, t = 1, \ldots, T$. Treating $\widehat{F}_T$ as the bootstrap population distribution, random samples $\{\varepsilon_t^*\}_{t=1}^T$ may be drawn from $\widehat{F}_T$. Thus, conditional on the data, the random variable ε_t^* is iid with distribution function $\widehat{F}_T$. Now construct the bootstrap sample $\{y_t^*\}_{t=1-p}^T$ recursively from $\hat{\phi}(L)y_t^* = \hat{\alpha} + \hat{\delta}t + \varepsilon_t^*$, given initial values $y_0^* = y_0, \ldots, y_{1-p}^* = y_{1-p}$. The bootstrap estimator $(\hat{\phi}_1^*, \ldots, \hat{\phi}_p^*)$ then may be obtained by OLS from $\{y_t^*\}_{t=1-p}^T$.

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

Consider an AR(p) data generating process (DGP)

$$\phi(L)y_t = \alpha + \delta t + \varepsilon_t, \tag{1}$$

which satisfies the following set of assumptions.

ASSUMPTION 1: (a) $\phi(z) = (1-z)\zeta(z)$ *with* $\zeta(z) = 1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_{p-1} z^{p-1}$ *and* $|\zeta(z)| \neq 0$ *for all* $|z| \leq 1$.

(b) $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$ *with* $E|\varepsilon_t|^{2+\kappa} < \infty$ *for some* $\kappa > 0$ *and* ε_t *is independent of* $y_0, y_1, \ldots, y_{1-p}$ *for* $t \geq 1$.

(c) *The distribution of* ε_t, F, *satisfies* $\sup_x |F'(x)| < \infty$.

(d) $y_j = 0_{as}(T^{1/2})$ *for* $j = 0, -1, -2, \ldots, 1-p$.

(e) $\alpha \neq 0$ *and* $\delta = 0$.

The process y_t can be written as

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

where $\rho = \phi_1 + \phi_2 + \cdots + \phi_p = 1$,

$$\beta = \begin{cases} (\rho, (1-\rho)\mu, \delta + \rho\mu)' & \text{if } p = 1, \\ (\rho, \zeta_1, \zeta_2, \ldots, \zeta_{p-1}, (1-\rho)\mu, \delta + \rho\mu)' & \text{if } p > 1, \end{cases}$$

$$x_t = \begin{cases} (y_{t-1} - \mu(t-1), 1, t)' & \text{if } p = 1, \\ (y_{t-1} - \mu(t-1), \Delta y_{t-1} - \mu, \\ \quad \Delta y_{t-2} - \mu, \ldots, \Delta y_{t-p+1} - \mu, 1, t)' & \text{if } p > 1, \end{cases}$$

$$\mu = \begin{cases} \alpha & \text{if } p = 1, \\ \alpha/(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1}) & \text{if } p > 1. \end{cases}$$

For notational convenience, we will often replace $p-1$ by q. Let $\hat{\beta}$ denote the OLS estimator of β and

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

It is well known that for $p = 1$

$$\Upsilon_T(\hat{\beta} - \beta) \xrightarrow{L} D_1 \begin{bmatrix} \int_0^1 B(r)^2 dr & \int_0^1 B(r) dr & \int_0^1 r B(r) dr \\ \int_0^1 B(r) dr & 1 & \frac{1}{2} \\ \int_0^1 r B(r) dr & \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \times \begin{bmatrix} \frac{1}{2}(B(1)^2 - 1) \\ B(1) \\ B(1) - \int_0^1 B(r) dr \end{bmatrix} \tag{3}$$

and that for $p > 1$

$$\Upsilon_T(\hat{\beta}-\beta) \xrightarrow{L} D_2 \begin{bmatrix} \int_0^1 B(r)^2 dr & 0_{1\times q} & \int_0^1 B(r)dr & \int_0^1 rB(r)dr \\ 0_{q\times 1} & \Gamma & 0_{q\times 1} & 0_{q\times 1} \\ \int_0^1 B(r)dr & 0_{1\times q} & 1 & \frac{1}{2} \\ \int_0^1 rB(r)dr & 0_{1\times q} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \times \begin{bmatrix} \frac{1}{2}(B(1)^2-1) \\ \Gamma^{\frac{1}{2}} W_q(1) \\ B(1) \\ B(1)-\int_0^1 B(r)dr \end{bmatrix}, \tag{4}$$

where B is a Brownian motion, W_q is a q-dimensional standard normal random vector, B and W_q are independent, $\Gamma = \{\mathrm{Cov}(\Delta y_{t-i}, \Delta y_{t-j})\}_{i,j=1,2,\ldots,q}$,

$$D_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{bmatrix}, \qquad D_2 = \begin{bmatrix} \zeta(1) & 0_{1\times q} & 0 & 0 \\ 0_{q\times 1} & \sigma I_q & 0_{q\times 1} & 0_{q\times 1} \\ 0 & 0_{1\times q} & \sigma & 0 \\ 0 & 0_{1\times q} & 0 & \sigma \end{bmatrix},$$

and $\zeta(1) = 1-\zeta_1-\zeta_2-\cdots-\zeta_{p-1}$. Let $\hat{\beta}^*$ denote the bootstrap OLS estimator of $\hat{\beta}$.

The standard bootstrap results for autoregressions by Bose (1988) do not cover this model. In Theorem 1, we prove that the standard bootstrap is valid for the non-unit-root parameters ζ_j, $j = 1, \ldots, p-1$, in the ADF representation (2), but not for ρ.

THEOREM 1(Asymptotic Properties of the Bootstrap in Integrated AR(p) Processes with Drift when the Regression Model Includes an Intercept and a Linear Time Trend): *Suppose that Assumption* 1 *holds*. *For* $p = 1$,

$$\widetilde{\Upsilon}_T(\hat{\beta}^*-\hat{\beta}) \xrightarrow{L} D_1 \begin{bmatrix} \int_0^1 \overline{S}(r,\tilde{\gamma}_0)^2 dr & \int_0^1 \overline{S}(r,\tilde{\gamma}_0)dr & \int_0^1 r\overline{S}(r,\tilde{\gamma}_0)dr \\ \int_0^1 \overline{S}(r,\tilde{\gamma}_0)dr & 1 & \frac{1}{2} \\ \int_0^1 r\overline{S}(r,\tilde{\gamma}_0)dr & \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \times \begin{bmatrix} \int_0^1 \overline{S}(r,\tilde{\gamma}_0)dB(r) \\ \overline{S}(1,\tilde{\gamma}_0) \\ \overline{S}(1,\tilde{\gamma}_0)-\int_0^1 \overline{S}(r,\tilde{\gamma}_0)dr \end{bmatrix} \quad \omega\text{-a.s.} \tag{5}$$

and for $p > 1$

$$\text{(6)} \qquad \widetilde{\Upsilon}_T(\hat{\beta}^* - \hat{\beta}) \xrightarrow{L} D_2 \begin{bmatrix} \int_0^1 \overline{S}(r,\tilde{\gamma}_0)^2 dr & 0_{1\times q} & \int_0^1 \overline{S}(r,\tilde{\gamma}_0)dr & \int_0^1 r\overline{S}(r,\tilde{\gamma}_0)dr \\ 0_{q\times 1} & \Gamma & 0_{q\times 1} & 0_{q\times 1} \\ \int_0^1 \overline{S}(r,\tilde{\gamma}_0)dr & 0_{1\times q} & 1 & \frac{1}{2} \\ \int_0^1 r\overline{S}(r,\tilde{\gamma}_0)dr & 0_{1\times q} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \int_0^1 \overline{S}(r,\tilde{\gamma}_0)dB(r) \\ \Gamma^{\frac{1}{2}} W_q(1) \\ \overline{S}(1,\tilde{\gamma}_0) \\ \overline{S}(1,\tilde{\gamma}_0) - \int_0^1 \overline{S}(1,\tilde{\gamma}_0)dr \end{bmatrix} \quad \omega\text{-a.s.}$$

where $q = p - 1$,

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

$$\overline{S}(r,\tilde{\gamma}_0) = \alpha\left[\int_0^r \exp(\tilde{\gamma}_0 s)ds - r\right],$$

and $\tilde{\gamma}_0$ *in* (5) *and* (6) *denote the first elements of the random vectors on the right-hand side* (*RHS*) *of* (3) *and* (4), *respectively.*

The fact that the bootstrap fails to mimic the limiting distribution of the OLS estimator for ρ, but that it recovers the limiting distribution of the OLS estimator for the non-unit-root parameters $(\zeta_1, \zeta_2, \dots, \zeta_{p-1})'$ has important implications for the distribution of the level slope parameters of the autoregressive process. The slope parameters of the level representation (1) can be expressed as linear combinations of ρ and of $\zeta_i, i = 1, 2 \dots, p-1$. Specifically, $\phi_1 = \rho + \zeta_1$, $\phi_j = \zeta_j - \zeta_{j-1}$ for $j = 2, \dots, p-1$, 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, for $p > 1$ the bootstrap provides an asymptotically valid approximation to the marginal distribution of the autoregressive slope parameters in the level representation, even in the presence of a unit root. This result is generalized in Corollary 1.

COROLLARY 1 (Validity of the Bootstrap for Slope Parameters in Integrated AR(p) Processes with Drift when the Regression Model Includes an Intercept and a Linear Time Trend): *Consider a linear combination of slope parameters* $c'\phi$ *where* $c = (c_1, c_2, \dots, c_p)' \neq (\lambda, \lambda, \dots, \lambda)'$ *for all* λ *and* $\phi = (\phi_1, \phi_2, \dots, \phi_p)'$. *Let* $\hat{\phi}$ *and* $\hat{\phi}^*$ *denote the OLS estimator and the bootstrap estimator in Theorem* 1, *respectively. Suppose that Assumption* 1 *holds with* $p > 1$. *Then* (*a*):

$$\text{(7)} \qquad T^{1/2} c'(\hat{\phi} - \phi) \xrightarrow{d} N(0, c'\Omega c),$$

$$\text{(8)} \qquad T^{1/2} c'(\hat{\phi}^* - \hat{\phi}) \xrightarrow{d} N(0, c'\Omega c) \qquad \omega\text{-a.s.},$$

and (*b*):

$$T^{1/2}c'(\hat{\phi}-\phi)/(c'\widehat{\Omega}c)^{1/2} \xrightarrow{d} N(0,1), \tag{9}$$

$$T^{1/2}c'(\hat{\phi}^*-\hat{\phi})/(c'\widehat{\Omega}^*c)^{1/2} \xrightarrow{d} N(0,1) \qquad \omega\text{-a.s.}, \tag{10}$$

where

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

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

$\widehat{\Omega}$ *is* $\hat{\sigma}^2(1/T)\sum_{t=1}^T \hat{\varepsilon}_t^2$ *times the upper-left* $p\times p$ *submatrix of* $((1/T)\sum_{t=1}^T w_t w_t')^{-1}$, $\widehat{\Omega}^*$ *is* $\hat{\sigma}^{2*} = (1/T)\sum_{t=1}^T \hat{\varepsilon}_t^{2*}$ *times the upper-left* $p\times p$ *submatrix of* $((1/T)\sum_{t=1}^T w_t^* w_t^{*\prime})^{-1}$, *and*

$$w_t = \begin{cases} (y_{t-1}, 1, t)' & \text{if } p=1, \\ (y_{t-1}, y_{t-2}, \ldots, y_{t-p}, 1, t)' & \text{if } p>1, \end{cases}$$

$$w_t^* = \begin{cases} (y_{t-1}^*, 1, t)' & \text{if } p=1, \\ (y_{t-1}^*, y_{t-2}^*, \ldots, y_{t-p}^*, 1, t)' & \text{if } p>1. \end{cases}$$

Corollary 1 validates the application of the standard nonparametric bootstrap to individual slope parameters, ϕ_j, for $j = 1, 2, \ldots, p$, and to linear combinations of slope parameters except to those proportional to $\rho = \phi_1 + \cdots + \phi_p$. When $c = (\lambda, \lambda, \ldots, \lambda)'$ for some $\lambda \neq 0$, $c'\phi = \lambda\rho$ and thus the bootstrap will be invalid. This result parallels similar results for the OLS estimator by Sims, Stock, and Watson (1990) and West (1988). Our results also provide the basis for bootstrap inference on smooth nonlinear functions of $(\rho, \zeta_1, \ldots, \zeta_{p-1})'$ such as impulse responses and half-lives, provided the limiting distribution is nondegenerate.

Note that we do not claim to have solved the unit root problem. Rather Corollary 1 shows that the bootstrap invalidity for ρ is irrelevant for many statistics of interest. 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, similar to Datta's (1996) result for the AR(1) model without intercept. Specifically, $T(\hat{\rho}^* - \hat{\rho})$ will converge to a random distribution.

The ultimate purpose of our bootstrap procedure is the construction of bootstrap confidence intervals (see Efron and Tibshirani (1993) for further discussion). Corollary 2 states the conditions under which the percentile interval and the equal-tailed percentile-t interval yield asymptotically correct coverage probabilities.

COROLLARY 2 (Asymptotic Coverage Probabilities of Percentile and Percentile-t Intervals): *Suppose that Assumption* 1 *holds with* $p > 1$. *Consider a linear combination of slope parameters* $c'\phi$ *where* $c = (c_1, c_2, \ldots, c_p)' \neq (\lambda, \lambda, \ldots, \lambda)'$ *for all* λ *and* $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$. *Let* $\hat{\phi}$ *and* $\hat{\phi}^*$ *denote the OLS estimator and the bootstrap estimator in Theorem* 1, *respectively.*

(a) *Let* $q_T^*(a)$ *denote the* $100a\%$ *quantile of the distribution of* $c'\hat{\phi}^*$. *Define the* $100(1-a)\%$ *percentile bootstrap confidence region for* $c'\phi$ *by*

$$K_T = \{c'\phi \in \Re : q_T^*(a/2) \le c'\phi \le q_T^*(1-a/2)\}. \tag{11}$$

Then

$$\lim_{T\to\infty} P(c'\phi_0 \in K_T) = 1-a, \tag{12}$$

where ϕ_0 *denotes the true value of the parameter vector.*

(b) *Let* $t_T^*(a)$ *denote the* $100a\%$ *quantile of the distribution of the studentized estimator* $\sqrt{T}c'(\hat{\phi}^*-\hat{\phi})/(c'\widehat{\Omega}^*c)^{1/2}$ *and define the* $100(1-a)\%$ *percentile-t bootstrap confidence region for* $c'\phi$ *by*

$$\widetilde{K}_T = \left\{c'\phi \in \Re : t_T^*(a/2) \le \frac{\sqrt{T}c'(\hat{\phi}-\phi)}{(c'\widehat{\Omega}c)^{\frac{1}{2}}} \le t_T^*(1-a/2)\right\}. \tag{13}$$

Then

$$\lim_{T\to\infty} P(c'\phi_0 \in \widetilde{K}_T) = 1-a. \tag{14}$$

Extensions of this result to smooth nonlinear functions are straightforward, provided the limit distribution is nondegenerate.

3. DISCUSSION AND CONCLUSION

The theoretical results of Section 2 imply that in higher-order models the standard bootstrap algorithm will provide an asymptotically valid approximation to the marginal distribution of the autoregressive slope parameters $\phi_1, \ldots, \phi_p$ even in the presence of a unit root. This result is important because it alleviates, at least asymptotically, the need for unit root pre-tests. Analogous results hold for the population model without drift (whether the regression model includes an intercept, an intercept and deterministic time trend, or no deterministic regressors) with the important difference that the bootstrap estimator of ρ converges at the same rate T as the usual OLS estimator (see Inoue and Kilian (2000)). Finally, the bootstrap is also asymptotically valid for slope parameters in the population model with drift when the regression model includes an intercept but no time trend. In the latter case the bootstrap estimator of the unit root parameter converges at rate $T^{3/2}$.

In this section, we provide some preliminary Monte Carlo evidence of the accuracy of the proposed bootstrap approximation in finite samples and contrast it with the usual asymptotic first-order approximation. We focus on second-order autoregressive models as the leading example of higher-order autoregressive processes. Table I shows the effective coverage probabilities of nominal 90% percentile and percentile-t bootstrap confidence intervals for ϕ_1 and ϕ_2 in selected integrated and near integrated AR(2) models. Both intervals yield asymptotically correct coverage probabilities under our assumptions as shown in Corollary 2. Analogous results for the stationary case may be derived based on Bose (1988). We also include intervals based on the conventional asymptotic normal approximation.

The simulation results are based on 4000 Monte Carlo trials with 1000 bootstrap replications each. We begin with the case of models with a known zero intercept. Although the assumption of a known zero intercept is clearly unrealistic, it provides a useful benchmark. Panel (a) suggests that the first-order asymptotic approximation is highly accurate even for sample sizes as small as $T = 100$ and works equally well for integrated and near-integrated processes. There is little to choose between the bootstrap approach and the conventional approach.

As an intercept is added to the regression model, the finite-sample coverage accuracy of both the percentile interval and the conventional asymptotic interval deteriorates somewhat. Panel (b) shows that for $T = 100$ the coverage probabilities may drop as low as 0.84 for the percentile interval and 0.85 for the conventional interval. The reason is

TABLE I

EFFECTIVE COVERAGE RATES OF NOMINAL 90% BOOTSTRAP CONFIDENCE INTERVALS INTEGRATED AND NEAR-INTEGRATED AR(2) PROCESSES

(a) Regression Model without Deterministic Regressors

DGP: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

	$T=100$			$T=300$			$T=500$		
	PER	PER-t	ASY	PER	PER-t	ASY	PER	PER-t	ASY
$\phi_1 = 0.80$	0.91	0.90	0.90	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_2 = 0.20$	0.90	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_1 = 0.79$	0.91	0.90	0.90	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_2 = 0.20$	0.90	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_1 = 0.78$	0.91	0.90	0.90	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_2 = 0.20$	0.90	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90

(b) Regression Model with Intercept

DGP: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

	$T=100$			$T=300$			$T=500$		
	PER	PER-t	ASY	PER	PER-t	ASY	PER	PER-t	ASY
$\phi_1 = 0.80$	0.84	0.89	0.85	0.88	0.89	0.88	0.89	0.90	0.89
$\phi_2 = 0.20$	0.88	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_1 = 0.79$	0.85	0.89	0.86	0.88	0.90	0.89	0.90	0.90	0.89
$\phi_2 = 0.20$	0.87	0.90	0.89	0.89	0.89	0.89	0.89	0.89	0.90
$\phi_1 = 0.78$	0.86	0.89	0.87	0.89	0.90	0.89	0.89	0.90	0.89
$\phi_2 = 0.20$	0.87	0.90	0.89	0.88	0.89	0.89	0.89	0.89	0.90

(c) Regression Model with Intercept

DGP: $y_t = 1 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

	$T=100$			$T=300$			$T=500$		
	PER	PER-t	ASY	PER	PER-t	ASY	PER	PER-t	ASY
$\phi_1 = 0.80$	0.90	0.89	0.90	0.90	0.89	0.89	0.90	0.90	0.90
$\phi_2 = 0.20$	0.90	0.89	0.89	0.90	0.89	0.89	0.90	0.90	0.90
$\phi_1 = 0.79$	0.89	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_2 = 0.20$	0.90	0.89	0.90	0.90	0.89	0.89	0.90	0.90	0.90
$\phi_1 = 0.78$	0.87	0.89	0.87	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_2 = 0.20$	0.89	0.90	0.89	0.89	0.89	0.89	0.89	0.90	0.90

TABLE I—*Continued*

(d) Regression Model with Intercept and Linear Time Trend

DGP: $y_t = 1 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$

	$T = 100$			$T = 300$			$T = 500$		
	PER	PER-t	ASY	PER	PER-t	ASY	PER	PER-t	ASY
$\phi_1 = 0.80$	0.69	0.87	0.77	0.82	0.89	0.85	0.85	0.90	0.88
$\phi_2 = 0.20$	0.80	0.89	0.87	0.87	0.88	0.88	0.88	0.89	0.89
$\phi_1 = 0.79$	0.73	0.88	0.80	0.87	0.89	0.88	0.89	0.90	0.89
$\phi_2 = 0.20$	0.82	0.89	0.87	0.89	0.89	0.89	0.90	0.90	0.90
$\phi_1 = 0.78$	0.74	0.88	0.81	0.86	0.89	0.88	0.88	0.90	0.89
$\phi_2 = 0.20$	0.81	0.89	0.87	0.87	0.89	0.89	0.88	0.89	0.89

Notes: Simulation results based on 4000 Monte Carlo trials with 1000 bootstrap replications each. ASY: conventional asymptotic interval. PER: Efron's percentile interval. PER-t: equal-tailed percentile-t interval. For details, see Efron and Tibshirani (1993).

the increase in OLS small sample bias associated with the estimation of the intercept. As the sample size increases, this bias shrinks and coverage rates improve. We also find that for $T = 100$ the percentile interval is not working as well as the conventional asymptotic approximation. For larger sample sizes the two intervals are about equally accurate. This finding is not surprising. It is well known that the percentile interval as originally devised is more susceptible to OLS small-sample bias than the conventional asymptotic normal approximation (see Kilian (1998)). In contrast, the percentile-t interval is designed to allow for small-sample bias. This fact explains its consistently high coverage accuracy for all three sample sizes.

For the population model with drift in panel (c), the differences between the conventional asymptotic interval and percentile interval are minor, and both intervals are reasonably accurate for all but the smallest sample sizes. The reason is the faster rate of convergence of ρ in the unit root case. This difference in convergence rates also explains why the accuracy of the percentile and conventional asymptotic intervals actually improves for $T = 100$, as the root approaches unity. As in panel (b), the percentile-t interval works equally well for all parameter values and sample sizes.

The advantages of the percentile-t interval over both the conventional asymptotic interval and the percentile interval are most pronounced if the regression model also includes a linear time trend as in panel (d). Similar results hold if the model with trend is fitted to data from the population model without drift. The reason is the further increase in small-sample bias that comes with the inclusion of a deterministic time trend in the regression. For example, in the first row of panel (d) the coverage probability of the percentile-t interval for $T = 100$ is 0.87 compared with only 0.69 for the percentile interval and 0.77 for the delta method. For larger sample sizes the differences in coverage accuracy are less pronounced, but the percentile-t interval is the only interval with consistently high coverage accuracy.

We tentatively conclude that the percentile-t interval works quite well in nearly all cases analyzed. This result is in sharp contrast to results for the parameter ϕ_1 in the AR(1) model. In the latter case, the coverage accuracy of the percentile-t interval tends to deteriorate sharply, not just in the random walk case, but also for models with roots near the unit circle (see Hansen (1999), Kilian (1999)). The percentile interval also is more accurate than in the AR(1) model, but may work poorly in small samples especially for the model with fitted time trend. In fact, its accuracy is often inferior to that of the conventional asymptotic normal approximation. Additional simulation evidence will be needed to assess the practical relevance of the theory we presented for nonlinear transformations of level slope parameters such as impulse responses and to explore further the relative merits of alternative bootstrap confidence intervals (see Kilian (1999) for related discussion).

There are two natural extensions of our analysis. One extension involves generalizing the results of this paper to possibly integrated and/or cointegrated 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). Both extensions are nontrivial and more appropriately dealt with in a separate paper.

Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695-8109, U.S.A.; atsushi@unity.ncsu.edu

and

Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, U.S.A.; lkilian@umich.edu

Manuscript received November, 1999; final revision received November, 2000.

APPENDIX

PROOF OF THEOREM 1: We adopt the following additional notation. Let $\Rightarrow$ denote the weak convergence of probability measures and $[x]$ denote the integer part of x. $\ell_{p\times 1}$ is a p-dimensional column vector of ones and $I(\cdot)$ is the indicator function. $O_{p,m\times n}[o_{p,m\times n}]$ is an $m\times n$ matrix of order $O_p[o_p]$, and O_p and o_p are with respect to the bootstrap conditional probability unless noted otherwise. Let $\mathcal{M}(X)$ denote the space of probability measures on X topologized by weak convergence. Given an $m\times n$ matrix $A=\{a_{ij}\}$, for $\mu,\nu\in\Gamma_r=\{\gamma\in\mathcal{M}(\Re^s):\int\|x\|^r\gamma(dx)<\infty\}$, let $d_r(\mu,\nu)$ be the infimum of $(E\|X-Y\|^r)^{1/r}$, over all possible joint distributions of two 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, and $a\vee b$ denotes the maximum of a and b.

First, consider the triangular array:

$$y_{T,t}=\rho_T y_{T,t-1}+\zeta_{T,1}(1-\rho_T L)y_{T,t-1}+\cdots+\zeta_{T,p-1}(1-\rho_T L)y_{T,t-p+1}+\alpha_T+\delta_T t+\varepsilon_t, \tag{15}$$

where $\rho_T=1+\gamma_0T^{-1}+o(T^{-1}),\zeta_{T,j}=\zeta_j+\gamma_jT^{-1/2}+o(T^{-1/2})$ for $j=1,2,\dots,p-1,\alpha_T=\alpha+\gamma_pT^{-1/2}+o(T^{-1/2}),\delta_T=\gamma_{p+1}T^{-3/2}+o(T^{-3/2}),\gamma=(\gamma_0,\gamma_1,\dots,\gamma_{p+1})'$ is fixed, $y_{T,0}=y_0,y_{T,-1}=y_{-1},\dots,y_{T,1-p}=y_{1-p}$ and $\varepsilon_t\sim\text{iid}(0,\sigma^2)$ for $t\geq 1$. Without loss of generality, we assume $p>1$ throughout the proof.

Let

$$z_{T,t}=[\xi_{T,t-1}\quad u_{T,t-1}\cdots u_{T,t-p+1}\quad 1\quad t]',$$

$$\zeta_{T,t}=\zeta_T(L)y_{T,t}=\rho_T^t\xi_{T,0}+\sum_{j=0}^{t-1}\rho_T^j\varepsilon_{t-j}+\alpha_T\sum_{j=0}^{t-1}\rho_T^j+\delta_T\sum_{j=0}^{t-1}\rho_T^j(t-j)-\alpha_T t,$$

$$u_{T,t}=(1-\rho_T L)y_{T,t}-(\alpha_T+\delta_T t)/\zeta_T(L).$$

By Propositions 1, 4, and 5 of Jeganathan (1991),

$$T^{-\frac{1}{2}}\left[\rho_T^{[Tr]}\xi_{T,0}+\sum_{j=0}^{[Tr]-1}\rho_T^j\varepsilon_{[T_r]-j}+\delta_T\sum_{j=0}^{[Tr]-1}\rho_T^j(t-j)\right] \tag{16}$$

$$\xrightarrow{L} B(r)+\gamma_0\int_0^r\exp((r-s)\gamma_0)B(s)ds+\gamma_{p+1}\int_0^r\exp(s\gamma_0)(r-s)ds,$$

$$T^{-1}\left[\alpha_T \sum_{j=0}^{[Tr]-1} \rho_T^j - \alpha_T[Tr]\right] \to \alpha\left[\int_0^r \exp(\gamma_0 s)ds - r\right] \equiv \overline{S}(r,\ \gamma_0), \tag{17}$$

and thus

$$T^{-1}\xi_{T,[Tr]} \Rightarrow \overline{S}(r,\gamma_0). \tag{18}$$

By applications of the continuous mapping theorem

$$\widetilde{\Upsilon}_T^{-1}\sum_{t=1}^{T} z_{T,t}z'_{T,t}\widetilde{\Upsilon}_T^{-1} \xrightarrow{L} \begin{bmatrix} \sigma^2\int_0^1 \overline{S}(r,\ \gamma_0)^2 dr & 0_{1\times q} & \sigma\int_0^1 \overline{S}(r,\ \gamma_0)dr & \sigma\int_0^1 r\overline{S}(r,\ \gamma_0)dr \\ 0_{q\times 1} & \Gamma & 0_{q\times 1} & 0_{q\times 1} \\ \sigma\int_0^1 \overline{S}(r,\ \gamma_0)dr & 0_{1\times q} & 1 & \frac{1}{2} \\ \sigma\int_0^1 r\overline{S}(r,\ \gamma_0)dr & 0_{1\times q} & \frac{1}{2} & \frac{1}{3} \end{bmatrix}, \tag{19}$$

$$\widetilde{\Upsilon}_T^{-1}\sum_{t=1}^{T} z_{T,t}\varepsilon_t \xrightarrow{L} \begin{bmatrix} \sigma^2\int_0^1 \overline{S}(r,\ \gamma_0)dB(r) \\ \sigma\Gamma^{\frac{1}{2}}W_q \\ \sigma\overline{S}(1,\ \gamma_0) \\ \sigma(\overline{S}(1,\ \gamma_0) - \int_0^1 \overline{S}(r,\ \gamma_0)dr) \end{bmatrix}. \tag{20}$$

In other words,

$$\lim_{T\to\infty} P\left(\widetilde{\Upsilon}_T\left(\sum_{t=1}^{T} z_{T,t}z'_{T,t}\right)^{-1}\sum_{t=1}^{T} z_{T,t}\varepsilon_t \le x\right) = H(\gamma, x) \tag{21}$$

where $H(\gamma, x)$ is the joint distribution function of the $p+2$-dimensional random vector $V_0^{-1}U_0$ and V_0 and U_0 are the random vectors on the RHS of (19) and (20), respectively.

Next, we shall derive the limiting distribution of the bootstrap OLS estimator when ρ_T, $\zeta_{T,1}, \dots, \zeta_{T,p-1}, \alpha_T, \delta_T$ in (15) are replaced with their OLS estimates. Let $\varepsilon_t^* \overset{\text{iid}}{\sim} \widehat{F}_T$ and

$$z_t^* = [\xi_{t-1}^* \quad u_{t-1}^* \cdots u_{t-p+1}^* \quad 1 \quad t]'.$$

Let η_T denote a random measure on $\Re^{p+2}$ defined by

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

where S is a Borel set in $\Re^{p+2}$, $H_T(x) = P^*(\widetilde{\Upsilon}_T(\sum_{t=1}^T z_t^* z_t^{*'})^{-1}\sum_{t=1}^T z_t^*\varepsilon_t^* \le x)$, and P^* is the probability measure induced by the bootstrap conditional on the original data. In other words,

$$\eta_T(S) = P^*\left(\widetilde{\Upsilon}_T\left(\sum_{t=1}^{T} z_t^* z_t^{*'}\right)^{-1}\sum_{t=1}^{T} z_t^*\varepsilon_t^* \in S\right).$$

Noting that $H(\gamma, x)$ is continuous in γ for each fixed x, we define a random measure η on $\Re^{p+2}$ by

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

where $\tilde{\gamma}$ is the p-dimensional random vector given by the RHS of (4). We will show that, as $T \to \infty$

$$\eta_T \Rightarrow \eta \quad \omega\text{-a.s.} \tag{22}$$

As Datta (1996, Remark 2.2) points out, η_T is not a function of $\hat{\beta}$ alone, but also a function of $\widehat{F}_T$, and thus it is difficult to prove (22) 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 the population $\phi(L)y_t = \alpha + \delta t + \varepsilon_t$ where $\varepsilon_t \overset{\text{iid}}{\sim} F$. 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.,

$$\hat{\phi}(L)\tilde{y}_t^* = \hat{\alpha} + \hat{\delta}t + \tilde{\varepsilon}_t^* \qquad (t = 1, 2, \dots, T), \tag{23}$$

where $\tilde{\varepsilon}_t^* \overset{\text{iid}}{\sim} F$, and $\tilde{y}_0^* = y_0, \dots, \tilde{y}_{1-p}^* = y_{1-p}$. The bootstrap analogue of η_T under this resampling scheme is given by

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

where $\widetilde{H}_T(x) = P^*(\widetilde{\Upsilon}_T(\sum_{t=1}^T \tilde{z}_t^* \tilde{z}_t^{*\prime})^{-1} \sum_{t=1}^T \tilde{z}_t^* \tilde{\varepsilon}_t^* \le x)$,

$$\tilde{z}_t^* = \begin{bmatrix} \tilde{\xi}_{t-1}^* & \tilde{u}_{t-1}^* \cdots \tilde{u}_{t-p+1}^* & 1 & t \end{bmatrix}',$$

$$\tilde{\xi}_t^* = \hat{\zeta}(L)\tilde{y}_t^* = \hat{\rho}^t \tilde{\xi}_0^* + \sum_{j=0}^{t-1} \hat{\rho}^j \tilde{\varepsilon}_{t-j}^* + \hat{\alpha} \sum_{j=0}^{t-1} \hat{\rho}^j + \hat{\delta} \sum_{j=0}^{t-1} \hat{\rho}^j (t-j) - \hat{\alpha}t,$$

$$\tilde{u}_t^* = (1 - \hat{\rho}L)\tilde{y}_t^* - (\hat{\alpha} + \hat{\delta}t)/\hat{\zeta}(L).$$

Next, consider the standard bootstrap resampling scheme

$$\hat{\phi}(L)y_t^* = \hat{\alpha} + \hat{\delta}t + \varepsilon_t^* \qquad (t = 1, 2, \dots, T),$$

where $y_0^* = y_0, \dots, y_{1-p}^* = y_{1-p}$. Let

$$\xi_t^* = \hat{\zeta}(L)y_t^* = \hat{\rho}^t \xi_0^* + \sum_{j=0}^{t-1} \hat{\rho}^j \varepsilon_{t-j}^* + \hat{\alpha} \sum_{j=0}^{t-1} \hat{\rho}^j + \hat{\delta} \sum_{j=0}^{t-1} \hat{\rho}^j (t-j) - \hat{\alpha}t,$$

$$u_t^* = (1 - \hat{\rho}L)y_t^* - (\hat{\alpha} + \hat{\delta}t)/\hat{\zeta}(L).$$

Keeping the difference in notation between these two resampling schemes in mind, let $\widetilde{U}_T^* = \widetilde{\Upsilon}_T^{-1} \sum_{t=1}^T \tilde{z}_t^* \tilde{\varepsilon}_t^*$, $U_T^* = \widetilde{\Upsilon}_T^{-1} \sum_{t=1}^T z_t^* \varepsilon_t^*$, $\widetilde{V}_T^* = \widetilde{\Upsilon}_T^{-1} \sum_{t=1}^T \tilde{z}_t^* \tilde{z}_t^{*\prime} \widetilde{\Upsilon}_T^{-1}$, and $V_T^* = \widetilde{\Upsilon}_T^{-1} \sum_{t=1}^T z_t^* z_t^{*\prime} \widetilde{\Upsilon}_T^{-1}$.

We will proceed as follows. First, in Lemma 1 we will provide bounds for the difference between $\widetilde{U}_T^*$ and U_T^* and between $\widetilde{V}_T^*$ and V_T^*. We show that this difference vanishes asymptotically for $\hat{\rho}$ close to unity. This fact allows us to substitute $\widetilde{U}_T^*$ and $\widetilde{V}_T^*$ for U_T^* and V_T^*, respectively, and to proceed as though F were known. Thus, in the subsequent analysis we can treat U_T^* and $\widetilde{U}_T^*$ and V_T^* and $\widetilde{V}_T^*$, respectively, as interchangeable. 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 (6). Finally, we will apply the continuous mapping theorem to show that $V_T^{*-1}U_T^*$ indeed converges to the limit distribution of the bootstrap estimator in (6).

LEMMA 1: *The following three statements are true*:

$$\begin{aligned} &d_1(\widetilde{V}_T^*, V_T^*) \\ &\quad \le \left[5 + 6q\left(\sum_{j=0}^{\infty} \hat{\theta}_j^2\right)^{\frac{1}{2}} + q^2 \sum_{j=0}^{\infty} \hat{\theta}_j^2\right](1 \vee |\hat{\rho}|^{2T})[T^{-1}(\sigma + \hat{\sigma}) + 2|\hat{\alpha}| + T|\hat{\delta}|]d_2(\varepsilon_1, \varepsilon_1^*) \\ &\qquad + o((1 \vee |\hat{\rho}|^{2T})d_2(\varepsilon_1, \varepsilon_1^*)) \quad \omega\text{-}a.s., \end{aligned} \tag{24}$$

$$(25)\qquad d_1(\widetilde{U}_T^*, U_T^*) \le \left[2+q\left(\sum_{k=0}^{\infty}\theta_k^2\right)^{\frac{1}{2}} + \frac{(T^{-\frac{1}{2}}+T|1-\hat{\rho}^2|)(1\vee|\hat{\rho}|^{2(T-1)})}{2|\hat{\rho}|}\right][1+\sigma+\hat{\sigma}+4|\hat{\alpha}|+2T|\hat{\delta}|] \times d_2(\varepsilon_1,\varepsilon_1^*)+o((1+T|1-\hat{\rho}^2|)(1\vee|\hat{\rho}|^{2(T-1)})d_2(\varepsilon_1,\varepsilon_1^*)) \quad \omega\text{-a.s.},$$

and

$$(26)\qquad d_1((\widetilde{U})_T^*, \widetilde{V}_T^*), (U_T^*, V_T^*)) \le RHS\ of\ (24)+RHS\ of\ (25) \quad \omega\text{-a.s.}$$

where $\hat{\theta}_j$ *is the jth moving average coefficient of the moving average representation* $\hat{\zeta}^{-1}(L)$.

For the proof of Lemma 1 see Inoue and Kilian (2000). In the rest of the proof, we treat $L^{\widetilde{U}_T^*,\widetilde{V}_T^*}$, $L^{U_T^*,V_T^*}$ and $L^{U,V}$ as $\mathcal{M}(\Re^{p+2}\times M_{(p+2)\times(p+2)})$-valued random elements, where $M_{(p+2)\times(p+2)}$ is a set of $(p+2)\times(p+2)$ 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:

$$(27)\qquad d_2(\varepsilon_1,\varepsilon_1^*)\to 0 \quad \omega\text{-a.s.}$$

LEMMA 3: *Let U and V be the RHS of* (19) *and* (20), *respectively, with* γ_0 *replaced by the first element of the random vector on the RHS of* (4). *Then*

$$(28)\qquad L^{U_T^*,V_T^*} \xrightarrow{L} L^{U,V} \quad \omega\text{-a.s.}$$

For the proofs of Lemma 2 and Lemma 3 see Inoue and Kilian (2000). Note that U_T^* and V_T^* correspond to the numerator and the denominator of the limiting expression of the bootstrap estimator in (6). Thus, we can derive the limit distribution of the estimator by analyzing the limit distribution of the ratio $V_T^{*-1}U_T^*$.

It follows from Lemma 3 that

$$(29)\qquad \widetilde{\Upsilon}_T^{-1}\sum_{t=1}^{T} z_t^* z_t^{*\prime}\,\widetilde{\Upsilon}_T^{-1} \xrightarrow{L} V \quad \omega\text{-a.s.}$$

$$(30)\qquad \widetilde{\Upsilon}_T^{-1}\sum_{t=1}^{T} z_t^*\varepsilon_t^* \xrightarrow{L} U \quad \omega\text{-a.s.}$$

Since V is nonsingular with probability one by Lemma 1 and Proposition 4 of Jeganathan (1991), it follows by the continuous mapping theorem that

$$(31)\qquad \widetilde{\Upsilon}_T\left(\sum_{t=1}^{T} z_t^* z_t^{*\prime}\right)^{-1}\sum_{t=1}^{T} z_t^*\varepsilon_t^* \xrightarrow{L} V^{-1}U \quad \omega\text{-a.s.}$$

Let

$$x_t^* = [y_{t-1}^* - \hat{\mu}(t-1)\quad \Delta y_{t-1}^* - \hat{\mu}\cdots\Delta y_{t-p+1}^* - \hat{\mu}\quad 1\quad t]',$$

$$N_T = \begin{bmatrix} \hat{\zeta}(1) & \sum_{j=1}^{p-1}\hat{\zeta}_j & \sum_{j=2}^{p-1}\hat{\zeta}_j & \cdots & \hat{\zeta}_{p-1} & \hat{\mu}\sum_{j=1}^{p-1} j\hat{\zeta}_j & 0 \\ 1-\hat{\rho} & \hat{\rho} & 0 & \cdots & 0 & 0 & (1-\hat{\rho})\hat{\mu} \\ 1-\hat{\rho} & \hat{\rho}-1 & \hat{\rho} & \cdots & 0 & -(1-\hat{\rho})\hat{\mu} & (1-\hat{\rho})\hat{\mu} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix},$$

where $\hat{\mu} = \hat{\alpha}/\hat{\zeta}(1)$. Then $y_t^* = \hat{\beta}' x_t^* + \varepsilon_t^*$ and

$$(32) \qquad z_t^* = N_T x_t^* + \begin{bmatrix} 0 \\ \hat{\delta}(t-1)/\hat{\zeta}(L) \\ \vdots \\ \hat{\delta}(t-p+1)/\hat{\zeta}(L) \\ 0_{2\times 1} \end{bmatrix} = N_T x_t^* + \begin{bmatrix} 0 \\ O_{p,(p-1)\times 1}(\frac{t}{T^{3/2}}) \\ 0_{2\times 1} \end{bmatrix} \quad \omega\text{-a.s.}$$

Thus, $\widetilde{\Upsilon}_T(\hat{\beta}_T^* - \hat{\beta})$ can be expressed in terms of z_t^* as

$$(33) \qquad \widetilde{\Upsilon}_T(\hat{\beta}_T^* - \hat{\beta}) = \widetilde{\Upsilon}_T \left(\sum_{t=1}^T x_t^* x_t^{*'} \right)^{-1} \sum_{t=1}^T x_t^* \varepsilon_t^*$$

$$= \widetilde{\Upsilon}_T N_T' \left(\sum_{t=1}^T z_t^* z_t^{*'} \right)^{-1} \sum_{t=1}^T z_t^* \varepsilon_t^* + o_p(1), \quad \omega\text{-a.s.}$$

Therefore, (6) follows from (31) and (33). *Q.E.D.*

PROOF OF COROLLARY 1(a): Since

$$\phi_1 = \rho + \zeta_1,$$

$$\phi_j = \zeta_j - \zeta_{j-1} \quad \text{for} \quad j = 2, \ldots, p-1,$$

$$\phi_p = -\zeta_{p-1},$$

(7) and (8) follow from (4) and Theorem 1, respectively. *Q.E.D.*

PROOF OF COROLLARY 1(b): The additional result required for proving Corollary 1(b) is the consistency of the bootstrap estimator of the variance of $\rho, \zeta_1, \ldots, \zeta_{p-1}$. By arguments similar to (32), it suffices to show that the upper-left $p \times p$ submatrix of $((1/T)\sum_{t=1}^T z_t^* z_t^{*'})^{-1}$ converges in probability to

$$(34) \qquad \begin{bmatrix} 0 & 0_{1\times q} \\ 0_{q\times 1} & \Gamma^{-1} \end{bmatrix}$$

ω-a.s. and that $\hat{\sigma}^{*2}$ converges in probability to $\sigma^2 \omega$-a.s. By (30), one can show that the upper-left $p \times p$ submatrix of $((1/T)\sum_{t=1}^T z_t^* z_t^{*'})^{-1}$ is

$$(35) \qquad \begin{bmatrix} O_p(T^{-1}) & O_{p,1\times p}(T^{-1/2}) \\ O_{p,p\times 1}(T^{-1/2}) & \Gamma^{-1} + o_{p,q\times q} \end{bmatrix} \quad \omega\text{-a.s.}$$

Thus, the upper-left $p \times p$ submatrix of $((1/T)\sum_{t=1}^T z_t^* z_t^{*'})^{-1}$ converges in probability to (34) ω-a.s. It remains to be shown that

$$(36) \qquad \hat{\sigma}^{*2} = \sigma^2 + o_p(1) \quad \omega\text{-a.s.}$$

By definition,

$$(37) \qquad \hat{\sigma}^{*2} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^{*2}$$

$$= \frac{1}{T} \sum_{t=1}^T \varepsilon_t^{*2} - \frac{1}{T} \sum_{t=1}^T \varepsilon_t^* \bar{x}_t^{*'} \left(\sum_{t=1}^T \bar{x}_t^* \bar{x}_t^{*'} \right)^{-1} \sum_{t=1}^T \bar{x}_t^* \varepsilon_t^*$$

$$= \frac{1}{T} \sum_{t=1}^T \varepsilon_t^{*2} - \frac{1}{T} \sum_{t=1}^T \varepsilon_t^* z_t^{*'} \left(\sum_{t=1}^T z_t^* z_t^{*'} \right)^{-1} \sum_{t=1}^T z_t^* \varepsilon_t^*.$$

By equation (55) of Inoue and Kilian (2000),

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^{*2} = E^*(\varepsilon_t^{*2}) + o(1) = \hat{\sigma}^2 + o(1) \quad \omega\text{-a.s.} \tag{38}$$

From equations (73) and (74) of Inoue and Kilian (2000), it follows that

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^{*} z_t^{*'}\left(\sum_{t=1}^{T} z_t^{*} z_t^{*'}\right)^{-1}\sum_{t=1}^{T} z_t^{*}\varepsilon_t^{*} = o_p(1) \quad \omega\text{-a.s.} \tag{39}$$

Therefore, (37), (38), and (39) imply (36). By the Slutsky Theorem, this result implies the asymptotic validity of bootstrapping the studentized estimators of linear combinations of level slope parameters, provided the limit distribution is nondegenerate. *Q.E.D.*

PROOF OF COROLLARY 2: For the proof of Corollary 2, see Inoue and Kilian (2000). *Q.E.D.*

REFERENCES

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.

——— (1991b): "Bootstrap Test of Significance and Sequential Bootstrap Estimation for Unstable First Order Autoregressive Processes," *Communications in Statistical Theory and Methods*, 20, 1015–1026.

BERAN, R., AND G. R. DUCHARME (1991): *Asymptotic Theory for Bootstrap Methods in Statistics*. Montreal, Canada: Les Publications CRM, Centre de Recherches Mathematiques, University of Montreal.

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

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.

EFRON, B., AND R. J. TIBSHIRANI (1993): *An Introduction to the Bootstrap.* New York: Chapman & Hall.

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," *Review of Economics and Statistics*, 81, 594–607.

HEIMANN, G., AND J.-P. KREISS (1996): "Bootstrapping General First-Order Autoregression," *Statistics and Probability Letters*, 30, 87–98.

INOUE, A., AND L. KILIAN (2000): "Bootstrapping Autoregressive Processes with Possible Unit Roots," Manuscript; http://www.econ.lsa.umich.edu/~lkilian.

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): "Small-Sample Confidence Intervals for Impulse Response Functions,"*Review of Economics and Statistics*, 80, 218–230.

——— (1999): "Finite-Sample Properties of Percentile and Percentile-*t* Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics*, 81, 652–660.

SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): "Inference in Linear Time Series with Some Unit Roots," *Econometrica*, 58, 113–144.

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