Econometrica, Vol. 69, No. 5 (September, 2001), 1283–1314

SUBSAMPLING INTERVALS IN AUTOREGRESSIVE MODELS WITH LINEAR TIME TREND

BY JOSEPH P. ROMANO AND MICHAEL WOLF[1]

A new method is proposed for constructing confidence intervals in autoregressive models with linear time trend. Interest focuses on the sum of the autoregressive coefficients because this parameter provides a useful scalar measure of the long-run persistence properties of an economic time series. Since the type of the limiting distribution of the corresponding OLS estimator, as well as the rate of its convergence, depend in a discontinuous fashion upon whether the true parameter is less than one or equal to one (that is, trend-stationary case or unit root case), the construction of confidence intervals is notoriously difficult. The crux of our method is to recompute the OLS estimator on smaller blocks of the observed data, according to the general subsampling idea of Politis and Romano (1994a), although some extensions of the standard theory are needed. The method is more general than previous approaches in that it works for arbitrary parameter values, but also because it allows the innovations to be a martingale difference sequence rather than i.i.d. Some simulation studies examine the finite sample performance.

KEYWORDS: Autoregressive time series, local-to-unity asymptotics, subsampling, trend-stationarity, unit roots.

1. INTRODUCTION

MUCH WORK IN THE RECENT LITERATURE has been devoted to the question of whether a macroeconomic time series is trend-stationary or whether it possesses a unit root. When the time series is modeled by an AR(p) sequence with linear time trend, the answer depends on whether the sum of the AR(p) coefficients, a parameter usually denoted by α, is less than one or equal to one. This parameter can be consistently estimated by applying ordinary least squares (OLS) to the usual Dickey-Fuller form regression model. Unfortunately, the construction of confidence intervals is nontrivial, since the type of the asymptotic distribution of the OLS estimator, as well as its rate of convergence, are different in the trend-stationary case as compared to the unit root case. When $\alpha < 1$, the limiting distribution is normal and the rate of convergence is the square root of the sample size. On the other hand, when $\alpha = 1$, the limiting distribution is nonstandard and the rate of convergence equals the sample size. This difficulty explains the emphasis in the unit root literature placed on hypothesis testing. However, confidence intervals provide much more information than knowing whether the null hypothesis of a unit root can be rejected or not, namely they serve as a measure of sampling uncertainty and describe the

[1] We thank Graham Elliot, Guido Imbens, and Halbert White for helpful discussions. The final version of this paper has substantially benefited from comments of three anonymous referees and the co-editor. Research of the second author was partly funded by the Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB98-0025.

range of models that are consistent with the observed data. This point was made in Stock (1991) and Andrews and Chen (1994), among others.

In this paper, we propose a novel approach for constructing confidence intervals for the parameter α, based on the subsampling method of Politis and Romano (1994a). The crux of the new approach is to recompute the OLS estimator on smaller blocks, or "subsamples," of the observed data sequence. Then, the empirical distribution of these subsample estimates, after an appropriate normalization, is used to approximate the sampling distribution of the estimator based on the entire data. Unlike the conventional bootstrap, the subsampling method can handle the discontinuity of the limiting distribution of the OLS estimator (as a function of α), since the subsamples are all generated by the true model rather than an approximating bootstrap distribution. While we will focus on the parameter α, the proposed method can equally well be applied to construct confidence intervals for alternative parameters of interest, such as the largest root of the AR(p) model, a particular AR coefficient, or the coefficient on the time trend.

In Section 2, the model and the parameter of principal interest are presented. Also, some previous methods for confidence interval construction are briefly described. The basic subsampling methodology and some necessary extensions are discussed in Section 3. The general approaches are applied to the problem at hand in Section 4, while Section 5 deals with issues concerning the practical implementation. Two simulation studies are presented in Section 6. Some conclusions are stated in Section 7. The proofs appear in an Appendix.

2. DEFINITIONS AND BACKGROUND

2.1. *Definition of the Model*

The model under consideration is an AR(p) model with intercept and linear time trend. The exposition of the model and the notation closely follow Andrews and Chen (1994) with the exception that we do not require the innovations to be i.i.d. and normal. The model can be written in an unobserved-components form and in a regression form. In the former, it is given by

$$(1) \qquad Y_t = \mu^* + \beta^* t + Y_t^* \quad \text{for} \quad t = 1, \ldots, n,$$
$$Y_t^* = \alpha Y_{t-1}^* + \psi_1 \Delta Y_{t-1}^* + \cdots + \psi_{p-1} \Delta Y_{t-1}^* + \epsilon_t,$$

where $\{\epsilon_t,\ t = p+1, \ldots, n\}$ is a strictly stationary, martingale difference innovation sequence and $\{Y_t,\ t = 1, \ldots, n\}$ is the observed series. The variable ΔY_t^* denotes $Y_t^* - Y_{t-1}^*$. The parameter α satisfies $\alpha \in (-1, 1]$; when $\alpha = 1$, the model is nonstationary. The parameters $(\psi_1, \ldots, \psi_{p-1})$ are such that the AR model for Y_t^* is stationary when $\alpha \in (-1, 1)$ and the AR model for ΔY_t^* is stationary when $\alpha = 1$. The starting values of Y_t^*—that is, $(Y_1^*, \ldots, Y_p^*)$—are taken such that $\{Y_t^*\}$ is stationary when $\alpha \in (-1, 1)$ and $\{\Delta Y_t^*\}$ is stationary

when $\alpha = 1$. The level of the ΔY_t^* series is arbitrary when $\alpha = 1$ (that is, when $\alpha = 1$, the initial random variable Y_1^* can be fixed or can have any distribution provided the subsequent Y_t^* values are such that ΔY_t^* is stationary).

The regression form of Model (1) is given by

$$(2) \qquad Y_t = \mu + \beta t + \alpha Y_{t-1} + \psi_1 \Delta Y_{t-1} \cdots + \psi_{p-1} \Delta Y_{t=p+1} + \epsilon_t \quad \text{for} \quad t = 1, \ldots, n,$$

$$\mu = \mu^*(1 - \alpha) + (\alpha - \psi_1 - \cdots - \psi_{p-1})\beta^* \quad \text{and} \quad \beta = \beta^*(1 - \alpha),$$

where $(Y_1, \ldots, Y_p)$ and $\{\epsilon_t, t = p + 1, \ldots, T\}$ are defined as in (1). Model (2) is the well-known augmented Dickey-Fuller regression form of the AR(p) model. The corresponding standard AR(p) regression form is given by

$$(3) \qquad Y_t = \mu + \beta t + \gamma_1 Y_{t-1} + \cdots + \gamma_p Y_{t-p} + \epsilon_t.$$

As is easy to see, the parameter α in the augmented Dickey-Fuller form equals the sum of the AR coefficients in the standard form, that is, $\alpha = \gamma_1 + \cdots + \gamma_p$. Moreover, it follows that $\psi_j = -(\gamma_{j+1} + \cdots + \gamma_p)$ for $j = 1, \ldots, p - 1$.

It should be pointed out that the time trend parameter β is necessarily equal to 0 when $\alpha = 1$ in both models (2) and (3). This desirable feature ensures that $E(Y_t)$ is a linear function of t for all $\alpha \in (-1, 1]$. If $\beta \neq 0$ were allowed when $\alpha = 1$, $E(Y_t)$ would be a quadratic function of t when $\alpha = 1$, so this discontinuity is naturally avoided.

2.2. *The Parameter of Interest and its Inference Problems*

The remainder of this paper will mainly focus on constructing confidence intervals for the parameter α. The motivation is that it provides a useful scalar measure for the long-run persistence properties of the time series $\{Y_t\}$. Indeed, in AR(p) models, $1/(1 - \alpha)$ equals the sum of the impulse response functions over all time horizons, that is, the cumulative impulse response; see Andrews and Chen (1994). An alternative scalar measure that has been considered in the literature is the largest root of the Ar(p) model, usually denoted by ρ. For example, Stock (1991) derived asymptotic confidence intervals for ρ based on a local-to-unity model and DeJong and Whiteman (1991a, b) discussed Bayes estimators of ρ. But, as was demonstrated in Andrews and Chen (1994), the persistence of two time series with the same value of ρ can be very different depending on the values of the other roots. Therefore, we have decided to focus on the parameter α instead. Note, however, that the methodology developed in this paper to construct confidence intervals for α can equally well be employed to construct confidence intervals for ρ.

As usual when inference for an unknown scalar parameter is desired, there exist two main avenues, namely hypothesis tests and confidence intervals. In accordance with many other authors, we feel that a confidence interval is much more informative than a test, since it not only states whether a specific parameter value is rejected or not by the observed data, but because it also

provides the range of all parameter values consistent with the data. In particular, this allows judgment of the degree of uncertainty about point estimates of the unknown parameter. So why is it that the main part of the unit root literature has been concerned with hypothesis tests for α, with the null hypothesis typically given by $\alpha = 1$? The reason for this preoccupation with the 'wrong' method is that hypothesis tests for α are by an order of magnitude easier to construct than confidence intervals. While α can be consistently estimated by applying OLS to the Dickey-Fuller form regression model (2), the form of the limiting distribution of the OLS estimator, as well as its rate of convergence depend in a discontinuous way on whether or not α equals 1; see the proof of Theorem 4.1 for details. Exactly this fact makes the construction of confidence intervals difficult. Clearly, the standard asymptotic approach—using the quantiles of the (estimated) limiting distribution—is rendered useless, since one has to know whether α is equal to 1 or not in order to know the quantiles of which distribution one should use. Moreover, the discontinuity of the form of the limiting distribution, as a function of α, causes the standard, residual-based bootstrap confidence intervals to fail; see Basawa et al. (1991). On the other hand, this dilemma does not affect hypothesis tests, since they only require specification of the limiting distribution of the test statistic under the null hypothesis. Despite the inherent difficulties in constructing confidence intervals for the parameter α, some notable progress has been made recently.

Stock (1991), focusing on the largest root ρ of the AR(p) model rather than on the sum of the AR(p) parameters, made use of local-to-unity asymptotics. To be more specific, he assumed that ρ shrinks towards one as the sample size tends to infinity in the linear fashion $\rho = 1 + c/n$, for some constant $c \leq 0$; note that the theory also works when $c > 0$. This model allows testing of the null hypothesis $c = c_0$ for any value c_0 and thereby finding a confidence interval for c as the collection of c_0 values that are not rejected by the test. Using the relation $\rho = 1 + c/n$, a confidence interval for ρ immediately ensues. The downside of this approach may be considered its 'breakdown' problem. The confidence intervals work well when ρ is 'close' to one, where 'close' depends on the sample size. Judging from the simulation studies in Stock (1991), the actual coverage is (nearly) equal to the nominal one when $\rho = 1$, that is, when $c = 0$ but deteriorates as ρ moves away from 1, that is, as c and/or n decrease.

Andrews and Chen (1994) based confidence intervals on approximately median-unbiased estimation in AR(p) models. This is an extension of previous work of Andrews (1993), where exactly median-unbiased estimation in AR(1) models was developed. The idea is to compute (or to simulate with arbitrary precision) the sampling distribution of the OLS estimator $\hat{\alpha}_n$ using model (2) but with i.i.d. innovations from a $N(0, \sigma^2)$ distribution. In the AR(1) case, this distribution can be shown to depend on α only, but not on μ, β, and σ^2; see Andrews (1993). In the general AR(p) case, the distribution also depends on $(\psi_1, \ldots, \psi_{p-1})$, so it can only be approximated. Given that one can (approximately) compute/simulate the sampling distribution of $\hat{\alpha}_n$ for any value of α_0, a confidence interval for α is obtained as the collection of all α_0 values whose

sampling distribution is "consistent" with the observed value of $\hat{\alpha}_n$; see Andrews and Chen (1994) for details. The obvious criticism of this method is that one has to specify the distribution of the innovations (such as normal) in order to calculate/simulate the sampling distribution of $\hat{\alpha}_n$. However, the method seems fairly robust against misspecification of this distribution, as appears from some simulations in Andrews and Chen (1994), as long as the innovations remain i.i.d. It will be clear from the proof of Theorem 4.1 that the method in general does not work when the innovations are dependent; specifically, see Remark 4.3.

As mentioned before, even when the innovations are assumed i.i.d., the conventional, residual-based bootstrap confidence intervals fail when $\alpha = 1$; see Basawa et al. (1991). On the other hand, it was shown by Hansen (1999) that one can construct 'indirect' bootstrap confidence intervals that are guaranteed to work for any $\alpha \in (-1, 1]$. The trick is to invert bootstrap tests, that is, to obtain a confidence interval for α as the collection of all α_0 values that are not rejected by a bootstrap test of the null hypothesis $\alpha = \alpha_0$. Hansen coined his method the "grid bootstrap" but it should be pointed out that the idea of inverting bootstrap tests to construct confidence intervals is time honored; for example, see DiCiccio and Romano (1988). A shortcoming of this approach is that it, too, is restricted to i.i.d. innovations.

The aim of this paper is to provide a new way for constructing confidence intervals for α that works for any $\alpha \in (-1, 1]$ and allows for stationary, dependent rather than i.i.d. innovations, though even the assumption of stationarity could be relaxed; see Remark 3.2. The new approach is based on the subsampling method of Politis and Romano (1994a). To make the paper self-contained, the general method pertaining to univariate parameters will be briefly described; broader methods, pertaining to multivariate or function-valued parameters can be found in Politis, Romano, and Wolf (1999). Then, some extensions of the standard theory that are necessary for the case $\alpha = 1$ will be presented.

3. THE SUBSAMPLING METHOD

3.1. *The Basic Method*

Subsampling is a general tool that allows one to construct asymptotically valid confidence intervals for unknown parameters under very weak assumptions.

Suppose $\{X_1, X_2, \ldots\}$ is a sequence of vector-valued random variables defined on a common probability space. Denote the joint probability law governing the infinite sequence by P. The goal is to construct a confidence interval for some real-valued parameter $\theta = \theta(P)$, on the basis of observing $\{X_1, \ldots, X_n\}$. We assume the existence of a sensible estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$.

For time series data, the gist of the subsampling method is to recompute the statistic of interest on smaller blocks of the observed sequence $\{X_1, \ldots, X_n\}$. Define $\hat{\theta}_{b,t} = \hat{\theta}_b(X_t, \ldots, X_{t+b-1})$, the estimator of θ based on the subsample $\{X_t, \ldots, X_{t+b-1}\}$. In this notation, b is the block size and t is the starting index of the block; note that $\hat{\theta}_{n,1} = \hat{\theta}_n$. Let $J_b(P)$ be the sampling distribution of

$\tau_b(\hat{\theta}_{b,t} - \theta)$, assuming that this distribution is independent of t. Here, τ_b is an appropriate normalizing constant. Also, define the corresponding cumulative distribution function:

$$J_b(x, P) = \text{Prob}_P\left\{\tau_b\left(\hat{\theta}_{b,t} - \theta\right) \leq x\right\}.$$

Note that with this notation $J_n(P)$ is the sampling distribution of $\tau_n(\hat{\theta}_n - \theta)$, that is, the sampling distribution of the (properly normalized) estimator based on the entire sample. A major assumption that is needed to construct asymptotically valid confidence intervals for θ is the following.

ASSUMPTION 3.1: *There exists a nondegenerate limiting law $J(P)$ such that $J_n(P)$ converges weakly to $J(P)$.*

This assumption states that the estimator, properly normalized, has a limiting distribution. It is hard to conceive of any asymptotic theory free of such a requirement. Also, it follows that the proper normalizing constant τ_n is the one ensuring a limiting distribution. In regular cases, the limiting distribution is normal and $\tau_n = n^{1/2}$.

The subsampling approximation to $J_n(x, P)$ is defined by

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1\left\{\tau_b\left(\hat{\theta}_{b,t} - \hat{\theta}_n\right) \leq x\right\}. \tag{4}$$

The motivation behind the method is the following. For any t, $\{X_t, \ldots, X_{t+b-1}\}$ is a 'true' subsample of size b. Hence, the *exact* distribution of $\tau_b(\hat{\theta}_{b,t} - \theta)$ is $J_b(P)$. If both b and n are large, then the empirical distribution of the $n-b+1$ values of $\tau_b(\hat{\theta}_{b,t} - \theta)$ should serve as a good approximation to $J_n(P)$. Replacing θ by $\hat{\theta}_n$ is permissible because $\tau_b(\hat{\theta}_n - \theta)$ is of order τ_b/τ_n in probability and we will assume that $\tau_b/\tau_n \to 0$.

For $J_n(x, P)$ to be approximated consistently by $L_{n,b}(x)$, both should have the same limit, namely $J(x, P)$. To ensure that $L_{n,b}(x)$ converges to $J(x, P)$ in probability, it is necessary that the information in the $n-b+1$ subsample statistics $\tau_b(\hat{\theta}_{b,t} - \hat{\theta}_n)$ tend to infinity with the sample size n. In previous theory (Politis and Romano (1994a); Politis, Romano, and Wolf (1997)), this followed from a weak dependence condition on the underlying sequence $\{Y_t\}$, namely an α-mixing condition (Rosenblatt (1956)).

DEFINITION 3.1: Given a stationary random sequence $\{X_t\}$, let $\mathcal{F}_t^s$ be the σ-algebra generated by the segment $\{X_t, X_{t+1}, \ldots, X_s\}$ and define the corresponding α-mixing sequence by

$$\alpha_X(h) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|,$$

where A and B vary over the σ-fields $\mathcal{F}_{-\infty}^t$ and $\mathcal{F}_{t+h}^{\infty}$, respectively. The sequence $\{X_t\}$ is called α-mixing or strong mixing if $\alpha_X(h) \to 0$ as $h \to \infty$.

For our applications, it will be convenient to have a more general theory that imposes a mixing condition on the subsample statistics only rather than on the underlying sequence. To this end, let $Z_{n,b,t} = \tau_b(\hat{\theta}_{b,t} - \theta)$ and denote by $\alpha_{a,b}(\cdot)$ the mixing coefficients corresponding to the sequence $\{Z_{n,b,t}, t = 1, \ldots, n - b + 1\}$.

The following theorem shows how subsampling can be used to construct asymptotically valid confidence intervals for θ.

THEOREM 3.1: *Assume Assumption* 3.1 *and that* $\tau_b/\tau_n \to 0$, $b/n \to 0$ *and* $b \to \infty$ *as* $n \to \infty$. *Also assume that* $n^{-1}\sum_{h=1}^{n} \alpha_{n,b}(h) \to 0$ *as* $n \to \infty$.

(i) *If* x *is a continuity point of* $J(\cdot, P)$, *then* $L_{n,b}(x) \to J(x, P)$ *in probability.*

(ii) *If* $J(\cdot, P)$ *is continuous, then* $\sup_x |L_{n,b}(x) - J(x, P)| \to 0$ *in probability.*

(iii) *For* $\lambda \in (0, 1)$, *let* $c_{n,b}(1 - \lambda) = \inf\{x : L_{n,b}(x) \geq 1 - \lambda\}$. *In other words,* $c_{n,b}(1 - \lambda)$ *serves as an* $(1 - \lambda)$ *quantile of the subsampling distribution* $L_{n,b}(\cdot)$. *Correspondingly, define* $c(1 - \lambda, P) = \inf\{x : J(x, P) \geq 1 - \lambda\}$. *If* $J(\cdot, P)$ *is continuous at* $c(1 - \lambda, P)$, *then*

$$\text{Prob}_P\left\{\tau_n\left(\hat{\theta}_n - \theta\right) \leq c_{n,b}(1 - \lambda)\right\} \to 1 - \lambda \quad \text{as} \quad n \to \infty.$$

Thus, the asymptotic coverage probability under P *of the interval*

$$I_1 = \left[\hat{\theta}_n - \tau_n^{-1} c_{n,b}(1 - \lambda), \infty\right)$$

is the nominal level $1 - \lambda$.

REMARK 3.1: The sufficient conditions on the block size b are very weak. In most applications, $\tau_n = n^{\kappa}$, for some constant $\kappa > 0$ and the conditions reduce to $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. As shown in Politis and Romano (1994a), the latter two conditions are in general not only sufficient but also necessary.

REMARK 3.2: The general theory presented here assumes that the subsample statistics are stationary, i.e., that the sequence $\{\hat{\theta}_{b,t}, t = 1, \ldots, n - b + 1\}$ is stationary. Note that this assumption could be relaxed to accommodate local heteroskedasticity and/or changing distributions of the subsample statistics along the lines of Politis, Romano, and Wolf (1997).

The interval I_1 in (iii) corresponds to a one-sided hybrid percentile interval in the bootstrap literature (e.g., Hall (1992)). A two-sided equal-tailed confidence interval can be obtained by forming the intersection of two one-sided intervals. The two-sided analogue of I_1 is

$$I_2 = \left[\hat{\theta}_n - \tau_n^{-1} c_{n,b}(1 - \alpha/2), \hat{\theta}_n - \tau_n^{-1} c_{n,b}(\alpha/2)\right].$$

I_2 is called equal-tailed because it has approximately equal probability in each tail:

$$\begin{aligned}&\text{Prob}_P\left\{\theta < \hat{\theta}_n - \tau_n^{-1} c_{n,b}(1 - \alpha/2)\right\} \\ &\quad \simeq \text{Prob}_P\left\{\theta > \hat{\theta}_n - \tau_n^{-1} c_{n,b}(\alpha/2)\right\} \simeq \alpha/2,\end{aligned}$$

where $\simeq$ denotes equality up to an additive $o(1)$ term. As an alternative approach, two-sided symmetric confidence intervals can be constructed. A two-sided symmetric confidence interval is given by $[\hat{\theta}_n - \hat{c}, \hat{\theta}_n + \hat{c}]$, where $\hat{c}$ is chosen so that $\text{Prob}_P\{|\hat{\theta}_n - \theta| > \hat{c}\} \simeq \alpha$. Hall (1988) showed that symmetric bootstrap confidence intervals enjoy enhanced coverage and, even in asymmetric circumstances, can be shorter than equal-tailed confidence intervals. An analogue for symmetric subsample confidence intervals, for the application of the sample mean, was provided by Politis, Romano, and Wolf (1999, Chapter 10). To construct two-sided symmetric subsampling intervals in practice, one estimates the two-sided distribution function

$$J_{n,|\cdot|}(x, P) = \text{Prob}_P\left\{\tau_n|\hat{\theta}_n - \theta| \le x\right\}.$$

The subsampling approximation to $J_{n,|\cdot|}(x, P)$ is defined by

$$L_{n,b,|\cdot|}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} 1\left\{\tau_b|\hat{\theta}_{b,t} - \hat{\theta}_n| \le x\right\}. \tag{5}$$

The asymptotic validity of two-sided symmetric subsampling intervals immediately follows from Theorem 3.1 and the continuous mapping theorem.

COROLLARY 3.1: *Make the same assumptions as in Theorem* 3.1. *Denote by* $J_{|\cdot|}(P)$ *the distribution of* $|Q|$, *where* Q *is a random variable with distribution* $J(P)$.

(i) *If* x *is a continuity point of* $J_{|\cdot|}(\cdot, P)$, *then* $L_{n,b,|\cdot|}(x) \to J_{|\cdot|}(x, P)$ *in probability*.

(ii) *If* $J_{|\cdot|}(\cdot, P)$ *is continuous, then* $\sup_x|L_{n,b,|\cdot|}(x) - J_{|\cdot|}(x, P)| \to 0$ *in probability*.

(iii) *For* $\lambda \in (0,1)$, *let* $c_{n,b,|\cdot|}(1-\lambda) = \inf\{x : L_{n,b,|\cdot|}(x) \ge 1-\lambda\}$. *Correspondingly, define* $c_{|\cdot|}(1-\lambda, P) = \inf\{x : J_{|\cdot|}(x, P) \ge 1-\lambda\}$. *If* $J_{|\cdot|}(\cdot, P)$ *is continuous at* $c_{|\cdot|}(1-\lambda, P)$, *then*

$$\text{Prob}_P\left\{\tau_n|\hat{\theta}_n - \theta| \le c_{n,b,|\cdot|}(1-\lambda)\right\} \to 1-\lambda \quad \textit{as} \quad n \to \infty.$$

Thus, the asymptotic coverage probability under P *of the interval* $I_{SYM} = [\hat{\theta}_n - \tau_n^{-1}c_{n,b,|\cdot|}(1-\lambda), \hat{\theta}_n + \tau_n^{-1}c_{n,b,|\cdot|}(1-\lambda)]$ *is the nominal level* $1-\lambda$.

The application of Theorem 3.1 or Corollary 3.1 requires knowledge of the rate of convergence τ_n. In standard cases, this is simply $n^{1/2}$. In nonstandard cases, it may be another power of n; for example, see Subsection 4.3. As long as the rate is known, nonstandard cases do not pose a problem.

On the other hand, for the parameter α it is well-known that the rate of convergence of the OLS estimator $\hat{\alpha}_n$ is given by $n^{1/2}$ when $\alpha < 1$ and by n when $\alpha = 1$, respectively. Hence, the application of the basic subsampling method would require the knowledge of whether the time series is trend-stationary or has a unit root! Fortunately, there is a way around this dilemma by considering a studentized statistic, namely the usual t-statistic for $\hat{\alpha}_n$. Indeed,

this statistic has a proper limiting distribution no matter what the value of α. The next subsection will provide the necessary theory to apply the subsampling idea in a studentized setting.

3.2. *Subsampling Studentized Statistics*

The focus is now on a studentized statistic $\tau_n^\bullet(\hat{\theta}_n - \theta)/\hat{\sigma}_n$, where $\hat{\sigma}_n = \hat{\sigma}_n(Y_1, \ldots, Y_n)$ is some positive estimate of scale. Note that the appropriate normalizing constant $\tau_n^\bullet$ may be different from its analogue τ_n in the nonstudentized case. Define $J_b^\bullet(P)$ to be the sampling distribution of $\tau_b^\bullet(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t}$ based on the subsample $Y_t, \ldots, Y_{t+b-1}$, assuming that this distribution is independent of t. Also, define the corresponding cumulative distribution function

$$J_b^\bullet(x, P) = \text{Prob}_P\left\{\tau_b^\bullet\left(\hat{\theta}_{b,t} - \theta\right)/\hat{\sigma}_{b,t} \le x\right\}.$$

The subsampling method is modified to the studentized case in the obvious way. Analogous to (4), define

$$L_{n,b}^\bullet(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} 1\left\{\tau_b^\bullet\left(\hat{\theta}_{b,t} - \hat{\theta}_n\right)/\hat{\sigma}_{b,t} \le x\right\}. \tag{6}$$

$L_{n,b}^\bullet(x)$ then represents the subsampling approximation to $J_n^\bullet(x, P)$.

The essential assumption needed to construct asymptotically valid confidence regions for θ now becomes more involved than for the nonstudentized case.

ASSUMPTION 3.2: *$J_n^\bullet(P)$ converges weakly to a nondegenerate limit law $J^\bullet(P)$. In addition, there exist positive sequences $\{a_n\}$ and $\{d_n\}$ such that $\tau_n^\bullet = a_n/d_n$, $a_n(\hat{\theta}_n - \theta)$ converges weakly to a limit law $V(P)$, and $d_n\hat{\sigma}_n$ converges weakly to a limit law $W(P)$ without positive mass at zero.*

THEOREM 3.2: *Assume Assumption* 3.2, $a_b/a_n \to 0$, $\tau_b^\bullet/\tau_n^\bullet \to 0$, $b/n \to 0$, *and* $b \to \infty$ *as* $n \to \infty$. *Also assume that $\{X_t\}$ is near epoch dependent of size $-q$, for some $q > 2$, on a basis process $\{V_t\}$ whose α-mixing coefficients satisfy* $\lim_{n\to\infty} n^{-1}\sum_{h=1}^{n} \alpha_V(h)^{1-2/r} < \infty$ *for some* $r > 0$.

(i) *If x is a continuity point of $J^\bullet(\cdot, P)$, then $L_{n,b}^\bullet(x) \to J^\bullet(x, P)$ in probability.*

(ii) *If $J^\bullet(\cdot, P)$ is continuous, then* $\sup_x |L_{n,b}^\bullet - J^\bullet(x, P)| \to 0$ *in probability.*

(iii) *For $\lambda \in (0,1)$, let* $c_{n,b}^\bullet(1-\lambda) = \inf\{x : L_{n,b}^\bullet(x) \ge 1-\lambda\}$. *Correspondingly, define* $c^\bullet(1-\lambda, P) = \inf\{x : J^\bullet(x, P) \ge 1-\lambda\}$. *If $J^\bullet(\cdot, P)$ is continuous at $c^\bullet(1-\lambda, P)$ then*

$$\text{Prob}_P\left\{\tau_n^\bullet\left(\hat{\theta}_n - \theta\right)/\hat{\sigma}_n \le c_{n,b}^\bullet(1-\lambda)\right\} \to 1-\lambda \quad \text{as} \quad n \to \infty.$$

Thus, the asymptotic coverage probability under P of the interval $I_1^\bullet = [\hat{\theta}_n - \hat{\sigma}_n(\tau_n^\bullet)^{-1}c_{n,b}^\bullet(1-\lambda), \infty)$ *is the nominal level* $1-\lambda$.

The issue of symmetric confidence intervals applies as well to studentized statistics. Let $J^{\bullet}_{n,|\cdot|}(P)$ be the sampling distribution of $\tau_n^{\bullet}|\hat{\theta}_n - \theta|/\hat{\sigma}_n$. Define

$$L^{\bullet}_{n,b,|\cdot|}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} 1\left\{\tau_b^{\bullet}|\hat{\theta}_{b,t} - \hat{\theta}_n|/\hat{\sigma}_{b,t} \le x\right\}. \tag{7}$$

$L^{\bullet}_{n,b,|\cdot|}(x)$ then represents the subsampling approximation to $J^{\bullet}_{n,|\cdot|}(x)$.

Theorem 3.2 and the continuous mapping theorem immediately imply the following corollary.

COROLLARY 3.2: *Make the same assumptions as in Theorem* 3.2. *Denote by* $J^{\bullet}_{|\cdot|}(P)$ *the distribution of* $|U|$, *where* U *is a random variable with distribution* $J^{\bullet}(P)$.

(i) *If* x *is a continuity point of* $J^{\bullet}_{|\cdot|}(\cdot, P)$, *then* $L^{\bullet}_{n,b,|\cdot|}(x) \to J^{\bullet}_{|\cdot|}(x, P)$ *in probability*.

(ii) *If* $J^{\bullet}_{|\cdot|}(\cdot, P)$ *is continuous, then* $\sup_x |L^{\bullet}_{n,b,|\cdot|} - J^{\bullet}_{|\cdot|}(x, P)| \to 0$ *in probability*.

(iii) *For* $\lambda \in (0,1)$, *let* $c^{\bullet}_{n,b,|\cdot|}(1-\lambda) = \inf\{x : L^{\bullet}_{n,b,|\cdot|}(x) \ge 1-\lambda\}$. *Correspondingly, define* $c^{\bullet}_{|\cdot|}(1-\lambda, P) = \inf\{x : J^{\bullet}_{|\cdot|}(x, P) \ge 1-\lambda\}$. *If* $J^{\bullet}_{|\cdot|}(\cdot, P)$ *is continuous at* $c^{\bullet}_{|\cdot|}(1-\lambda, P)$, *then*

$$\mathrm{Prob}_P\left\{\tau_n^{\bullet}|\hat{\theta}_n - \theta|/\hat{\sigma}_n \le c^{\bullet}_{n,b,|\cdot|}(1-\lambda)\right\} \to 1-\lambda \quad \text{as} \quad n \to \infty.$$

Thus, the asymptotic coverage probability under P *of the interval*

$$I^{\bullet}_{SYM} = \left[\hat{\theta}_n - \hat{\sigma}_n(\tau_n^{\bullet})^{-1} c^{\bullet}_{n,b,|\cdot|}(1-\lambda), \hat{\theta}_n + \hat{\sigma}_n(\tau_n^{\bullet})^{-1} c^{\bullet}_{n,b,|\cdot|}(1-\lambda)\right]$$

is the nominal level $1-\lambda$.

4. SUBSAMPLING INFERENCE

4.1. *Confidence Intervals for* α *in the Full Model*

We will now demonstrate that the subsampling approach of Subsection 3.2 can be applied to construct asymptotically valid confidence intervals for the parameter α. Hence, α will play the role of the general parameter θ of the previous section.

The estimator $\hat{\alpha}_n$ is the OLS estimator for α based on the Dickey-Fuller regression form (2); note that it would be numerically equivalent to compute $\hat{\alpha}_n$ as $\sum_{i=1}^{p} \hat{\gamma}_{i,n}$ where the $\hat{\gamma}_{i,n}$ are the OLS estimators of model (3). Consequently, $\hat{\alpha}_{b,t}$ is the OLS estimator for α based on the block of data $\{Y_t, \ldots, Y_{t+b-1}\}$. Denote the corresponding OLS standard errors by $SE_{OLS}(\hat{\alpha}_n)$ and $SE_{OLS}(\hat{\alpha}_{b,t})$. For reasons to become apparent shortly, define $\hat{\sigma}_n = n^{1/2} SE_{OLS}(\hat{\alpha}_n)$ and $\hat{\sigma}_{b,t} = b^{1/2} SE_{OLS}(\hat{\alpha}_{b,t})$. To apply the methodology of Subsection 3.2, it is left to specify the appropriate normalizing constant $\tau_n^{\bullet}$. With the definition of $\hat{\sigma}_n$ above, this turns out to be $n^{1/2}$ no matter what the value of α; see the proof of the following theorem.

Denote the mixing coefficients corresponding to sequence $\{Y_t^*\}$, which is stationary when $\alpha < 1$, by α_{Y^*} and the mixing coefficients corresponding to sequence $\{\Delta Y_t^*\}$, which is stationary when $\alpha = 1$, by $\alpha_{\Delta Y^*}$.

THEOREM 4.1: *Assume that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$ *and that the stationary sequence* $\{\epsilon_t\}$ *is a martingale difference sequence with* $E|\epsilon_t|^{4+\delta} < \infty$ *for some* $\delta > 0$.

- *When* $\alpha < 1$ *assume* $\sum_{h=1}^{\infty}(h+1)^2 \alpha_{Y^*}^{\delta/(4+\delta)}(h) < \infty$.
- *When* $\alpha = 1$ *assume that* ΔY_t^* *is strong mixing*.

Let $\theta = \alpha$, $\hat{\theta}_n = \hat{\alpha}_n$, $\hat{\sigma}_n = n^{1/2} SE_{OLS}(\hat{\alpha}_n)$, *and* $\tau_n^{\bullet} = n^{1/2}$. *Then, conclusions* (*i*)–(*iii*) *of Theorem* 3.2 *and Corollary* 3.2 *hold*.

REMARK 4.1: Note that unlike the (conventional) bootstrap, the subsampling method can handle discontinuities of the limiting distribution of estimators as a function of the underlying model parameters. The intuition is that the subsampling approximation of the sampling distribution of an estimator is based on subsample statistics computed from smaller blocks of the observed data. The subsample statistics are therefore always generated from the true model. The bootstrap, on the other hand, bases its approximation on pseudo statistics computed from pseudo data according to a bootstrap distribution, which was estimated from the observed time series. The bootstrap data come from a model close to the truth, but not exactly the truth and this can cause the bootstrap to fail. A case in point is the parameter α in AR(p) models, where the (conventional) bootstrap is inconsistent; see Basawa et al. (1991). However, if one is willing to assume i.i.d. residuals, it is possible to invert bootstrap tests for α to construct asymptotically valid confidence intervals; see Hansen (1999).

REMARK 4.2: We have presented a result that allows construction of asymptotically valid confidence intervals for any *fixed* $\alpha \in (-1, 1]$. Strictly speaking, this problem is already solved by a pretest method, at least when the residuals are assumed i.i.d. (we would like to thank an anonymous referee for pointing this out). The idea is to test for a unit root—using a significance level tending to zero with the sample size—and to base the confidence interval on the normal approximation, when the test rejects the null, or to set it equal to the singleton unity, otherwise. However, it is well-known that this method has terrible finite sample properties; this is one of the reasons for considering local-to-unity asymptotics such as in Stock (1991). The problem with the pretest method is seen by the fact that it applies one of two inherently *different* types of confidence intervals—normal interval or singleton unity—depending on the outcome of a test with low power in finite samples. Hence, quite often the "false" interval will be used, resulting in poor coverage. On the other hand, the subsampling method avoids this pitfall, since it employs one *unique* construction that works both when $\alpha < 1$ and when $\alpha = 1$. The intuition that subsampling should therefore lead to good finite sample properties is confirmed by some simulation studies in Section 6.

REMARK 4.3: An important "byproduct" of the proof of Theorem 4.1 is the fact that, when $\alpha < 1$, the t-statistic for $\hat{\alpha}_n$ has a limiting normal distribution with mean 0 but with variance that can be arbitrarily different from 1 if the

innovations are allowed to be a martingale difference sequence (m.d.s.) rather than i.i.d.; note that m.d.s.-type innovations cannot be transformed to i.i.d. innovations by increasing the order of the AR(p) model. This is an important result, since inference for α—at least when it is known that $\alpha < 1$—is often based on standard OLS output. However, this inference can be arbitrarily misleading, unless the innovations are known to be i.i.d. In the same way, any other inference method for α that assumes i.i.d. innovations—such as Andrews and Chen (1994) or Hansen (1999)—is equally affected in the trend-stationary case $\alpha < 1$. Subsampling, on the other hand, offers a safety net against innovations that are a m.d.s. Note that when attention is restricted to the case $\alpha < 1$ the assumption of a m.d.s. could be relaxed to an uncorrelated innovation sequence.

REMARK 4.4: The assumption of stationarity of the innovation process $\{\epsilon_t\}$ is made to ensure the stationarity of the subsample statistics $\{\hat{\alpha}_{b,t}, t = 1, \ldots, n - b + 1\}$. However, by extending the general theory along the lines of Politis, Romano, and Wolf (1997), this assumption could be relaxed to allow for heteroskedasticity and/or changing distributions of the ϵ_t; see Remark 3.2.

4.2. *Confidence Intervals for α in Models without Time Trend*

Sometimes it may be known a priori that $\beta^* = 0$ in model (1). It is then desirable to incorporate this knowledge in making inference on α. But, the above restriction implies $\beta = 0$ in model (2). Therefore, the knowledge can be incorporated through computing the restricted version of $\hat{\alpha}_n$ by applying OLS to model (2) excluding the time trend. Denote the restricted version of $\hat{\alpha}_n$ by $\hat{\alpha}_n^{\beta=0}$. Also, denote the corresponding OLS standard error by $SE_{OLS}(\hat{\alpha}_n^{\beta=0})$.

The application of the subsampling method is analogous to the general model and is based on computing the restricted version $\hat{\alpha}_b^{\beta=0}$ on all the subsamples of size b. The following corollary shows that the ensuing confidence intervals also have asymptotically correct coverage probability.

COROLLARY 4.1: *Make the same assumptions as in Theorem* 4.1. *In addition, assume that* $\beta^* = 0$ *in model* (1). *Let* $\theta = \alpha$, $\hat{\theta}_n = \hat{\alpha}_n^{\beta=0}$, $\hat{\sigma}_n = n^{1/2} SE_{OLS}(\hat{\alpha}_n^{\beta=0})$, *and* $\tau_n^{\bullet} = n^{1/2}$. *Then, conclusions* (i)–(iii) *of Theorem* 3.2 *and Corollary* 3.2 *hold*.

REMARK 4.5: Consider the case where a restricted model holds, that is, where the time trend is equal to zero. One can then base the estimation of α on the restricted model or on the full model. The proof of the above corollary shows that when $\alpha < 1$ the asymptotic distribution of $\hat{\alpha}_n^{\beta=0}$ is equal to that of $\hat{\alpha}_n$. Therefore, there is no asymptotic efficiency loss for unnecessarily including the time trend in the estimation process; it stands to reason, though, that finite sample performance will be affected. On the other hand, when $\alpha = 1$ the asymptotic distribution of $\hat{\alpha}_n^{\beta=0}$ does differ from that of $\hat{\alpha}_n$. Therefore, even asymptotically it is beneficial to exclude the time trend from the estimation procedure if it is equal to zero indeed.

4.3. *Confidence Intervals for Other Parameters*

The general results of Section 3 also allow for the construction of confidence intervals for parameters of interest other than α and the details are straightforward and left to the reader.

To give only one example, consider confidence intervals for any regression coefficient of model (3) in the trend-stationary case $\alpha < 1$. The common inference is based on the limiting standard normality of the t-statistic of the corresponding OLS estimator. However, as the proof of Proposition A.1 in the Appendix shows, if the innovations are a m.d.s., this inference is again misleading because the limiting variance of the t-statistic is then in general not equal to 1. On the other hand, subsampling is robust in this respect. Since the rate of convergence is well-known for all parameters—given by $n^{1/2}$ for the γ_i and μ and by $n^{3/2}$ for β, respectively—the basic subsampling approach of Subsection 3.1 can be employed. Alternatively, the studentized approach of Subsection 3.2 is available as well.

5. CHOICE OF THE BLOCK SIZE

A practical issue in constructing subsampling intervals is the choice of the block size b and it can be compared to the problem of choosing the bandwidth for kernel methods. In this section, we propose two methods to select b in practice. The first one is very general and can be used whenever subsampling applies. The second one tries to exploit the semi-parametric structure of the AR(p) model with linear time trend.

5.1. *Minimizing Confidence Interval Volatility*

This general approach is of a heuristic nature and we do not claim any optimality properties. It is based on the fact that, in order for the subsampling method to be consistent, the block size b needs to tend to infinity with the sample size n, but at a smaller rate satisfying $b/n \to 0$. Indeed, for b too close to n all subsample statistics $\hat{\theta}_{b,t}$ will almost equal to $\hat{\theta}_n$, resulting in the subsampling distribution being too tight and in undercoverage of subsampling confidence intervals. If b is too small, the intervals can undercover or overcover depending on the state of nature (e.g., Politis, Romano, and Wolf (1997)). This leaves a number of b values in the "right range" where we would expect almost correct results, at least for big sample sizes. This idea is exploited by computing subsampling intervals for a number of block sizes b and then looking for a region where the intervals do not change very much. Within this region, an interval is then picked according to some arbitrary criterion.

This idea is illustrated in Figure 1. For two data sets, symmetric subsampling intervals are computed for a wide range of block sizes b. The "right ranges" extend from $b = 10$ to about $b = 130$ for the first data set and from $b = 10$ to about $b = 80$ for the second data set. The fact that for very large block sizes the

confidence intervals will shrink towards the singleton $\hat{\alpha}_n$ is a consequence of the fact that the subsampling approximation of the sampling distribution of $n^{1/2}(\hat{\alpha}_n - \alpha)/\hat{\sigma}_n$ collapses to a point mass at zero as the block size b tends to n.

While this method can be carried out by "visual inspection," it is desirable to also have some automatic selection procedure, especially when simulation studies are to be carried out. The procedure we propose is based on minimizing a running standard deviation. Assume we compute subsampling intervals for block sizes b in the range of b_{small} to b_{big}. The endpoints of the confidence intervals will change in a smooth fashion as b changes. A running standard deviation applied to the endpoints then determines the volatility around a specific b value. We choose the value of b with the smallest volatility. Here is a more formal description of the algorithm.

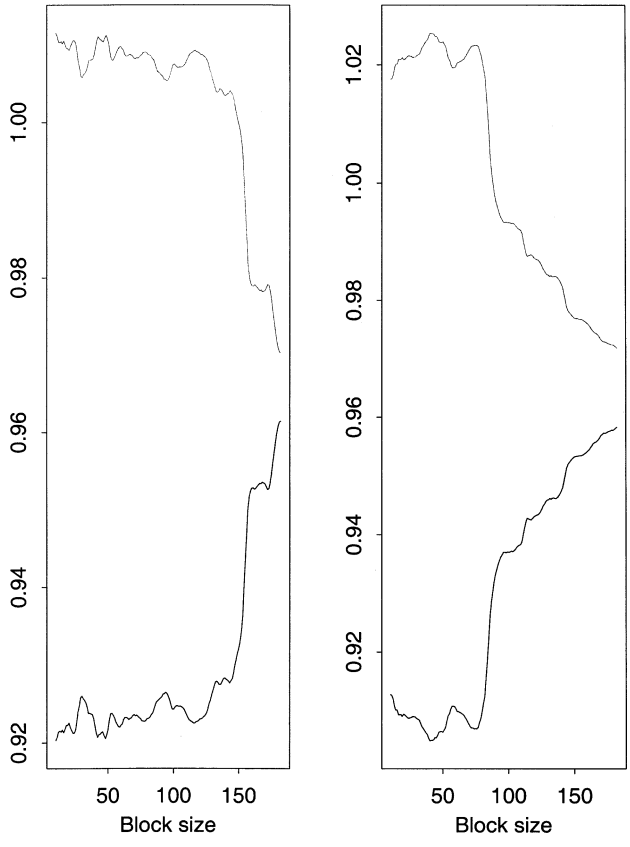


FIGURE 1.—Confidence intervals as function of the block size b for two AR(1) data sets; the x-axis shows the block size b while the y-axis shows the upper and lower confidence interval endpoints. The data were generated according to model (1) with $\mu^* = \beta^* = 0$, $\alpha = 0.99$, $n = 200$, and i.i.d. standard normal innovations. The intervals are nominal 95% two-sided symmetric intervals based on the studentized approach of Subsection 3.2 and the block size ranges from $b = 10$ to $b = 180$.

ALGORITHM 5.1 (*Minimizing Confidence Interval Volatility*):

1. For $b = b_{small}$ to $b = b_{big}$ compute a subsampling interval for θ at the desired confidence level, resulting in endpoints $I_{b,low}$ and $I_{b,up}$.

2. For each b compute a volatility index VI_b as the standard deviation of the interval endpoints in a neighborhood of b. More specifically, for a small integer k, let VI_b be equal to the standard deviation of $\{I_{b-k,low}, \ldots, I_{b+k,low}\}$ plus the standard deviation of $\{I_{b-k,up}, \ldots, I_{b+k,up}\}$.

3. Pick the value b^* with the smallest volatility index and report $[I_{b^*,low}, I_{b^*,up}]$ as the final confidence interval.

Some remarks concerning the implementation of this algorithm are in order.

REMARK 5.1: The range of b values, determined by b_{small} and b_{big}, which is included in the minimization algorithm, is not of crucial importance. On the other hand, to keep the computational cost down as well as to "enforce" the requirements $b \to \infty$ and $b/n \to 0$ as $n \to \infty$, it is sensible to choose $b_{small} = c_1 n^\eta$ and $b_{big} = c_2 n^\eta$ for constants $0 < c_1 < c_2$ and $0 < \eta < 1$. We recommend $c_1 \in [0.5, 1]$, $c_2 \in [2, 3]$, and $\eta = 0.5$.

REMARK 5.2: The algorithm contains a model parameter k. Simulation studies have shown that the algorithm is very insensitive to its choice. We recommend $k = 2$ or $k = 3$.

We now illustrate how the algorithm works with the help of two simulated data sets. First, we generated a time series of size according to model (1) with $\mu^* = \beta^* = 0$, $\alpha = 0.95$, $n = 200$, and i.i.d. standard normal innovations. The range of b values was chosen as $b_{small} = 10$ and $b_{big} = 40$. The minimization of the volatility in Step 2 was done using $k = 2$. The results are shown at the top of Figure 2. The left plot corresponds to equal-tailed confidence intervals while the right plot corresponds to symmetric confidence intervals. The block sizes b chosen by the algorithm are highlighted by a star. The resulting final confidence intervals are included in the plots together with the point estimate $\hat{\alpha}_n$.

This exercise was repeated for another data set according to model (1) with $\mu^* = \beta^* = 0$, $\alpha = 1$, $n = 500$, and i.i.d. standard normal innovations. The range of b there was chosen as $b_{small} = 15$ and $b_{big} = 60$. The results are shown at the bottom of Figure 2.

The plots show that symmetric intervals are somewhat more stable, that is, the endpoints change less as b is varied. This behavior is typical and was observed for many other simulations as well.

5.2. *Choosing b according to an Estimated Model*

The idea underlying the second approach is that the optimal finite sample block size for a specific nominal coverage probability $1 - \lambda$ could be calculated, or at least simulated, if the true data generating mechanism was known. Using

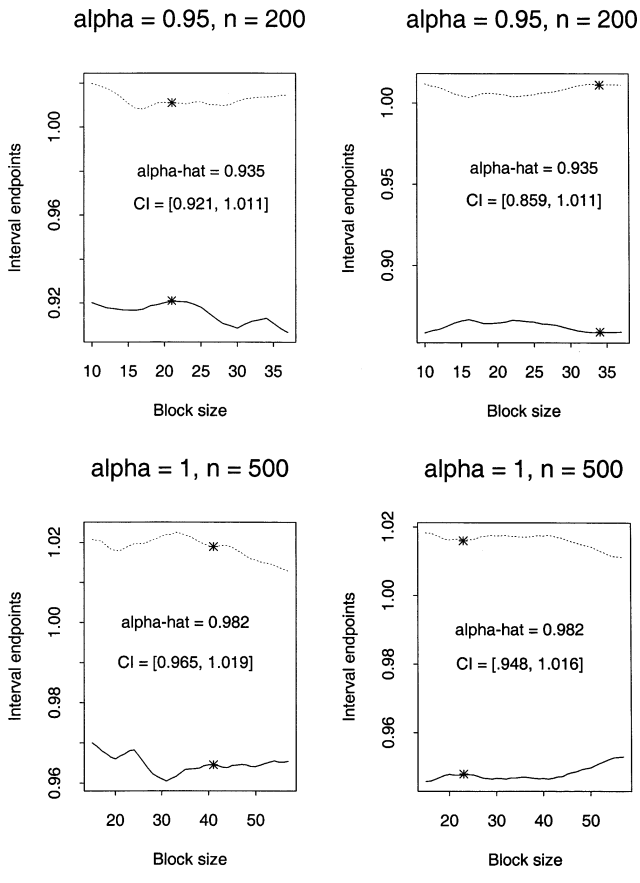


FIGURE 2.—Illustration of the Minimizing Confidence Interval Volatility Algorithm for two data sets. The plots on the left correspond to equal-tailed confidence intervals, while the plots on the right correspond to symmetric confidence intervals; both interval types are based on the studentized approach of Subsection 3.2. The block sizes selected by the algorithm are highlighted by a star. The final confidence intervals appear within the plots together with the point estimates.

the simulation method, one would generate a large number K, say, of time series according to the true mechanism (with the same sample size as the observed series), construct subsampling intervals using a number of different block sizes for each generate series, and compute the estimated coverage probability for each block size as the fraction of the corresponding B intervals that contain the true parameter. One then would use the block size whose estimated coverage probability is closest to $1 - \lambda$. Of course, this method is not feasible, since the true data generating process is generally unknown.

However, it is reasonable to hope that a feasible variant of this method will still yield useful results in case the true data generating mechanism can be consistently estimated. In that case one would use the above algorithm with the

estimated process in place of the true process. For a completely nonparametric application, it is in general not clear how to consistently estimate the underlying mechanism. Our application, on the other hand, is of semi-parametric nature depending on $p+2$ real-valued parameters, each of which can be consistently estimated by OLS, say, and the probability mechanism of the white noise innovation sequence, which can be consistently estimated by applying a time series bootstrap to the estimated innovations, say the moving blocks bootstrap (Künsch (1989), Liu and Singh (1992)) or the stationary bootstrap (Politis and Romano (1994b)). While it is well-known that this residual based bootstrap yields inconsistent results when used directly, that is, to approximate the sampling distribution of $\hat{\alpha}_n$ (e.g. Basawa et al. (1991)), it yields consistent results when used indirectly, that is, to estimate the optimal block size of the subsampling method. This is a simple consequence of the fact that any method of picking one of several "competing" block sizes, even coin-tossing, would yield consistent results as long as the block sizes included in the 'contest' satisfy the regularity conditions $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. The point is that when using an estimated model in picking the block size, one should expect better finite sample properties as compared to coin tossing.

To provide a somewhat more formal description, introduce the notion of a *calibration* function h: $b \to -\kappa$ that expresses the true coverage probability of a nominal $1-\lambda$ confidence interval as a function of the block size b that is used in constructing the interval. If $h(\cdot)$ was known, one could construct an interval with perfect coverage by employing a block size b with $h(b) = 1-\lambda$ (provided that such a solution exists). While the true $h(\cdot)$ is unknown, we can approximate it as previously suggested. The estimated data generating mechanism is based on OLS estimation of model (3) and the resulting estimates $\hat{\mu}, \hat{\beta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_p$, $\hat{\epsilon}_{p+1}, \ldots, \hat{\epsilon}_n$ where the subscript n corresponding to estimation based on n data points has been suppressed. To generate a corresponding "estimated" or pseudo sequence, we start by applying a time series bootstrap to the estimated innovations to obtain pseudo innovations $\epsilon^*_{p+1}, \ldots, \epsilon^*_n$. The pseudo sequence is then defined by the recursive relation

$$Y_t^* = Y_t \qquad (t = 1, \ldots, p), \tag{8}$$

$$Y_t^* = \hat{\mu} + \hat{\beta} t + \hat{\gamma}_1 Y^*_{t-1} + \cdots \hat{\gamma}_p Y^*_{t-p} + \epsilon_t^* \qquad (t = p+1, \ldots, n).$$

The following then is the algorithm corresponding to the above calibration idea. It is stated for a general parameter θ and its corresponding estimate $\hat{\theta}_n$. Of course, we are mainly interested in α and $\hat{\alpha}_n$, but the algorithm equally applies to any other parameters of interest; see Subsection 4.3.

ALGORITHM 5.2 (*Block Size Calibration*):

1. Generate K pseudo sequences $Y^*_{1,k}, \ldots, Y^*_{n,k}$, according to (8). For each sequence $k = 1, \ldots, K$:

1a. Compute a $1-\lambda$ level confidence interval $\mathrm{CI}^k_{b_j}$, for a grid of block sizes $b_{min} \le b_j \le b_{max}$.

2. For each b_j compute $\hat{h}(b_j) = \#\{\hat{\theta}_n \in \mathrm{CI}_{b_j}^k\}/K$.
3. Find the value of b_j with $\hat{h}(b_j)$ closest to $1 - \lambda$.
4. Construct a confidence interval using the block size b_j.

REMARK 5.3: Algorithm 5.2 is related to adjusting the nominal level of a confidence interval so that its actual level better matches the desired level in finite samples, an idea that dates back to Loh (1987). However, to this end the (standardized) sampling distribution of $\hat{\theta}_n$ under the estimated mechanism must be a consistent approximation of the (standardized) sampling distribution of $\hat{\theta}_n$ under the true distribution for the resulting confidence intervals to have asymptotically correct coverage probability. As mentioned before, this condition is violated in our application.

REMARK 5.4: It is clear that the grid of (subsampling) block sizes to be used in Algorithm 5.2 should be as fine as possible within the limitations $b_{min} \to \infty$ and $b_{max}/n \to 0$ as $n \to \infty$. Moreover, b_{min} and b_{max} play roles analogous to b_{small} and b_{big} in Algorithm 5.1 and can be picked in a similar fashion. However, at least for simulation studies, including every integer number between b_{min} and b_{max} might be computationally too expensive, in which case an appropriate subset can be selected.

REMARK 5.5: As far as the choice of the time series bootstrap is concerned, we prefer the stationary bootstrap, since it is well-known to be less sensitive to the choice of the (bootstrap) block size than the moving blocks bootstrap and also because it does not suffer from the "end effects" of the latter; see Politis and Romano (1994b).

6. SMALL SAMPLE PERFORMANCE

The purpose of this section is to examine the small sample performance of the subsampling confidence intervals via some simulation studies. Performance is mainly measured by coverage probability of two-sided nominal 95% intervals for the parameter α. We also look at median length. The approach of Subsection 3.2, subsampling a studentized statistic, is employed, using both equal-tailed and symmetric intervals (denoted by indices ET and SYM). Moreover, the two methods of Section 5 for choosing the block size are employed (denoted by indices MV and CA standing for Minimum Volatility and CAlibration). The four resulting interval types are labeled $\mathrm{Sub}_{MV,ET}$, $\mathrm{Sub}_{MV,SYM}$, $\mathrm{Sub}_{CA,ET}$, and $\mathrm{Sub}_{CA,SYM}$. For comparison, the normal method, which bases the confidence interval on asymptotic standard normality of the t-statistic for $\hat{\alpha}_n$, and the method of Stock (1991) are also included. These two intervals are labeled CLT and Stock, respectively.

Some brief remarks concerning Stock's intervals are in order. First, his intervals are for the largest root ρ instead of the sum of the AR(p) coefficients α. These two parameters only coincide when $p = 1$. Hence, Stock's intervals are

only included in the simulations for the AR(1) case. Next, Table A.1 of Stock (1991) allows one, up to some minor interpolation, to check whether a particular value of ρ is contained in the confidence intervals in a way that can be automated for simulations. However, the computation of the actual intervals—and thus their length—requires a graphical device (Figures 1 and 2 of Stock (1991)) the automization of which seems very cumbersome. For this reason, only coverage, but not median length, is reported.

6.1. *AR*(1) *Model*

We start with the most simple model, namely AR(1). The data are generated according to model (1) with $\mu^* = \beta^* = 0$ and α one of the following: 1, 0.99, 0.95, 0.9, or 0.6; note that the value $\alpha = 0.6$ is too far away from 1 to be handled by Stock's intervals. The statistic $\hat{\alpha}_n$ is the OLS estimator $\hat{\gamma}_{1,n}$ based on model (3). The innovations ϵ_t are either i.i.d. standard normal or of the form $\epsilon_t = Z_t Z_{t-1}$ with the Z_t i.i.d. standard normal. In the latter case, the innovations are a martingale difference sequence but dependent. The sample sizes considered are $n = 120$ and $n = 240$. The range of b values used for Algorithm 5.1 is determined by $b_{small} = 5$ and $b_{big} = 25$ when $n = 120$ and by $b_{small} = 10$ and $b_{big} = 40$ when $n = 240$, respectively. The grid of block sizes b_j for Algorithm 5.2 is 5, 8, 12, 18, 26, 40 when $n = 120$ and 10, 15, 20, 30, 40, 60 when $n = 240$. The reason for not employing an equally spaced grid is that some prior simulations (with finer grids) showed that the estimated calibration function $\hat{h}(\cdot)$ changes more rapidly for smaller block sizes.

The results are presented in Table I. It is seen that equal-tailed subsampling intervals perform worse than symmetric intervals in general. The two methods for choosing the block size are comparable. Next, one notes the well-known fact that the CLT approach does not work when $\alpha = 1$ and that it works poorly when α is close to 1. Stock's intervals, on the other hand, have rather accurate coverage when they apply. If dependent innovations of the form $\epsilon_t = Z_t Z_{t-1}$ are employed, the CLT breaks down. For $\alpha = 0.6$, the coverage drops to about 80%. In fact it can be shown that, for innovations of the form $\epsilon_t = Z_t Z_{t-1} \cdots Z_{t-k}$, the coverage of CLT intervals will tend to 0 as k tends to infinity; see Romano and Thombs (1996). Also, it appears that Stock's intervals are somewhat less robust against dependence as compared to subsampling intervals (especially for $n = 120$).

The difference in empirical coverage between equal-tailed and symmetric subsampling intervals is noteworthy. A possible explanation is that the equal-tailed interval is based on estimating a 2.5% and a 97.5% quantile while the symmetric interval is based on estimating a (single) 95% quantile, and it is conceivable that the latter can be estimated with higher precision. One way to examine this issue would be to redo the above table for a number of different confidence levels, such as 90% and 80%. Instead, we opt for considering all levels "simultaneously" by exploiting the duality between confidence intervals

TABLE I

ESTIMATED COVERAGE PROBABILITIES OF VARIOUS NOMINAL 95% CONFIDENCE INTERVALS BASED ON 1000 REPLICATIONS FOR EACH SCENARIO

α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT	Stock
			AR(1) Model, $n = 120$, $\epsilon_t = Z_t$			
1	0.90	0.91	0.95	0.90	0.38	0.96
0.99	0.90	0.93	0.95	0.92	0.51	0.95
0.95	0.83	0.95	0.84	0.94	0.78	0.94
0.9	0.82	0.96	0.71	0.95	0.84	0.94
0.6	0.77	0.95	0.66	0.95	0.93	NA
			AR(1) Model, $n = 240$, $\epsilon_t = Z_t$			
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT	Stock
1	0.91	0.92	0.88	0.91	0.37	0.96
0.99	0.91	0.94	0.85	0.93	0.62	0.94
0.95	0.80	0.96	0.78	0.95	0.83	0.95
0.9	0.75	0.96	0.73	0.96	0.88	0.97
0.6	0.84	0.96	0.79	0.95	0.95	NA
			AR(1) Model, $n = 120$, $\epsilon_t = Z_t Z_{t-1}$			
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT	Stock
1	0.95	0.94	0.96	0.94	0.36	0.92
0.99	0.95	0.96	0.97	0.96	0.51	0.93
0.95	0.89	0.95	0.94	0.97	0.75	0.91
0.9	0.82	0.96	0.87	0.96	0.81	0.91
0.6	0.75	0.94	0.78	0.95	0.81	NA
			AR(1) Model, $n = 240$, $\epsilon_t = Z_t Z_{t-1}$			
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT	Stock
1	0.94	0.93	0.97	0.94	0.36	0.93
0.99	0.94	0.97	0.95	0.96	0.63	0.94
0.95	0.81	0.97	0.84	0.97	0.83	0.93
0.9	0.78	0.96	0.78	0.96	0.85	0.95
0.6	0.74	0.93	0.72	0.94	0.80	NA

The data were generated according to model (1) with $\mu^* = \beta^* = 0$ and $p = 1$. The innovations are either i.i.d. standard normal ($\epsilon_t = Z_t$) or of the form $\epsilon_t = Z_t Z_{t-1}$. All intervals were obtained by including a time trend in the fitted model.

and hypothesis tests to calculate corresponding *P*-values. Hence, the *P*-value is given by 1 minus the confidence level of the interval that "barely excludes" a hypothesized value α_0. Fortunately, this number can be directly computed. For example, the symmetric studentized *P*-value is given by

$$P\text{-value}^{\bullet}_{n,b}(\alpha_0) = \frac{\#\{b^{1/2}(\hat{\alpha}_{b,t} - \hat{\alpha}_n)/\hat{\sigma}_{b,t} \geq n^{1/2}(\hat{\alpha}_n - \alpha_0)/\hat{\sigma}_n\}}{n - b + 1}$$

and the remaining *P*-values are computed analogously. It is well known that if α_0 is equal to the true parameter α, then the distribution of a *P*-value corresponding to an exact hypothesis test (or, equivalently, to an exact confidence interval) is given by Uniform[0, 1], provided that this distribution is continuous. Hence, one can judge the accuracy of hypothesis test (or the corresponding confidence intervals) by a Q-Q-plot of a large number of simu-

lated P-values (for the true parameter) against the Uniform[0, 1] distribution. We do this in Figure 3 for the two subsampling interval types and the CLT interval; Stock's method is excluded, since the tables in Stock (1991) do not allow computation of P-values. To generate the data, we use i.i.d. standard normal innovations, $n = 240$, and the three values $\alpha = 1$, 0.95, and 0.6. A fixed block size of $b = 25$ is used for all subsampling intervals; this is somewhat suboptimal, since the best fixed block size in general depends on the approach used as well as on the true underlying parameter (and possibly even on the nominal level of the interval).

The plots show that the two subsampling interval types are qualitatively rather different. The equal-tailed studentized intervals work well when $\alpha = 1$ but generally undercover when $\alpha < 1$. On the other hand, symmetric studentized intervals are relatively accurate at large confidence levels throughout, but at smaller confidence level (70%, say) undercover when $\alpha = 1$ and overcover when $\alpha < 1$. In addition, once more we observe the well-known fact that the CLT intervals work well for α far away from 1, but undercover increasingly as α gets closer to one.

As discussed in Subsection 4.2, when it is known a priori that $\beta^* = 0$ in model (1) this knowledge should and can be incorporated in constructing confidence intervals for α. It is of interest to compare the gain in efficiency, that is, in interval length in those instances. We do this by computing median length of the 1000 confidence intervals in each scenario for the two methods of using the full model and of using the restricted model without time trend (the latter is appropriate, since we employ $\mu^* = \beta^* = 0$ when generating the data). Of course, it is also of interest to compare the median length of the various interval types. The results are presented in Table II. The gain from excluding the time trend when it is indeed not needed is substantial when α is equal to or close to 1, but it generally decreases as α gets further away from 1. This is not surprising because asymptotically there is a gain in rightfully omitting the time trend from the model when $\alpha = 1$ but there is none when $\alpha < 1$; see Remark 4.5. Moreover, symmetric studentized subsampling intervals are about as good as the CLT intervals when both have approximately correct coverage, that is, when $\alpha = 0.6$. Note that we also computed empirical coverage for confidence intervals constructed without time trend. The results were similar to those of Table I and thus are not reported.

6.2. *AR*(2) *Model*

The data generating mechanism is now model (1), with $p = 2$, α one of the values 1, 0.99, 0.95, 0.9, and 0.6, and ψ_1 equal to either 0.4 or to -0.4. The statistic $\hat{\alpha}_n$ is computed as $\hat{\gamma}_{1,n} + \hat{\gamma}_{2,n}$ where the OLS estimation is based on model (3) again. Everything else is as in Subsection 6.1. Tables III and IV provide empirical coverage for nominal 95% confidence intervals. The results are qualitatively comparable to the AR(1) case.

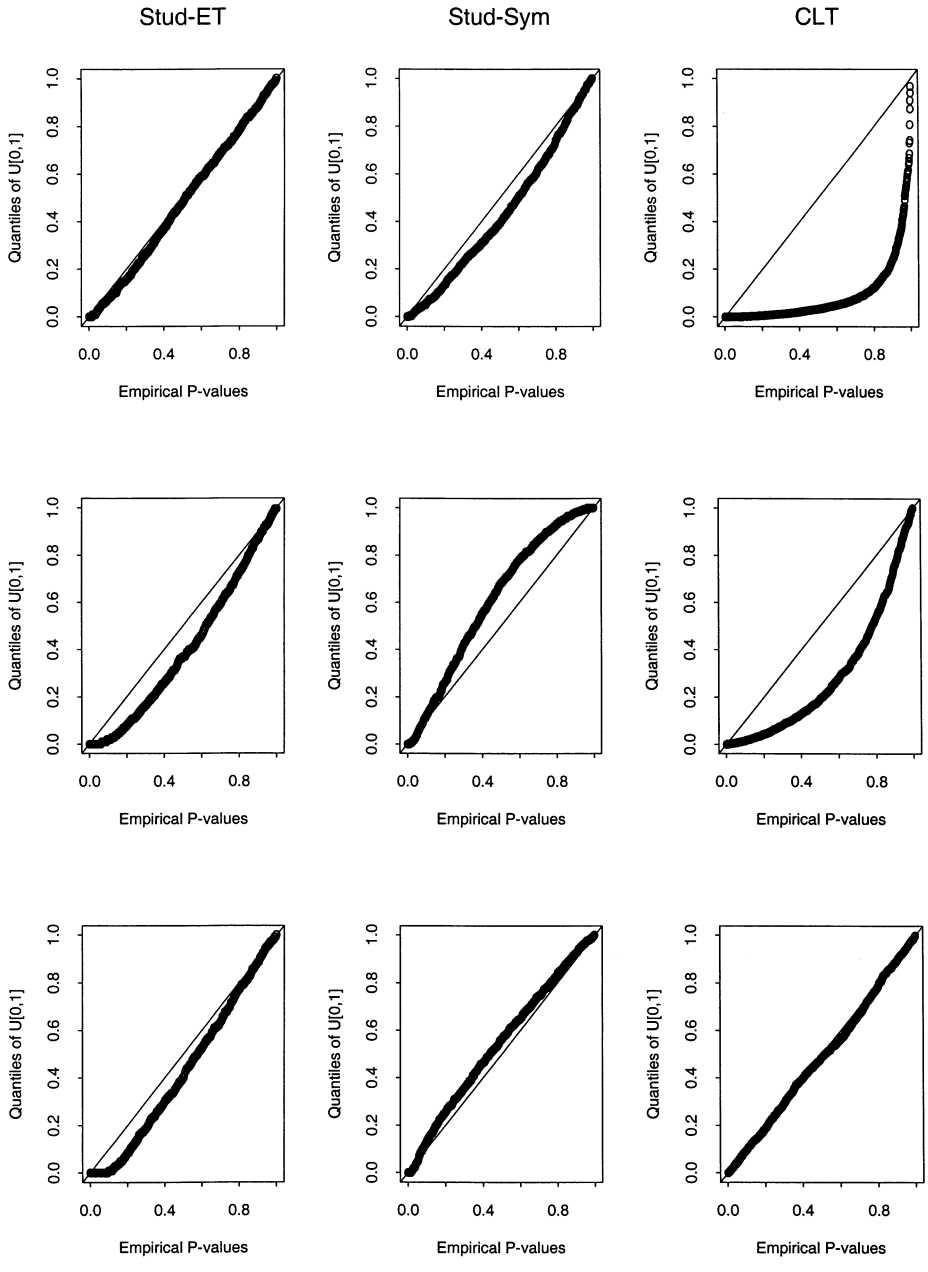


FIGURE 3.—Q-Q-plots of 1000 empirical P-values against Uniform$[0,1]$. The data were generated according to model (1) with $\mu^* = \beta^* = 0$, $n = 240$, and i.i.d. standard normal innovations. The three values for α, from top to bottom, are 1, 0.95, and 0.6. A straight line through the origin with slope 1 is included in all plots.

TABLE II

MEDIAN LENGTH OF VARIOUS NOMINAL 95% CONFIDENCE INTERVALS BASED ON 1000 REPLICATIONS FOR EACH SCENARIO

AR(1) Model including Time Trend, $n = 120$, $\epsilon_t = Z_t$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.12	0.24	0.24	0.24	0.14
0.99	0.13	0.23	0.24	0.23	0.14
0.95	0.14	0.26	0.22	0.25	0.17
0.9	0.17	0.29	0.13	0.28	0.20
0.6	0.26	0.36	0.18	0.37	0.31
AR(1) Model excluding Time Trend, $n = 120$, $\epsilon_t = Z_t$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.09	0.13	0.20	0.21	0.09
0.99	0.10	0.15	0.21	0.22	0.11
0.95	0.13	0.19	0.28	0.23	0.14
0.9	0.16	0.22	0.34	0.30	0.18
0.6	0.25	0.31	0.28	0.35	0.30
AR(1) Model including Time Trend, $n = 240$, $\epsilon_t = Z_t$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.06	0.11	0.05	0.12	0.07
0.99	0.06	0.12	0.06	0.12	0.07
0.95	0.08	0.15	0.07	0.15	0.10
0.9	0.10	0.18	0.09	0.17	0.12
0.6	0.18	0.23	0.15	0.24	0.21
AR(1) Model excluding Time Trend, $n = 240$, $\epsilon_t = Z_t$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.04	0.06	0.04	0.07	0.05
0.99	0.05	0.08	0.05	0.09	0.06
0.95	0.08	0.11	0.07	0.12	0.09
0.9	0.10	0.14	0.09	0.15	0.12
0.6	0.18	0.21	0.17	0.23	0.21

The data were generated according to model (1) with $\mu^* = \beta^* = 0$ and $p = 1$. The innovations are i.i.d. standard normal.

7. SUMMARY

In this paper, we proposed a new way of constructing confidence intervals in AR(p) models with linear time trend. While the focus was on the parameter α, the sum of the AR(p) coefficients, the method is general enough to cover essentially any other parameter of interest as well. The crux of the new approach is to recompute an estimator on smaller blocks of the observed data to approximate the sampling distribution of the estimator computed from the entire sequence. This is the general idea of the subsampling method of Politis and Romano (1994a).

The subsampling method overcomes the notorious difficulty in the construction of confidence intervals for α, namely that the limiting distribution, as well as the rate of convergence, of the OLS estimator $\hat{\alpha}_n$ depend in a discontinuous way upon whether or not $\alpha < 1$. Some extensions of previous theory were necessary to handle the unknown convergence rate. The approach suggested is

TABLE III

ESTIMATED COVERAGE PROBABILITIES OF VARIOUS NOMINAL 95% CONFIDENCE INTERVALS BASED ON 1000 REPLICATIONS FOR EACH SCENARIO

α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
AR(2) Model, $\psi_1 = 0.4$, $n = 120$, $\epsilon_t = Z_t$					
1	0.93	0.92	0.97	0.91	0.37
0.99	0.92	0.95	0.96	0.93	0.58
0.95	0.89	0.96	0.90	0.94	0.85
0.9	0.82	0.97	0.85	0.95	0.89
0.6	0.80	0.96	0.82	0.96	0.94
AR(2) Model, $\psi_1 = 0.4$, $n = 240$, $\epsilon_t = Z_t$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.91	0.92	0.94	0.91	0.37
0.99	0.90	0.94	0.89	0.93	0.57
0.95	0.81	0.96	0.80	0.95	0.80
0.9	0.78	0.96	0.80	0.95	0.87
0.6	0.79	0.97	0.78	0.96	0.94
AR(2) Model, $\psi_1 = 0.4$, $n = 120$, $\epsilon_t = Z_t Z_{t-1}$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.95	0.92	0.97	0.95	0.37
0.99	0.94	0.95	0.97	0.95	0.60
0.95	0.89	0.97	0.93	0.96	0.84
0.9	0.86	0.97	0.87	0.96	0.89
0.6	0.80	0.96	0.79	0.95	0.91
AR(2) Model, $\psi_1 = 0.4$, $n = 240$, $\epsilon_t = Z_t Z_{t-1}$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.94	0.93	0.96	0.94	0.39
0.99	0.93	0.96	0.90	0.95	0.68
0.95	0.87	0.96	0.84	0.96	0.88
0.9	0.85	0.95	0.78	0.95	0.89
0.6	0.82	0.94	0.70	0.96	0.91

The data were generated according to model (1) with $\mu^* = \beta^* = 0$ and $p = 2$. The innovations are either i.i.d. standard normal ($\epsilon_t = Z_t$) or of the form $\epsilon_t = Z_t Z_{t-1}$. All intervals were obtained by including a time trend in the fitted model.

based on subsampling the t-statistic for $\hat{\alpha}_n$, which has a nondegenerate limiting distribution no matter what the value of α. The theory is flexible enough to allow innovations of the AR(p) series to be a stationary martingale difference sequence rather than an i.i.d. sequence (but even the assumption of stationarity could be relaxed). This flexibility was seen to be of practical relevance, since in the trend-stationary case $\alpha < 1$, the standard inference on α—but as well on other parameters of interest—can be arbitrarily misleading when the innovations are not independent.

Finite sample performance was examined through some simulation studies and was seen to be satisfactory, at least when symmetric subsampling intervals are used. The results were most favorable in the case of dependent innovations, since in this case the CLT intervals break down (even for α far away from 1) and Stock's (1991) intervals, which are asymptotically valid as well, seem somewhat more affected than subsampling intervals in small samples.

Finally, it should be pointed out that subsampling is a very general and powerful technique and not restricted to inference in AR(p) models. Basically,

TABLE IV

ESTIMATED COVERAGE PROBABILITIES OF VARIOUS NOMINAL 95% CONFIDENCE INTERVALS BASED ON 1000 REPLICATIONS FOR EACH SCENARIO.

α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
AR(2) Model, $\psi_1 = -0.4$, $n = 120$, $\epsilon_t = Z_t$					
1	0.93	0.91	0.97	0.94	0.40
0.99	0.92	0.93	0.96	0.96	0.50
0.95	0.90	0.95	0.90	0.97	0.71
0.9	0.81	0.95	0.85	0.97	0.82
0.6	0.77	0.96	0.82	0.95	0.91
AR(2) Model, $\psi_1 = -0.4$, $n = 240$, $\epsilon_t = Z_t$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.94	0.93	0.97	0.93	0.37
0.99	0.90	0.95	0.93	0.94	0.71
0.95	0.84	0.97	0.81	0.95	0.89
0.9	0.78	0.96	0.79	0.97	0.92
0.6	0.84	0.95	0.82	0.96	0.95
AR(2) Model, $\psi_1 = -.04$, $n = 120$, $\epsilon_t = Z_t Z_{t-1}$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.93	0.92	0.97	0.93	0.40
0.99	0.92	0.93	0.93	0.94	0.50
0.95	0.87	0.95	0.87	0.95	0.71
0.9	0.82	0.95	0.83	0.95	0.78
0.6	0.76	0.93	0.78	0.94	0.86
AR(2) Model, $\psi_1 = -0.4$, $n = 240$, $\epsilon_t = Z_t Z_{t-1}$					
α	$\text{Sub}_{MV,ET}$	$\text{Sub}_{MV,SYM}$	$\text{Sub}_{CA,ET}$	$\text{Sub}_{CA,SYM}$	CLT
1	0.93	0.93	0.98	0.94	0.38
0.99	0.91	0.94	0.91	0.95	0.56
0.95	0.80	0.96	0.83	0.96	0.82
0.9	0.75	0.96	0.79	0.96	0.86
0.6	0.76	0.95	0.73	0.95	0.87

The data were generated according to model (1) with $\mu^* = \beta^* = 0$ and $p = 2$. The innovations are either i.i.d. standard normal ($\epsilon_t = Z_t$) or of the form $\epsilon_t = Z_t Z_{t-1}$. All intervals were obtained by including a time trend in the fitted model.

subsampling can be applied beneficially whenever the limiting distribution of an estimator depends on underlying model parameters in a complicated and maybe even discontinuous way; one of many examples is inference in models with integrated or nearly integrated regressors as discussed in Elliot and Stock (1994) and Elliot (1998). The main condition for subsampling to work is that the estimator, properly normalized, has a nondegenerate limiting distribution and that the subsample statistics are weakly dependent in a sufficient way.

Dept. of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305, U.S.A.; romano@stat.stanford.edu

and

Dept. of Economics and Business, Universitat Pompeu Fabra, 08005, Barcelona, Spain

Manuscript received June, 1998; final revision received August, 2000.

APPENDIX: PROOFS OF TECHNICAL RESULTS

PROOF OF THEOREM 3.1: Without loss of generality, we may think of b as a function of n. Therefore, the notational burden can be reduced by omitting the b subscripts. For example, $L_n(\cdot) \equiv L_{n,b}(\cdot)$, $c_n(\lambda) \equiv c_{n,b}(\lambda)$, etc. To simplify the notation further, introduce $q \equiv q_n \equiv n - b + 1$. Let

$$U_n(x) = \frac{1}{q} \sum_{t=1}^{q} 1\{\tau_b(\hat{\theta}_{b,t} - \theta) \le x\}.$$

To prove (i), it suffices to show that $U_n(x)$ converges in probability to $J(x, P)$ for every continuity point x of $J(\cdot, P)$. This can be seen by noting that

$$L_n(x) = \frac{1}{q} \sum_{t=1}^{q} 1\{\tau_b(\hat{\theta}_{b,t} - \theta) + \tau_b(\theta - \hat{\theta}_n) \le x\},$$

so that for every $\epsilon > 0$,

$$U_n(x - \epsilon)1\{E_n\} \le L_n(x)1\{E_n\} \le U_n(x + \epsilon),$$

where $1\{E_n\}$ is the indicator of the event $E_n = \{\tau_b|\theta - \hat{\theta}_n| \le \epsilon\}$. But, the event E_n has probability tending to one. So, with probability tending to one,

$$U_n(x - \epsilon) \le L_n(x) \le U_n(x + \epsilon).$$

Thus, if $x + \epsilon$ and $x - \epsilon$ are continuity points of $J(\cdot, P)$, then $U_n(x \pm \epsilon) \to J(x \pm \epsilon, P)$ in probability implies

$$J(x - \epsilon, P) - \epsilon \le L_n(x) \le J(x + \epsilon, P) + \epsilon$$

with probability tending to one. Now, let $\epsilon \to 0$ such that $x \pm \epsilon$ are continuity points of $J(\cdot, P)$ to conclude that $L_n(x) \to J(x, P)$ in probability as well. Therefore, we may restrict our attention to $U_n(x)$.

Since $E(U_n(x)) = J_b(x)$, the proof of (i) reduces by Assumption 3.1 to showing that $\text{var}(U_n(x))$ tends to zero. Define

$$I_{b,t} = 1\{\tau_b(\hat{\theta}_{b,t} - \theta) \le x\} \qquad (t = 1, \ldots, q),$$

$$s_{q,h} = \frac{1}{q} \sum_{t=1}^{q-h} \text{cov}(I_{b,t}, I_{b,t+h}).$$

Due to a standard mixing inequality for bounded random variables,

$$|\text{cov}(I_{b,t}, I_{b,t+h})| \le 4\alpha_{n,b}(h)$$

and therefore,

$$\text{var}(U_n(x)) = \frac{1}{q}\left(s_{q,0} + 2\sum_{h=1}^{q-1} s_{q,h}\right)$$

$$\le \frac{1}{q}\left(1 + 2\sum_{h=1}^{q-1} \alpha_{n,b}(h)\right) \to 0.$$

This completes the proof of (i).

To prove (ii), given any subsequence $\{n_k\}$, one can extract a further subsequence $\{n_{k_j}\}$ such that $L_{n_{k_j}}(x) \to J(x, P)$ for all x in some countable dense set of the real line, almost surely. It then follows

that, on a set of probability one, $L_{n_{k_j}}(x)$ tends weakly to $J(x, P)$. By the continuity of $J(\cdot, P)$ this convergence is uniform by Polya's theorem.

The proof of (iii) is very similar to the proof of Theorem 1 of Beran (1984) given (i) and is thus omitted. *Q.E.D.*

PROOF OF THEOREM 3.2: Again, let $q = n - b + 1$. To prove (i), note that

$$L^{\bullet}_{n,b}(x) = \frac{1}{q}\sum_{t=1}^{q} 1\{\tau_b^{\bullet}(\hat{\theta}_{b,t} - \hat{\theta}_n)/\hat{\sigma}_{b,t} \le x\} \tag{9}$$

$$= \frac{1}{q}\sum_{t=1}^{q} 1\{\tau_b^{\bullet}(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t} \le x + \tau_b^{\bullet}(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,t}\}.$$

We want to show that the terms $\tau_b^{\bullet}(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,t}$ are negligible in the last equation. To this end, for $u > 0$, let

$$R_n(u) = \frac{1}{q}\sum_{t=1}^{q} 1\{\tau_b^{\bullet}(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,t} \le u\}$$

$$= \frac{1}{q}\sum_{t=1}^{q} 1\{d_b\hat{\sigma}_{b,t} \ge a_b(\hat{\theta}_n - \theta)/u\}.$$

Here, we have used the assumption that both the sequences $\{a_n\}$ and $\{d_n\}$ are positive. By Assumption 3.2 and $a_b/a_n \to 0$, we have for any $\delta > 0$ that $a_b(\theta_n - \theta) \le \delta$ with probability tending to one. Therefore, with probability tending to one

$$R_n(u) \ge \frac{1}{q}\sum_{t=1}^{q} 1\{d_b\hat{\sigma}_{b,t} \ge \delta/u\}.$$

We need to consider the case $u > 0$ only, as the scale estimates $\hat{\sigma}_{b,t}$ are positive. Due to the usual subsampling argument of Theorem 3.1, $(1/q)\sum_{t=1}^{q} 1\{d_b\hat{\sigma}_{bta} \ge \delta/u\}$ converges in probability to $1 - W(\delta/u, P)$, as long as δ/u is a continuity point of $W(P)$; note that we do not require $d_b/d_n \to 0$ here since the subsample statistics are of the form $d_b\hat{\sigma}_{b,t}$ rather than $d_b(\hat{\sigma}_{b,t} - \hat{\sigma}_n)$. Hence, we can make sure that $R_n(u)$ is arbitrarily close to one by choosing δ small enough; recall that we assume $W(P)$ does not have positive mass at zero. In other words, for any $u > 0$, we have $R_n(u) \to 1$ in probability. Let us now rewrite (9) in the following way:

$$L^{\bullet}_{n,b}(x) = \frac{1}{q}\sum_{t=1}^{q} 1\{\tau_b^{\bullet}(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t} \le x + \tau_b^{\bullet}(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,t}\}$$

$$\le \frac{1}{q}\sum_{t=1}^{q} 1\{\tau_b^{\bullet}(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t} \le x + u\} + (1 - R_n(u)),$$

for any positive number u. The last inequality follows because the tth term in (9) is less than or equal to

$$1\{\tau_b^{\bullet}(\hat{\theta}_{b,t} - \theta)/\hat{\sigma}_{b,t} \le x + u\} + 1\{\tau_b^{\bullet}(\hat{\theta}_n - \theta)/\hat{\sigma}_{b,t} > u\};$$

then, sum over all t. We have seen that $(1 - R_n(u)) \to 0$ in probability and hence by a standard subsampling argument again we get, for any $\epsilon > 0$, $L^{\bullet}_{n,b}(x) \le J^{\bullet}(x + u, P) + \epsilon$ with probability tending to one, provided that $x + u$ is a continuity point of $J^{\bullet}(\cdot, P)$. Letting u tend to zero shows that $L^{\bullet}_{n,b}(x) \le J^{\bullet}(x, P) + \epsilon$ with probability tending to one. A similar argument leads to $L^{\bullet}_{n,b}(x) \ge J^{\bullet}(x, P) - \epsilon$ with probability tending to one. Since ϵ is arbitrary, this implies $L^{\bullet}_{n,b} \to J^{\bullet}(x, P)$ in probability, and thus we have proved (i).

The proofs of (ii) and (iii) given (i) are very similar to the proofs of (ii) and (iii) given (i) in Theorem 3.1 and thus are omitted. *Q.E.D.*

PROOF OF THEOREM 4.1: The subsampling theory presented in Section 3 assumes that the distributions of $(\hat{\alpha}_{b,t} - \alpha)$ and of $(\hat{\alpha}_{b,t} - \alpha)/\hat{\sigma}_{b,t}$ are independent of t. But this follows immediately by Appendix A.1 of Andrews and Chen (1994) given the stationarity of the innovation sequence $\{\epsilon_t\}$.

Therefore, we are left to show that the conditions of Theorem 3.2 are satisfied, no matter what the value of α. In the proof, we will distinguish the two cases $\alpha < 1$ and $\alpha = 1$.

Case $\alpha < 1$:

By the arguments of Appendix A.1 of Andrews and Chen (1994) again, in deriving the asymptotic distributions of $\hat{\alpha}_n$ and $(\hat{\alpha}_n - \alpha)/\hat{\sigma}_n$, we may assume without loss of generality that $\mu^* = \beta^* = 0$ in model (1). Hence, for the trend-stationary case $\alpha < 1$, we may assume that $Y_t = Y_t^*$ and so that Y_t is a stationary, mean-zero sequence of random variables.

Next, note that $\hat{\alpha}_n = \hat{\gamma}_{1,n} + \hat{\gamma}_{2,n} + \cdots + \hat{\gamma}_{p,n}$, where $\hat{\gamma}_{i,n}$ is the OLS estimator of γ_i in model (3). Denote $\underline{\gamma} = (\gamma_1, \ldots, \gamma_p)'$. As will be shown in Proposition A.1, the asymptotics of $\underline{\hat{\gamma}}_n$ and the OLS estimator of its covariance matrix are given by

$$n^{1/2}(\underline{\hat{\gamma}}_n - \underline{\gamma}) \overset{\mathscr{L}}{\Rightarrow} N(0, \Pi^{-1}\Omega\Pi^{-1}) \tag{10}$$

and

$$n\widehat{\mathrm{Cov}}(\underline{\hat{\gamma}}_n) \overset{\mathscr{P}}{\to} \sigma_\epsilon^2 \Pi^{-1}. \tag{11}$$

Here, Ω and Π are two symmetric $p \times p$ matrices of full rank, $\sigma_\epsilon^2 = E(\epsilon_t^2)$, $\overset{\mathscr{L}}{\Rightarrow}$ denotes convergence in distribution, and $\overset{\mathscr{P}}{\to}$ denotes convergence in probability. This implies, letting $\underline{1} = (1, \ldots, 1)'$,

$$n^{1/2}(\hat{\alpha}_n - \alpha) \overset{\mathscr{L}}{\Rightarrow} N(0, \underline{1}\Pi^{-1}\Omega\Pi^{-1}\underline{1}'), \tag{12}$$

$$\hat{\sigma}_n^2 \overset{\mathscr{P}}{\to} \sigma_\epsilon^2 \underline{1}\Pi^{-1}\underline{1}', \tag{13}$$

and

$$n^{1/2}(\hat{\alpha}_n - \alpha)/\hat{\sigma}_n \overset{\mathscr{L}}{\Rightarrow} N\left(0, \frac{\underline{1}\Pi^{-1}\Omega\Pi^{-1}\underline{1}'}{\sigma_\epsilon^2 \underline{1}\Pi^{-1}\underline{1}'}\right). \tag{14}$$

Moreover, as follows from Proposition A.1, in the special case of i.i.d innovations ϵ_t, we have $\Omega = \sigma_\epsilon^2 \Pi$ and the asymptotics simplify to

$$n^{1/2}(\hat{\alpha}_n - \alpha) \overset{\mathscr{L}}{\Rightarrow} N(0, \sigma_\epsilon^2 \underline{1}\Pi^{-1}\underline{1}'),$$

$$\hat{\sigma}_n^2 \overset{\mathscr{P}}{\to} \sigma_\epsilon^2 \underline{1}\Pi^{-1}\underline{1}',$$

and

$$n^{1/2}(\hat{\alpha}_n - \alpha)/\hat{\sigma}_n \overset{\mathscr{L}}{\Rightarrow} N(0,1).$$

It is, however, important to note that the t-statistic for $\hat{\alpha}_n$ does in general not have a limiting standard normal distribution. Indeed, the limiting variance can have any arbitrary value as determined by the the dependence structure of the innovations ϵ_t; see Romano and Thombs (1996) for some explicit examples. Therefore, the standard inference on α in the trend-stationary case (even if α is close to 0) can be very misleading. On the other hand, subsampling inference on α is not affected by uncorrelated rather than i.i.d. innovations, as will be demonstrated now.

To check the conditions of Theorem 3.2, first note that, due to (12)–(14), Assumption 3.2 holds with $\tau_n^\bullet = n^{1/2}$, $a_n = n^{1/2}$, and $d_n = 1$. Next, both $\hat{\alpha}_{b,t}$ and $\hat{\sigma}_{b,t}$ are functions of $(Y_t, \ldots, Y_{t+b-1})$ and thus $\alpha_{n,b}^\bullet(h) \le \min\{1, \alpha_{Y^*}(h-b)\}$; recall that for the proof we may assume $Y_t = Y_t^*$. This implies the mixing condition $n^{-1}\sum_{h=1}^n \alpha_{n,b}(h) \to 0$ as $n \to \infty$, since $\alpha_{Y^*}(h) \to 0$ as $h \to \infty$ as well as $b/n \to 0$ as $n \to \infty$.

Case $\alpha = 1$:

As shown in Stock (1991),

$$n(\hat{\alpha}_n - 1) \overset{\mathscr{L}}{\Rightarrow} b(1)\left(\int_0^1 W^t(s)^2\, ds\right)^{-1} \int_0^1 W^t(s)\, dW(s), \tag{15}$$

$$n^{1/2}\hat{\sigma}_n \overset{\mathscr{L}}{\Rightarrow} b(1)\left(\int_0^1 W^t(s)^2\, ds\right)^{-1/2}, \quad \text{and} \tag{16}$$

$$n^{1/2}(\hat{\alpha}_n - 1)/\hat{\sigma}_n \overset{\mathscr{L}}{\Rightarrow} \left(\int_0^1 W^t(s)^2\, ds\right)^{-1/2} \int_0^1 W^t(s)\, dW(s). \tag{17}$$

Here, $b(1) = 1 - \sum_{i=1}^{p-1}\psi_i$, $W(s)$ is a standard Brownian Motion, and

$$W^t(s) = W(s) - \int_0^1 (4 - 6r)W(r)\, dr - s\int_0^1 (12r - 6)W(r)\, dr$$

is a "detrended" Brownian Motion.

Again, we may assume without loss of generality that $Y_t = Y_t^*$ and hence that $\Delta Y_t = \Delta Y_t^*$.

To check the conditions of Theorem 3.2, first note that, due to (15)–(17), Assumption 3.2 holds with $\tau_n^\bullet = n^{1/2}$, $a_n = n$, and $d_n = n^{1/2}$. Next, $\hat{\alpha}_{b,t}$ and $\hat{\sigma}_{b,t}$ are functions of $(Y_t, \ldots, Y_{t+b-1})$ and therefore functions of $(Y_t, \Delta Y_{t+1}, \ldots, \Delta Y_{t+b-1})$. As noted in Appendix A.1 of Andrews and Chen (1994), the value of $\hat{\alpha}_{b,t}$ does not depend on the initial random variable Y_t but only on the differences $\Delta Y_{t+1}, \ldots, \Delta Y_{t+b-1}$; the same can be seen true for the value of $\hat{\sigma}_{b,t}$. This means that we are able to reconstruct the numerical values $\hat{\alpha}_{b,t}$ and $\hat{\sigma}_{b,t}$ from $(\Delta Y_{t+1}, \ldots, \Delta Y_{t+b-1})$ alone. Thus, $\alpha_{n,b}^\bullet(h) \le \min\{1, \alpha_{\Delta Y^*}(h-b)\}$. This implies the mixing condition $n^{-1}\sum_{h=1}^n \alpha_{n,b}^\bullet(h) \to 0$ as $n \to \infty$, since $\alpha_{\Delta Y^*}(h) \to 0$ as $h \to \infty$ as well as $b/n \to 0$ as $n \to \infty$. *Q.E.D.*

PROPOSITION A.1: *Assume the assumptions of Theorem* 4.1 *and that* $\alpha < 1$. *Let* Ω *and* Π *be two* $p \times p$ *matrices defined by their* (i,j)*th entries*

$$\omega_{i,j} = E(Y_p^* Y_{p-|i-j|}^* \epsilon_{p+1}^2)$$

and

$$\pi_{i,j} = E(Y_p^* Y_{p-|i-j|}^*).$$

Then, the convergences (10) *and* (11) *hold. Moreover, and immediately clear, if in addition the innovations* ϵ_t *are i.i.d., then* $\Omega = \sigma_\epsilon^2 \Pi$, *where* $\sigma_\epsilon^2 = E(\epsilon_t^2)$.

PROOF: As before, assume without loss of generality that $\mu^* = \beta^* = 0$ in model (1) and hence that $Y_t = Y_t^*$ is a stationary, mean-zero process. Define $\theta = (\gamma_1, \ldots, \gamma_p, \mu, \beta)'$ and

$$\Xi_n = \begin{pmatrix} n^{1/2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & n^{1/2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & n^{1/2} & 0 & 0 \\ 0 & 0 & \cdots & 0 & n^{1/2} & 0 \\ 0 & 0 & \cdots & 0 & 0 & n^{3/2} \end{pmatrix} \quad \text{and} \quad V_t = \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \\ 1 \\ t \end{pmatrix}.$$

Then, for the OLS estimator $\hat{\theta}_n$, we have

$$\hat{\theta}_n - \theta = \left[\sum_{t=p+1}^{n} V_t V_t' \right]^{-1} \sum_{t=p+1}^{n} V_t \epsilon_t. \tag{18}$$

Premultiplying (18) by Ξ_n yields

$$\Xi_n(\hat{\theta}_n - \theta) = \left[\Xi_n^{-1} \sum_{t=p+1}^{n} V_t V_t' \Xi_n^{-1} \right]^{-1} \left[\Xi_n^{-1} \sum_{t=p+1}^{n} V_t \epsilon_t \right]. \tag{19}$$

We claim that

$$\Xi_n^{-1} \sum_{t=p+1}^{n} V_t V_t' \Xi_n^{-1} \xrightarrow{\mathscr{P}} \begin{pmatrix} \Pi & 0 \\ 0 & \Lambda \end{pmatrix} \quad \text{and thus} \tag{20}$$

$$\left[\Xi_n^{-1} \sum_{t=p+1}^{n} V_t V_t' \Xi_n^{-1} \right]^{-1} \xrightarrow{\mathscr{P}} \begin{pmatrix} \Pi^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix},$$

where

$$\Lambda = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}.$$

To see why, note that

$$\Xi_n^{-1} \sum_{t=p+1}^{n} V_t V_t' \Xi_n^{-1} = \begin{pmatrix} n^{-1}\Sigma Y_{t-1}^2 & n^{-1}\Sigma Y_{t-1}Y_{t-2} & \cdots & n^{-1}\Sigma Y_{t-1}Y_{t-p} & n^{-1}\Sigma Y_{t-1} & n^{-2}\Sigma t Y_{t-1} \\ n^{-1}\Sigma Y_{t-2}Y_{t-1} & n^{-1}\Sigma Y_{t-2}^2 & \cdots & n^{-1}\Sigma Y_{t-2}Y_{t-p} & n^{-1}\Sigma Y_{t-2} & n^{-2}\Sigma t Y_{t-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ n^{-1}\Sigma Y_{t-p}Y_{t-1} & n^{-1}\Sigma Y_{t-p}Y_{t-2} & \cdots & n^{-1}\Sigma Y_{t-p}^2 & n^{-1}\Sigma Y_{t-p} & n^{-2}\Sigma t Y_{t-p} \\ n^{-1}\Sigma Y_{t-1} & n^{-1}\Sigma Y_{t-2} & \cdots & n^{-1}\Sigma Y_{t-p} & n^{-1}\cdot n & n^{-2}\Sigma t \\ n^{-2}\Sigma t Y_{t-1} & n^{-2}\Sigma t Y_{t-2} & \cdots & n^{-2}\Sigma t Y_{t-p} & n^{-2}\Sigma t & n^{-3}\Sigma t^2 \end{pmatrix}.$$

The convergence part pertaining to Λ is obvious. Next, to show the convergence part pertaining to Π, focus on element $\pi_{i,j}$. First, the sequence $Y_{t-i}Y_{t-j}$ has a bounded $2+\delta$th absolute moment (by Cauchy-Schwarz). Second, it is strong mixing with mixing coefficients $\alpha_{i,j}(h) \le \min\{1, \alpha_{Y^*}(h - |i - j|)\}$. Third, the expectation of the sample mean has limit $\pi_{i,j}$ as n tends to infinity (actually, it is constantly equal to $\pi_{i,j}$). The convergence of $n^{-1}\Sigma Y_{t-i}Y_{t-j}$ to $\pi_{i,j}$ now follows with Corollary 3.48 of White (1984), $(n-p)/n \to 1$, and Slutzky's Theorem. Next, the convergence of $n^{-1}\Sigma Y_{t-i}$ to 0 in probability is analogous. Finally, the convergence of $n^{-2}\Sigma t Y_{t-i}$ to 0 in probability is analogous as well by writing it as $n^{-1}\Sigma(t/n)Y_{t-i}$. Therefore, (20) is proved.

We claim now that

$$\operatorname{cov}\left(\Xi_n^{-1} \sum_{t=p+1}^{n} V_t \epsilon_t \right) \to \begin{pmatrix} \Omega & \Upsilon \\ \Upsilon' & \Gamma \end{pmatrix} \quad \text{as} \quad n \to \infty, \tag{21}$$

where

$$\Upsilon = \begin{pmatrix} E(Y_p \epsilon_{p+1}^2) & (1/2)E(Y_p \epsilon_{p+1}^2) \\ \vdots & \vdots \\ E(Y_1 \epsilon_{p+1}^2) & (1/2)E(Y_1 \epsilon_{p+1}^2) \end{pmatrix} \quad \text{and} \quad \Gamma = \sigma_\epsilon^2 \Lambda.$$

To see why, note that

$$\Xi_n^{-1} \sum_{t=p+1}^{n} V_t \epsilon_t = n^{-1/2} \sum_{t=p+1}^{n} \xi_t \quad \text{with} \quad \xi_t = \begin{pmatrix} Y_{t-1}\epsilon_t \\ Y_{t-2}\epsilon_t \\ \vdots \\ Y_{t-p}\epsilon_t \\ \epsilon_t \\ (t/n)\epsilon_t \end{pmatrix}.$$

Note that the sequence ξ_t is uncorrelated and strong mixing with mixing coefficients $\alpha_\xi(\cdot)$ satisfying $\alpha_\xi(h) \le \min\{1, \alpha_{Y^*}(h-p)\}$. Moreover, ξ_t has expectation zero and covariance matrix

$$\operatorname{cov}(\xi_t) = \begin{pmatrix} \omega_{1,1} & \cdots & \omega_{1,p} & E(Y_{t-1}\epsilon_t^2) & (t/n)E(Y_{t-1}\epsilon_t^2) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \omega_{p,1} & \cdots & \omega_{p,p} & E(Y_{t-p}\epsilon_t^2) & (t/n)E(Y_{t-p}\epsilon_t^2) \\ E(Y_{t-1}\epsilon_t^2) & \cdots & E(Y_{t-p}\epsilon_t^2) & E(\epsilon_t^2) & (t/n)E(\epsilon_t^2) \\ (t/n)E(Y_{t-1}\epsilon_t^2) & \cdots & (t/n)E(Y_{t-p}\epsilon_t^2) & (t/n)E(\epsilon_t^2) & (t/n)^2E(\epsilon_t^2) \end{pmatrix}.$$

Since the sequence ξ_t is uncorrelated, it therefore follows that

$$\operatorname{cov}\left(n^{-1/2} \sum_{t=p+1}^{n} \xi_t \right) = n^{-1} \sum_{t=p+1}^{n} \operatorname{cov}(\xi_t) \to \begin{pmatrix} \Omega & \Upsilon \\ \Upsilon' & \Gamma \end{pmatrix} \quad \text{as} \quad n \to \infty$$

and so (21) is proved.

The convergences (21) together with the mixing and moment conditions of Theorem 4.1 imply that

$$\Xi_n(\hat{\theta}_n - \theta) \overset{\mathscr{L}}{\Rightarrow} N\left[0, \begin{pmatrix} \Pi^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix} \begin{pmatrix} \Omega & \Upsilon \\ \Upsilon' & \Gamma \end{pmatrix} \begin{pmatrix} \Pi^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix}\right]. \tag{22}$$

Indeed, the convergence (22) follows from the proof of Theorem 3.4 of Politis, Romano, and Wolf (1997). Only a slight modification is needed, since the last element of $\hat{\theta}_n$ (namely $\hat{\beta}_n$) converges at rate $n^{3/2}$ rather than $n^{1/2}$, whereas in said theorem all regression coefficients converge at rate $n^{1/2}$. However, the necessary changes are minor and straightforward and so the details are omitted.

Recalling that $\theta = (\underline{\gamma}, \mu, \beta)'$ and that Ξ_n is a diagonal matrix with the first p diagonal entries equal to $n^{1/2}$, we immediately have that

$$n^{1/2}(\hat{\underline{\gamma}} - \underline{\gamma}) \overset{\mathscr{L}}{\Rightarrow} N(0, \Pi^{-1}\Omega\Pi^{-1}),$$

which demonstrates (10).

To show (11), note that

$$\widehat{n\operatorname{Cov}}(\hat{\underline{\gamma}}_n) = \hat{\sigma}_{\epsilon,n}^2 \left[\Xi_n^{-1} \sum_{t=p+1}^{n} V_t V_t' \Sigma_n^{-1} \right]^{-1}_{(1\ldots p, 1\ldots p)},$$

where $\hat{\sigma}_{\epsilon,n}^2$ is the OLS estimator of σ_ϵ^2 and $[\cdot]_{(1\ldots p, 1\ldots p)}$ is the $p \times p$ "upper-left" submatrix of $[\cdot]$. The consistency of $\hat{\sigma}_{\epsilon,n}^2$, which is standard, and (20) now imply (11). *Q.E.D.*

PROOF OF COROLLARY 4.1: The proof is analogous to the proof of Theorem 4.1 with the major difference being that Assumption 3.2 has to be rechecked. The results for the case $\alpha < 1$ are identical, given by (12)–(14), and are derived in a similar fashion as in the proof of Theorem 4.1. The

results for the case $\alpha = 1$ are given by (15)–(17) when the "detrended" Brownian Motion is replaced by a "demeaned" Brownian Motion; see Stock (1991). *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1993): "Exactly Median-unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165.

ANDREWS, D. W. K., AND H.-Y. CHEN (1994): "Approximately Median-unbiased Estimation of Autoregressive Models, *Journal of Business and Economic Statistics*, 12, 187–204.

BASAWA, I. V., A. K. MALLIK, W. P. MCCORMICK, J. H. REEVES, AND R. L. TAYLOR (1991): "Bootstrapping Unstable First-order Autoregressive Processes, *Annals of Statistics*, 19, 1008–1101.

BERAN, R. (1984): "Bootstrap Methods in Statistics," *Jahresberichte des Deutschen Mathematischen Vereins*, 86, 14–30.

DEJONG, D., AND C. H. WHITEMAN (1991a): "Reconsidering Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 221–254.

——— (1991b): "The Temporal Stability of Dividends and Stock Prices: Evidence from the Likelihood Function," *American Economic Review*, 81, 600–617.

DICICCIO, T. J., AND J. P. ROMANO (1988): "A Review of Bootstrap Confidence Intervals" (with Discussion), *Journal of the Royal Statistical Society*, *B*, 50, 338–370.

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

ELLIOT, G., AND J. H. STOCK (1994): "Inference in Time Series Regression when the Order of Integration of a Regressor Is Unknown," *Econometric Theory*, 10, 672–700.

HALL, P. (1988): "On Symmetric Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society*, *B*, 50, 35–45.

——— (1992): *The Bootstrap and Edgeworth Expansion*. New York: Springer.

HANSEN, B. E. (1999): "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics*, 81, 594–607.

KÜNSCH, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

LIU, R. Y., AND K. SINGH (1992): "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap*, ed. by R. LePage and L. Billard. New York: John Wiley, pp. 225–248.

LOH, W.-Y. (1987): "Calibrating Confidence Coefficients," *Journal of the American Statistical Association*, 82, 155–162.

POLITIS, D. N., AND J. P. ROMANO (1994a): "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050.

——— (1994b): "The Stationary Boostrap," *Journal of the American Statistical Association*, 89, 1303–1313.

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1997): "Subsampling for Heteroskedastic Time Series," *Journal of Econometrics*, 81, 281–317.

——— (1999): *Subsampling*. New York: Springer.

ROMANO, J. P., AND L. A. THOMAS (1996): "Inference for Autocorrelations under Weak Assumptions, *Journal of the American Statistical Association*, 91, 590–600.

ROSENBLATT, M. (1956): "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences*, 42, 43–47.

STOCK, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459.

WHITE, H. (1984): *Asymptotic Theory for Econometricians*. Orlando: Academic Press.