Bootstrap Standard Error Estimates for Linear Regression

Sílvia GONÇALVES and Halbert WHITE

Standard errors of parameter estimates are widely used in empirical work. The bootstrap often can provide a convenient means of estimating standard errors. The conditions under which bootstrap standard error estimates are theoretically justified have not received much attention, however. This article establishes conditions for the consistency of the moving blocks bootstrap estimators of the variance of the least squares estimator in linear dynamic models with dependent data. We discuss several applications of this result, in particular, the use of bootstrap standard error estimates for bootstrapping Studentized statistics. A simulation study shows that inference based on bootstrap standard error estimates may be considerably more accurate in small samples than inference based on closed-form asymptotic estimates.

KEY WORDS: Moving blocks bootstrap; Studentized statistic.

1. INTRODUCTION

The bootstrap is a general method for estimating the sampling distribution of a statistic. Under suitable conditions, the bootstrap distribution is asymptotically first-order equivalent to the asymptotic distribution of the statistic of interest. The consistency of the bootstrap distribution, however, does not guarantee the consistency of the variance of the bootstrap distribution (the "bootstrap variance") as an estimator of the asymptotic variance, because it is well known that convergence in distribution of a random sequence does not imply convergence of moments (see, e.g., Billingsley 1995, thm. 25.12). For the sample median and smooth functions of sample means, examples of the inconsistency of bootstrap variance estimators in the iid context have been given by Ghosh, Parr, Singh, and Babu (1984) and Shao (1992).

For time series observations, the moving blocks bootstrap (MBB) introduced by Künsch (1989) and Liu and Singh (1992a) has been shown to consistently estimate the variance of the sample mean under weak dependence and heterogeneity assumptions (see Gonçalves and White 2002). For more general statistics, conditions for the consistency of the bootstrap variance estimator do not appear to be available.

The main purpose of this article is to provide sufficient conditions for the consistency of MBB variance estimators when the statistic of interest is the least squares (LS) estimator in possibly misspecified linear regression models with dependent data. Our framework includes linear regression with iid observations as a special case. In related work, Liu and Singh (1992b) showed the consistency of the iid bootstrap variance estimator for regressions with fixed regressors and iid errors. Our results allow for stochastic regressors and autocorrelated errors. Although the consistency of the MBB distribution of the LS estimator is well established in the literature (see, e.g., Fitzenberger 1997; Politis, Romano, and Wolf 1997), the consistency of the bootstrap variance of the LS estimator has not received much attention. As we remarked earlier, the former does not necessarily imply the latter, so that currently available results do not justify bootstrapping the standard errors of the LS estimates using the MBB.

Our result is important in that many applied studies have used bootstrap standard error estimates as a measure of the precision of their parameter estimates (see, e.g., Efron 1979; Freedman and Peters 1984; Efron and Tibshirani 1986; Li and Maddala 1999). We also emphasize that this result plays an important role in justifying bootstrap applications based on Studentized statistics, for which asymptotic refinements of the bootstrap can be expected. The construction of Studentized statistics involves normalization by the standard error of the estimator. Our results formally justify using the bootstrap in computing such standard errors. This feature is especially convenient in cases when asymptotic closed-form solutions are not available or are too cumbersome to be calculated. In addition, we present simulation evidence that suggests that inference based on bootstrap estimates of standard errors may be considerably more accurate in small samples than inference based on asymptotic closed-form standard error estimates. For a multiple linear regression model with autocorrelated (and heteroscedastic) errors, we find that confidence intervals that rely on bootstrap standard errors tend to perform better than confidence intervals that rely on asymptotic closed-form variances. In particular, the coverage errors of symmetric MBB percentile-t confidence intervals based on bootstrap standard error estimates are substantially smaller than the coverage errors typically found for other (asymptotic theory-based and bootstrap-based) confidence intervals in this setting, especially under strong autocorrelation.

The remainder of the article is organized as follows. Section 2 presents the theoretical results. Section 3 compares the accuracy of the bootstrap estimator with that of closed-form estimators of the variance. Section 4 provides concluding remarks, and an Appendix gives all of the proofs.

2. LINEAR REGRESSION

In this section we prove the asymptotic validity of the MBB for variance estimation in the context of linear regressions when the data-generating process (DGP) is near–epoch-dependent (NED) on a mixing process (Billingsley 1968; McLeish 1975; Gallant and White 1988). NED processes allow for considerable dependence and heterogeneity. They include as a special case the more conventional mixing processes, which can

Sílvia Gonçalves is Assistant Professor, Département de Sciences Économiques, CIRANO and CIREQ, Université de Montréal, Québec H3C 3J7, Canada (E-mail: *silvia.goncalves@umontreal.ca*). Halbert White is Professor, Economics Department, University of California San Diego, La Jolla, CA 92093 (E-mail: *hwhite@weber.ucsd.edu*). Gonçalves acknowledges financial support from Praxis XXI and a Sloan Doctoral Dissertation fellowship. White's participation was supported by National Science Foundation grants SBR-9811562 and SES-0111238. The authors thank Lutz Kilian for many helpful discussions, which led to a significant improvement in a previous version of this article, and to two anonymous referees for their valuable comments.

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Theory and Methods
DOI 10.1198/016214504000002087

be overly restrictive for applications in economics [see, e.g., Andrews 1984 for an example of a simple AR(1) process that fails to be strong mixing]. NED processes cover a variety of nonlinear time series models, including the bilinear, generalized autoregressive conditional heteroscedastic, and threshold autoregressive models (see Davidson 2002).

We define $\{\mathbf{Z}_t\}$ to be L_q-NED on a mixing process $\{\mathbf{V}_t\}$ provided that $E(\mathbf{Z}_t^q) < \infty$ and $v_k \equiv \sup_t \|\mathbf{Z}_t - E_{t-k}^{t+k}(\mathbf{Z}_t)\|_q$ tends to 0 as $k \to \infty$ at an appropriate rate, where $q \geq 2$. In particular, if $v_k = O(k^{-a-\delta})$ for some $\delta > 0$, then we say that $\{\mathbf{Z}_t\}$ is L_q-NED (on $\{\mathbf{V}_t\}$) of size $-a$. Here and in what follows, $\|\mathbf{Z}_t\|_q \equiv (E|\mathbf{Z}_t|^q)^{1/q}$ denotes the L_q norm of the random vector $\mathbf{Z}_t$, with $|\mathbf{Z}_t|$ its Euclidean norm, and $E_{t-k}^{t+k}(\cdot) \equiv E(\cdot|\mathcal{F}_{t-k}^{t+k})$, where $\mathcal{F}_{t-k}^{t+k} \equiv \sigma(\mathbf{V}_{t-k}, \ldots, \mathbf{V}_{t+k})$ is the σ-field generated by $\mathbf{V}_{t-k}, \ldots, \mathbf{V}_{t+k}$. The sequence $\{\mathbf{V}_t\}$ is assumed to be strong mixing, that is, $\alpha_k \equiv \sup_m \sup_{\{A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+k}^{\infty}\}} |P(A \cap B) - P(A)P(B)| \to 0$ as $k \to \infty$ at an appropriate rate.

Gallant and White (1988) studied the asymptotic properties of quasi–maximum likelihood estimators (QMLEs) for heterogeneous NED data and nonlinear dynamic models. Recently, Gonçalves and White (2004) established the first-order asymptotic validity of the MBB for the framework of Gallant and White (1988). In particular, Gonçalves and White (2004) showed that the MBB consistently estimates the asymptotic distribution of the QMLE. But as Gonçalves and White (2004) remarked, their results do not justify using the variance of the bootstrap distribution to consistently estimate the asymptotic variance of the QMLE. Here we fill this gap for the special case of the LS estimator for linear dynamic models. In particular, we give explicit conditions that justify bootstrapping the variance of the LS estimator in possibly misspecified linear dynamic models when the DGP is NED on a mixing process.

Assumption 1 is a version of the Gallant and White (1988) and Gonçalves and White (2004) assumptions specialized to the case of linear dynamic models.

Assumption 1. a. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. The observed data are a realization of a strictly stationary stochastic process $\{\mathbf{Z}_t = (Y_t, \mathbf{X}_t')' : \Omega \to \mathbb{R}^{p+1}, t = 1, 2, \ldots\}$, $p \in \mathbb{N}$; $\mathbf{Z}_t(\omega) = \mathbf{W}_t(\ldots, \mathbf{V}_{t-1}(\omega), \mathbf{V}_t(\omega), \mathbf{V}_{t+1}(\omega), \ldots)$, $\omega \in \Omega$, where $\mathbf{V}_t : \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, and $\mathbf{W}_t : \times_{\tau=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^{p+1}$ are such that $\mathbf{Z}_t$ is measurable, $t = 1, 2, \ldots$.

b. $Y_t = \mathbf{X}_t'\boldsymbol{\beta}^o + \varepsilon_t$, $t = 1, 2, \ldots$, for some $\boldsymbol{\beta}^o \in \mathbb{R}^p$, where $\mathbf{X}_t' = (\mathbf{X}_{t1}, \ldots, \mathbf{X}_{tp})$ and $E(\mathbf{X}_t \varepsilon_t) = 0$.

c. For some $r > 2$, $\|Y_t\|_{6r} \leq \Delta < \infty$, $\|X_{ti}\|_{6r} \leq \Delta < \infty$, for $i = 1, \ldots, p$, $t = 1, 2, \ldots$.

d. For some small $\delta > 0$, the elements of $\{\mathbf{Z}_t\}$ are $L_{2+\delta}$-NED on $\{\mathbf{V}_t\}$ with NED coefficients v_k of size $-\frac{4(r-1)^2}{(r-2)^2}$; $\{\mathbf{V}_t\}$ is an α-mixing sequence with α_k of size $-\frac{(2+\delta)r}{r-2}$.

e. $\mathbf{A}^o \equiv E(\mathbf{X}_t\mathbf{X}_t')$ is nonsingular, that is, $\lambda_{\min}(\mathbf{A}^o) \geq \eta > 0$ for some $\eta > 0$, where $\lambda_{\min}(\mathbf{A}^o)$ denotes the smallest eigenvalue of $\mathbf{A}^o$.

f. $\mathbf{B}^o \equiv \lim_{n\to\infty} \mathbf{B}_n^o$ is positive definite, where $\mathbf{B}_n^o = \operatorname{var}(n^{-1/2}\sum_{t=1}^n \mathbf{X}_t\varepsilon_t)$.

According to Assumption 1.a, we observe data on $(p+1) \times 1$ random vectors $\mathbf{Z}_t = (Y_t, \mathbf{X}_t')'$, each of which is viewed as a transformation of some underlying process $\{\mathbf{V}_t\}$. Here Y_t denotes the observation t on the dependent variable and $\mathbf{X}_t \equiv (X_{t1}, \ldots, X_{tp})'$ is the $p \times 1$ vector of regressors for observation t; $\mathbf{X}_t$ may include lagged dependent variables. For simplicity, we assume that the DGP for $\mathbf{Z}_t$ is strictly stationary. Without stationarity, results analogous to ours can still be derived under additional conditions controlling the degree of heterogeneity in the data. Assumption 1.b specifies a linear dynamic model that may be misspecified in the sense that for all $\boldsymbol{\beta} \in \mathbb{R}^p$, it is true that $P(E(Y_t|\mathbf{X}_t) \neq \mathbf{X}_t'\boldsymbol{\beta}) > 0$. Such models are relevant for forecasting, because in this misspecified context $\boldsymbol{\beta}^o$ is the parameter that minimizes the mean squared error of the linear approximation to the unknown $E(Y_t|\mathbf{X}_t)$. In particular, under Assumption 1.e, $\boldsymbol{\beta}^o$ is uniquely defined by $\boldsymbol{\beta}^o = (E(\mathbf{X}'\mathbf{X}))^{-1}E(\mathbf{X}'\mathbf{Y})$, where we let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)'$.

We estimate $\boldsymbol{\beta}^o$ by the LS estimator $\hat{\boldsymbol{\beta}}_n = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Under our assumptions, $\hat{\boldsymbol{\beta}}_n$ consistently estimates $\boldsymbol{\beta}^o$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^o) \Rightarrow \mathrm{N}(\mathbf{0}, \mathbf{A}^{o-1}\mathbf{B}^o\mathbf{A}^{o-1})$; that is, the limiting distribution of the LS estimator $\hat{\boldsymbol{\beta}}_n$ is the multivariate normal distribution with asymptotic variance–covariance matrix $\mathbf{C}^o \equiv \mathbf{A}^{o-1}\mathbf{B}^o\mathbf{A}^{o-1}$. The bootstrap can be used to estimate the distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^o)$ and to estimate $\mathbf{C}^o$.

Let $\hat{\boldsymbol{\beta}}_n^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^*$ be the LS estimator of $\boldsymbol{\beta}^o$ based on the bootstrap data $\{\mathbf{Z}_{nt}^* = (Y_{nt}^*, \mathbf{X}_{nt}^{*\prime})'\}$ obtained with the MBB as follows. Let $\ell = \ell_n \in \mathbb{N}$ $(1 \leq \ell < n)$ denote the length of the blocks and let $\mathbf{B}_{t,\ell} = \{\mathbf{Z}_t, \mathbf{Z}_{t+1}, \ldots, \mathbf{Z}_{t+\ell-1}\}$ be the block of ℓ consecutive observations starting at $\mathbf{Z}_t$; $\ell = 1$ corresponds to the standard iid bootstrap. The MBB resamples $k = n/\ell$ blocks randomly with replacement from the set of $n - \ell + 1$ overlapping blocks $\{\mathbf{B}_{1,\ell}, \ldots, \mathbf{B}_{n-\ell+1,\ell}\}$, where for simplicity we assume that $n = k\ell$.

One bootstrap variance–covariance matrix estimator of $\hat{\boldsymbol{\beta}}_n$ is given by the bootstrap population variance–covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$, conditional on the original data, $\hat{\mathbf{C}}_n^* = \operatorname{var}^*(\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n))$. Because in general there is no closed-form expression for $\hat{\mathbf{C}}_n^*$, we compute an approximation to $\hat{\mathbf{C}}_n^*$ by Monte Carlo simulation, that is, $\hat{\mathbf{C}}_n^* = \lim_{B\to\infty} \frac{n}{B}\sum_{i=1}^B (\hat{\boldsymbol{\beta}}_n^{*(i)} - \bar{\hat{\boldsymbol{\beta}}}_n^*)(\hat{\boldsymbol{\beta}}_n^{*(i)} - \bar{\hat{\boldsymbol{\beta}}}_n^*)'$, where $\bar{\hat{\boldsymbol{\beta}}}_n^* = \frac{1}{B}\sum_{i=1}^B \hat{\boldsymbol{\beta}}_n^{*(i)}$, with $\hat{\boldsymbol{\beta}}_n^{*(i)}$ the bootstrap LS estimator evaluated on the ith bootstrap replication and B the total number of bootstrap replications.

In this article we focus on an alternative bootstrap variance–covariance matrix estimator of $\hat{\boldsymbol{\beta}}_n$, namely the bootstrap population variance–covariance matrix of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$. Following Liu and Singh (1992b) and Shao and Tu (1995, chap. 7, sec. 7.2.2), we define $\tilde{\boldsymbol{\beta}}_n^*$ as

$$\tilde{\boldsymbol{\beta}}_n^* = \begin{cases} (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^* & \text{if } \lambda_{\min}\left(\dfrac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right) \geq \dfrac{\delta}{2} \\ \hat{\boldsymbol{\beta}}_n & \text{otherwise} \end{cases} \tag{1}$$

for some $\delta > 0$, where $\lambda_{\min}(\mathbf{A})$ denotes the smallest eigenvalue of $\mathbf{A}$ for any matrix $\mathbf{A}$. Given the foregoing definition, $\tilde{\boldsymbol{\beta}}_n^*$ is equal to $\hat{\boldsymbol{\beta}}_n^*$ whenever $\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}$ is nonsingular. Because for any $\varepsilon > 0$ and sufficiently large n, there exists $\delta > 0$ such that

$$P\left[P^*\left(\lambda_{\min}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right) \geq \frac{\delta}{2}\right) > 1 - \varepsilon\right] > 1 - \varepsilon, \tag{2}$$

this modification affords considerable convenience without adverse practical consequences by greatly simplifying the theoretical study of the bootstrap variance estimator, $\tilde{\mathbf{C}}_n^* \equiv \text{var}^*(\sqrt{n}\tilde{\boldsymbol{\beta}}_n^*)$.

An important intermediate step toward proving the consistency of $\tilde{\mathbf{C}}_n^*$ for $\mathbf{C}^o$ is to establish the first-order asymptotic validity of the bootstrap distribution of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$. This follows by an application of theorem 2.2 of Gonçalves and White (2004) for the special case of the LS estimator of (possibly) misspecified linear dynamic models and the fact that condition (2) holds under Assumption 1. Assumption 1 also ensures that the elements of $\{\mathbf{X}_t\varepsilon_t\}$ satisfy assumptions 2.1 and 2.2 of Gonçalves and White (2004). In particular, assumption 2.2 of Gonçalves and White (2004) is automatically satisfied because $E(\mathbf{X}_t\varepsilon_t) = \mathbf{0}$ for all t given the stationarity assumption and the definition of $\boldsymbol{\beta}^o$. Alternatively, for assumption 2.1 we could have assumed that $\|X_{ti}\varepsilon_t\|_{3r} \le \Delta < \infty$ for some $r > 2$, and that for some $\delta > 0$, the elements of $\{\mathbf{X}_t\varepsilon_t\}$ are $L_{2+\delta}$-NED on $\{\mathbf{V}_t\}$ with NED coefficients v_k of size $-\frac{2(r-1)}{r-2}$.

Our main result establishes the consistency of $\tilde{\mathbf{C}}_n^*$ for $\mathbf{C}^o \equiv \mathbf{A}^{o-1}\mathbf{B}^o\mathbf{A}^{o-1}$ under Assumption 1.

Theorem 1. Under Assumption 1, if $\ell_n = o(n^{1/2})$ and $\ell_n \to \infty$, then $\tilde{\mathbf{C}}_n^* \xrightarrow{P} \mathbf{C}^o$, where $\tilde{\mathbf{C}}_n^* \equiv \text{var}^*(\sqrt{n}\tilde{\boldsymbol{\beta}}_n^*)$ and $\mathbf{C}^o \equiv \mathbf{A}^{o-1}\mathbf{B}^o\mathbf{A}^{o-1}$.

Theorem 1 justifies the use of $\tilde{\mathbf{C}}_n^*$ as an heteroscedasticity- and autocorrelation-consistent (HAC) variance estimator of $\mathbf{C}^o$. Given the first-order asymptotic validity of the bootstrap distribution of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$, we show that $E^*(|\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)|^{2+\delta}) = O_P(1)$, which is a sufficient condition for the uniform integrability of the sequence $\{|\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)|^2\}$.

Although we have focused on the LS estimator for linear regression models, several extensions of our results are possible. First, we can generalize our results to the k-step QMLEs as proposed by Davidson and MacKinnon (1999). To illustrate, consider the one-step QMLE as defined by Gonçalves and White (2004),

$$\hat{\theta}_{1n}^* = \hat{\theta}_n - A_n^*(\hat{\theta}_n)^{-1} n^{-1} \sum_{t=1}^{n} s_{nt}^*(\hat{\theta}_n).$$

We use the same notation as Gonçalves and White (2004). In particular, $\hat{\theta}_n$ is the QMLE of a pseudoparameter θ_n^o, $A_n^*(\hat{\theta}_n) = n^{-1}\sum_{t=1}^n \nabla^2 \log f_{nt}^*(\hat{\theta}_n)$ is the MBB resampled estimated Hessian, and $\{s_{nt}^*(\hat{\theta}_n)\}$ are the MBB resampled estimated scores. Under the assumptions of Gonçalves and White (2004), the bootstrap distribution of $\sqrt{n}(\hat{\theta}_{1n}^* - \hat{\theta}_n)$ is first-order asymptotic equivalent to the asymptotic normal distribution of the QMLE $\sqrt{n}(\hat{\theta}_n - \theta_n^o)$ (cf. their thm. 2.2 and cor. 2.1). Therefore, it suffices to show that $E^*(|\sqrt{n}(\hat{\theta}_{1n}^* - \hat{\theta}_n)|^{2+\delta}) = O_P(1)$. As with the LS estimator, here it is convenient to consider a truncated version of the one-step bootstrap estimator, namely $\widetilde{\theta}_{1n}^* = \hat{\theta}_{1n}^*$ when $A_n^*(\hat{\theta}_n)^{-1}$ exists and $\widetilde{\theta}_{1n}^* = \hat{\theta}_n$ otherwise. To prove that $E^*(|\sqrt{n}(\widetilde{\theta}_{1n}^* - \hat{\theta}_n)|^{2+\delta}) = O_P(1)$, we can use reasoning similar to that underlying the proof of our Theorem 1. In particular, it suffices to show that $E^*|n^{-1/2}\sum_{t=1}^n s_{nt}^*(\hat{\theta}_n)|^{2+\delta} = O_P(1)$. To maintain our focus on the case of linear regression, we do not provide further details here, but we will take up formal statements of k-step QMLE results elsewhere.

Another useful extension of the results presented here is to quantile regression. Because the asymptotic variance–covariance matrix of the quantile regression estimator depends on the density of the error term, bootstrapping the standard error estimate is particularly convenient, because it avoids nonparametric density estimation. For cross-sectional quantile regression, Buchinsky (1995) investigated the finite-sample performance of several bootstrap standard error estimates, including a pairwise bootstrap standard error estimate. Nevertheless, no formal justification for these bootstrap applications was provided. Also in the cross-sectional context, Hahn (1995) proved the first-order asymptotic validity of the bootstrap distribution of the quantile regression estimator. As Hahn (1995, p. 107) remarked, his results provide a theoretical justification for bootstrap percentile confidence intervals, but they do not justify using the bootstrap to estimate standard errors. Similarly, although Fitzenberger (1997) proved that the MBB consistently estimates the asymptotic distribution of the quantile regression estimator, his results do not apply to bootstrap standard error estimates. Thus, establishing theoretical results that justify the application of the bootstrap to variance estimation for the quantile regression estimator is an important area of future research. In his study, Fitzenberger (1997) treated the quantile regression estimator in a setting analogous to the LS case. Therefore, we conjecture that verification of the uniform integrability condition for the quantile regression estimator could be pursued along the same lines as for the LS estimator in Theorem 1. As for the k-step QMLE results, we take up formal treatment of quantile regression elsewhere.

3. MONTE CARLO RESULTS

In this section we conduct a Monte Carlo experiment that highlights the potential gains in accuracy from using bootstrap standard error estimates in the context of a multiple linear regression with serially dependent and heteroscedastic errors. Important examples of linear regression models in the applied econometrics literature are long-horizon regressions. Such regression models have been applied in, for example, the context of testing the predictability of exchange returns or, more generally, asset returns based on economic fundamentals (see Mark 1995; Hodrick 1992; Kirby 1998; Kilian 1999).

We consider the problem of building a confidence interval for a single regression parameter. We use the finite-sample coverage probability of symmetric 95% confidence intervals as our performance criterion. Our study is analogous to the simulation studies of Fitzenberger (1997) and Politis et al. (1997), following the basic setup of Andrews (1991) (see also Romano and Wolf 2003 for a similar design).

In particular, we consider the linear regression model $Y_t = \mathbf{X}_t'\boldsymbol{\beta}^o + \varepsilon_t$, where $\mathbf{X}_t' = (X_{t1}, \mathbf{X}_{t2}')$ contains five regressors, the first of which is a constant ($X_{t1} \equiv 1$). We consider two of the DGPs proposed by Andrews (1991), namely AR(1)–HOMO and AR(1)–HET2. The regressors and errors of the DGP AR(1)–HOMO are generated as mutually independent AR(1) models with variance 1 and AR parameter ρ,

$$X_{ti} = \rho X_{t-1,i} + \sqrt{1-\rho^2}\, v_{ti}, \qquad i = 2, \ldots, 5,$$

and

$$\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \sqrt{1-\rho^2}\, u_t,$$

where $\varepsilon_t = \tilde{\varepsilon}_t$. The innovations v_{ti} and u_t are generated as independent standard normal distributions. We set the true parameter $\boldsymbol{\beta}^o$ equal to $\mathbf{0}$ (without loss of generality) and consider the following values for the AR parameter ρ: .3, .5, .9, and .95. The DGP AR(1)–HET2 is obtained from the AR(1)–HOMO model by introducing conditional heteroscedastiticy in the errors ε_t. In particular, we let $\varepsilon_t = |\mathbf{X}'_{t2}\boldsymbol{\gamma}|\tilde{\varepsilon}_t$ with $\boldsymbol{\gamma} = (.5, .5, .5, .5)$. In the simulations, 5,000 random samples are generated for the sample sizes $n \in \{64, 128, 256, 512, 1{,}024\}$. The bootstrap intervals are based on 999 replications for each sample.

The goal is to build a confidence interval for β_2^o. The asymptotic normal theory-based confidence interval for β_2^o is given by $\hat{\beta}_{2n} \pm n^{-1/2}1.96\sqrt{\hat{C}_{n,22}}$, where $\hat{C}_{n,22}$ is the element $(2, 2)$ of $\hat{\mathbf{C}}_n$, a consistent estimator of the asymptotic variance–covariance matrix $\mathbf{C}^o = \mathbf{A}^{o-1}\mathbf{B}^o\mathbf{A}^{o-1}$. We consider two different choices of $\hat{\mathbf{C}}_n$. Our first choice exploits the sandwich form of $\mathbf{C}^o$ and is given by $\hat{C}_{n,\mathrm{QS}} = \hat{\mathbf{A}}_n^{-1}\hat{\mathbf{B}}_{n,\mathrm{QS}}\hat{\mathbf{A}}_n^{-1}$, where $\hat{\mathbf{A}}_n = \frac{\mathbf{X}'\mathbf{X}}{n}$ and $\hat{\mathbf{B}}_{n,\mathrm{QS}}$ is the quadratic spectral (QS) kernel variance estimator of Andrews (1991). This yields the following asymptotic normal theory-based confidence interval for β_2^o:

$$CI_{\mathrm{QS}} = \hat{\beta}_{2n} \pm n^{-1/2}1.96\sqrt{\hat{C}_{n,\mathrm{QS},22}}.$$

A second choice of $\hat{\mathbf{C}}_n$ is $\tilde{\mathbf{C}}_n^* = \mathrm{var}^*(\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n))$, the bootstrap covariance matrix of the distribution of $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$. Our Theorem 1 provides a formal justification for this choice. Here $\tilde{\boldsymbol{\beta}}_n^*$ is the truncated version of the LS estimator $\hat{\boldsymbol{\beta}}_n^*$, which replaces $\hat{\boldsymbol{\beta}}_n^*$ with $\hat{\boldsymbol{\beta}}_n$ whenever $(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$ does not exist. As it turned out, for our Monte Carlo design we never encountered any singularity problems. Thus in our simulations, $\tilde{\boldsymbol{\beta}}_n^* = \hat{\boldsymbol{\beta}}_n^*$, and $\tilde{\mathbf{C}}_n^*$ coincides with $\hat{\mathbf{C}}_n^* = \mathrm{var}^*(\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n))$. Notice that $\tilde{\mathbf{C}}_n^*$ does not rely on the sandwich form of $\mathbf{C}_n^o$ and is typically evaluated with Monte Carlo simulation methods. In particular, the bootstrap variance estimator based on B bootstrap replications is

$$\tilde{C}_{n,22,B}^* = \frac{n}{B}\sum_{i=1}^{B}\left(\tilde{\beta}_{2n}^{*(i)} - \overline{\tilde{\beta}_{2n}^*}\right)^2,$$

where $\tilde{\beta}_{2n}^{*(i)}$ denotes the (truncated) LS estimator of β_2^o evaluated on the ith bootstrap replication and $\overline{\tilde{\beta}_{2n}^*} = \frac{1}{B}\sum_{i=1}^{B}\tilde{\beta}_{2n}^{*(i)}$. When $B \to \infty$, $\tilde{C}_{n,22,B}^*$ approximates $\tilde{C}_{n,22}^*$, the "ideal" bootstrap variance estimator based on $B = \infty$. Here we let $B = 999$. A bootstrap variance, asymptotic normal theory-based confidence interval for β_2^o can be obtained as

$$CI_{\mathrm{var}^*} = \hat{\beta}_{2n} \pm n^{-1/2}1.96\sqrt{\tilde{C}_{n,22,B}^*},$$

where the critical value of the t-statistic is still obtained with the asymptotic normal distribution.

We also consider bootstrap percentile-t confidence intervals, for which asymptotic refinements can be expected. A 95% level symmetric bootstrap percentile-t confidence interval for β_2^o takes the form

$$CI_{\mathrm{per}\text{-}t} = \hat{\beta}_{2n} \pm q_{\mathrm{stud},.95}^*\sqrt{\hat{C}_{n,22}}, \qquad (3)$$

where $q_{\mathrm{stud},.95}^*$ is the 95% bootstrap percentile of the absolute value of the Studentized bootstrap statistic

$$t_{\tilde{\beta}_{2n}^*} = \frac{\sqrt{n}(\tilde{\beta}_{2n}^* - \hat{\beta}_{2n})}{\sqrt{\overline{C}_{n,22}^*}}. \qquad (4)$$

Here $\overline{C}_{n,22}^*$ is a consistent estimator of the bootstrap population variance of $\sqrt{n}\tilde{\beta}_{2n}^*$. (Note the use of $\overline{C}_{n,22}^*$ rather than $\tilde{C}_{n,22}^*$, for reasons elaborated on later.) A bootstrap percentile-t confidence interval requires the choice of two standard error estimates, one for studentizing the t-statistic evaluated on the real data [cf. $\sqrt{\hat{C}_{n,22}}$ in (3)] and the other for studentizing the t-statistic evaluated on the bootstrap data [cf. $\sqrt{\overline{C}_{n,22}^*}$ in (4)].

As discussed by Davison and Hall (1993) and Götze and Künsch (1996), for the MBB with dependent data, a careful choice of these standard error estimates is crucial if asymptotic refinements are to be expected. In particular, for smooth functions of means of stationary mixing data, to studentize the bootstrap statistic, Götze and Künsch (1996) suggested a variance estimator that exploits the independence of the bootstrap blocks and that can be interpreted as the sample variance of the bootstrap block means. To studentize the original statistic, Götze and Künsch (1996) used a kernel variance estimator with rectangular weights and warned that triangular weights would destroy second-order properties of the block bootstrap.

In our Monte Carlo simulations, to studentize the original t-statistic, we consider the same two choices as before, namely $\hat{C}_{n,\mathrm{QS},22}$, which relies on the sandwich form of C^o and uses the QS-kernel to estimate $\mathbf{B}^o$, and $\tilde{C}_{n,22,B}^*$, which estimates the standard error of $\hat{\beta}_{2n}$ with the bootstrap. To studentize the bootstrap t-statistic, we use the multivariate analog of the Götze and Künsch (1996) variance estimator, adapted to the LS context. In particular, we let $\overline{C}_{n,22}^*$ be the element $(2, 2)$ of $\overline{\mathbf{C}}_n^* = \tilde{\mathbf{A}}_n^{*-1}\tilde{\mathbf{B}}_n^*\tilde{\mathbf{A}}_n^{*-1}$, where $\tilde{\mathbf{A}}_n^* = \frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}$ and

$$\tilde{\mathbf{B}}_n^* = k^{-1}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell}\mathbf{X}_{I_i+t}\left(Y_{I_i+t} - \mathbf{X}'_{I_i}\tilde{\boldsymbol{\beta}}_n^*\right)\right) \times \left(\ell^{-1/2}\sum_{t=1}^{\ell}\mathbf{X}'_{I_i+t}\left(Y_{I_i+t} - \mathbf{X}'_{I_i}\tilde{\boldsymbol{\beta}}_n^*\right)\right),$$

where $\{I_i\}$ are iid random variables uniformly distributed on $\{0, 1, \ldots, n - \ell\}$. Another possibility would be to use the bootstrap to estimate the bootstrap variance of $\sqrt{n}\tilde{\beta}_{2n}^*$. This would correspond to a double bootstrap, where the bootstrap is used to simulate the distribution of the Studentized estimator, which is based on a standard error estimate that in turn has been estimated by the bootstrap. Implementing the double bootstrap would be extremely computationally intensive, and therefore we do not consider this alternative here. Nevertheless, we note that our theoretical results formally justify such an approach.

To summarize, we consider the following two 95% level symmetric bootstrap percentile-t confidence intervals:

$$CI_{\mathrm{per}\text{-}t,\mathrm{QS}} = \hat{\beta}_{2n} \pm q_{\mathrm{stud},.95}^*\sqrt{\hat{C}_{n,\mathrm{QS},22}}$$

and

$$CI_{\mathrm{per}\text{-}t,\mathrm{var}^*} = \hat{\beta}_{2n} \pm q_{\mathrm{stud},.95}^*\sqrt{\tilde{C}_{n,22,B}^*}.$$

For comparison purposes, we also include the 95% bootstrap percentile confidence interval given by

$$CI_{\text{per}} = \hat{\beta}_{2n} \pm q^*_{.95},$$

where $q^*_{.95}$ is the 95% bootstrap percentile of the absolute value of $\sqrt{n}(\tilde{\beta}^*_{2n} - \hat{\beta}_{2n})$. In contrast to the bootstrap percentile-t confidence interval, the bootstrap percentile confidence interval does not require any standard error estimate. However, because it is not based on an asymptotically pivotal statistic, this bootstrap method does not promise any asymptotic refinements.

The choices of the bandwidth for the QS-based confidence interval and of the block size for the MBB confidence intervals are critical. We use Andrews's (1991) automatic procedure based on approximating AR(1) models for the elements of $\mathbf{X}_t\hat{\varepsilon}_t$ to compute a data-driven bandwidth for the QS kernel. Given the asymptotic equivalence between the MBB and the Bartlett-kernel variance estimators, we choose the block length as the integer part of the data-driven bandwidth chosen by Andrews's automatic procedure for the Bartlett kernel. This choice is easy to implement and affords meaningful comparison of our results.

Figures 1 and 2 present results for the DGP AR(1)–HOMO, and Figures 3 and 4 present results for the DGP AR(1)–HET2. Each figure contains two panels, corresponding to two different values of ρ. Each panel depicts the actual coverage rate of each confidence interval as a function of the sample size.

All methods tend to undercover; the larger the ρ, the worse the undercoverage. One exception is $CI_{\text{per-}t,\text{var}^*}$, which shows a slight tendency to overcover for small n. The QS kernel-based confidence interval shows the worst performance among all methods. The bootstrap variance-based confidence interval CI_{var^*} shows improved coverage rates when compared with CI_{QS}, especially for small n and large ρ. This improvement may be quite substantial. For instance, for DGP AR(1)–HOMO, when $n = 64$ and $\rho = .9$, the coverage rate of CI_{QS} is 67.34%, whereas that of CI_{var^*} is 79.06%. Because both confidence intervals rely on the asymptotic normal approximation, using the bootstrap does not eliminate the undercoverage. However, these results suggest that replacing the asymptotic closed-form standard error estimates by bootstrap standard error estimates may by itself significantly improve the finite-sample performance of asymptotic normal theory-based confidence intervals. The finite-sample performance of CI_{var^*} is similar to that of CI_{per}.

As expected from the bootstrap theory, bootstrap percentile-t confidence intervals have smaller coverage distortions compared with the percentile confidence interval and the asymptotic normal theory-based confidence intervals. For AR(1)–HOMO, when the degree of autocorrelation is weak (i.e., for $\rho = .3$ and $\rho = .5$), $CI_{\text{per-}t,\text{var}^*}$ tends to overcover for the smaller sample sizes, whereas $CI_{\text{per-}t,\text{QS}}$ always undercovers. Both methods tend to be within one percentage point of the desired 95% level. When the degree of autocorrelation is strong (i.e., $\rho = .9$ and $\rho = .95$), the undercoverage of $CI_{\text{per-}t,\text{QS}}$ worsens. In contrast, $CI_{\text{per-}t,\text{var}^*}$ shows coverage rates that are closer to the nominal 95% level, with slight overcoverages for $n = 64$ and $n = 128$ and slight undercoverages for the larger sample sizes. Thus our results show that the choice of the standard error estimate used to studentize the t-statistic evaluated on the original

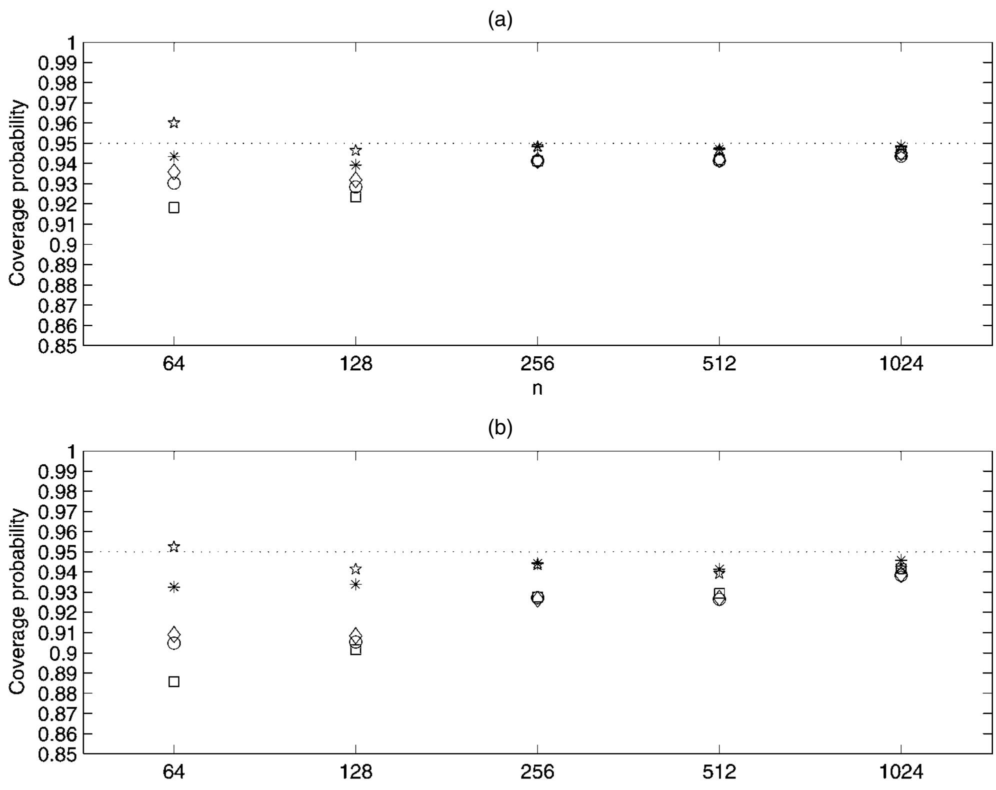

Figure 1. Coverage Probabilities of 95% Nominal Symmetric Confidence Intervals. Regression errors are homoscedastic AR(1) with autoregression coefficient equal to ρ. For (a), $\rho = .30$; for (b), $\rho = .50$ (□ CI_{QS}; ◇ CI_{var^*}; ○ CI_{per}; ∗ $CI_{per\text{-}t,QS}$; ☆ $CI_{per\text{-}t,var^*}$).

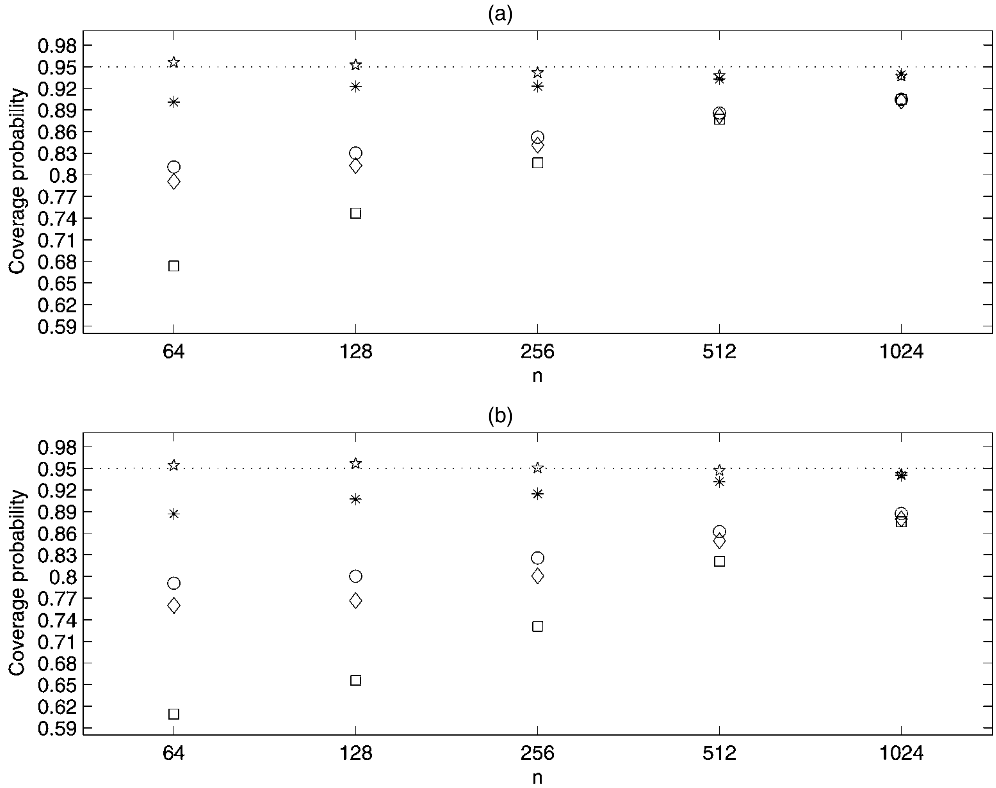

Figure 2. Coverage Probabilities of 95% Nominal Symmetric Confidence Intervals. Regression errors are homoscedastic AR(1) with autoregression coefficient equal to ρ. For (a), $\rho = .90$; for (b), $\rho = .95$ ($\sqsubset$ CI_{QS}; $\diamond$ CI_{var^*}; $\circ$ CI_{per}; $*$ $CI_{per\text{-}t,QS}$; $\star$ $CI_{per\text{-}t,var^*}$).

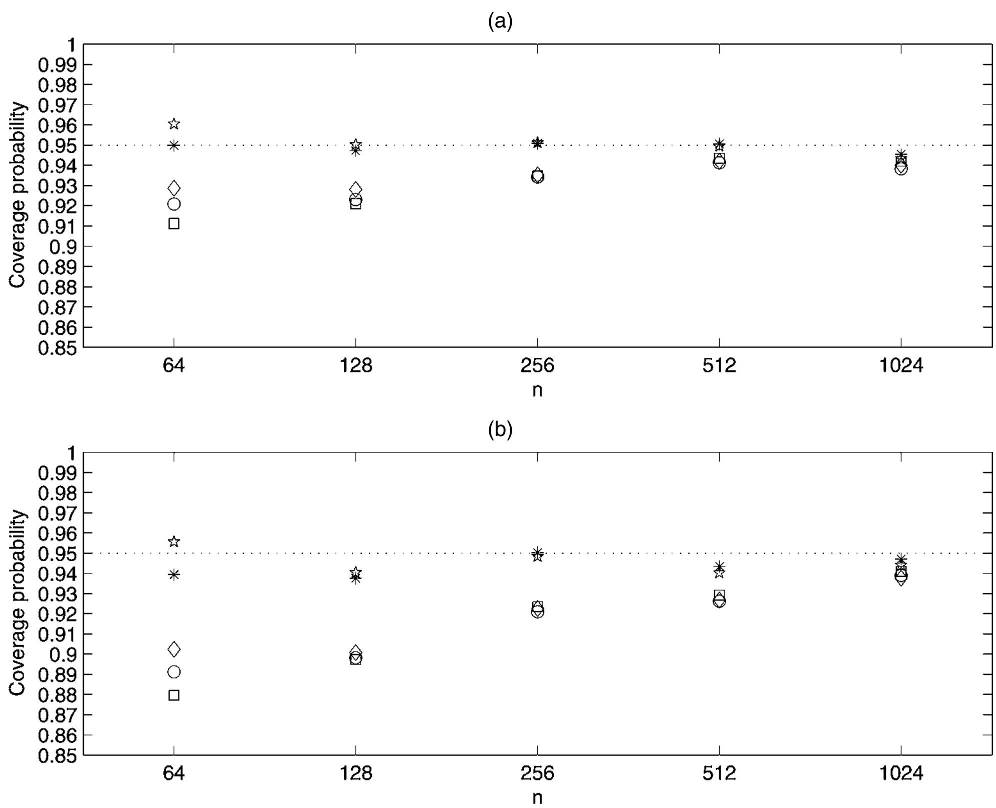

Figure 3. Coverage Probabilities of 95% Nominal Symmetric Confidence Intervals. Regression errors are heteroscedastic AR(1) with autoregression coefficient equal to ρ. For (a), $\rho = .30$; for (b), $\rho = .50$ ($\sqsubset$ CI_{QS}; $\diamond$ CI_{var^*}; $\circ$ CI_{per}; $*$ $CI_{per\text{-}t,QS}$; $\star$ $CI_{per\text{-}t,var^*}$).

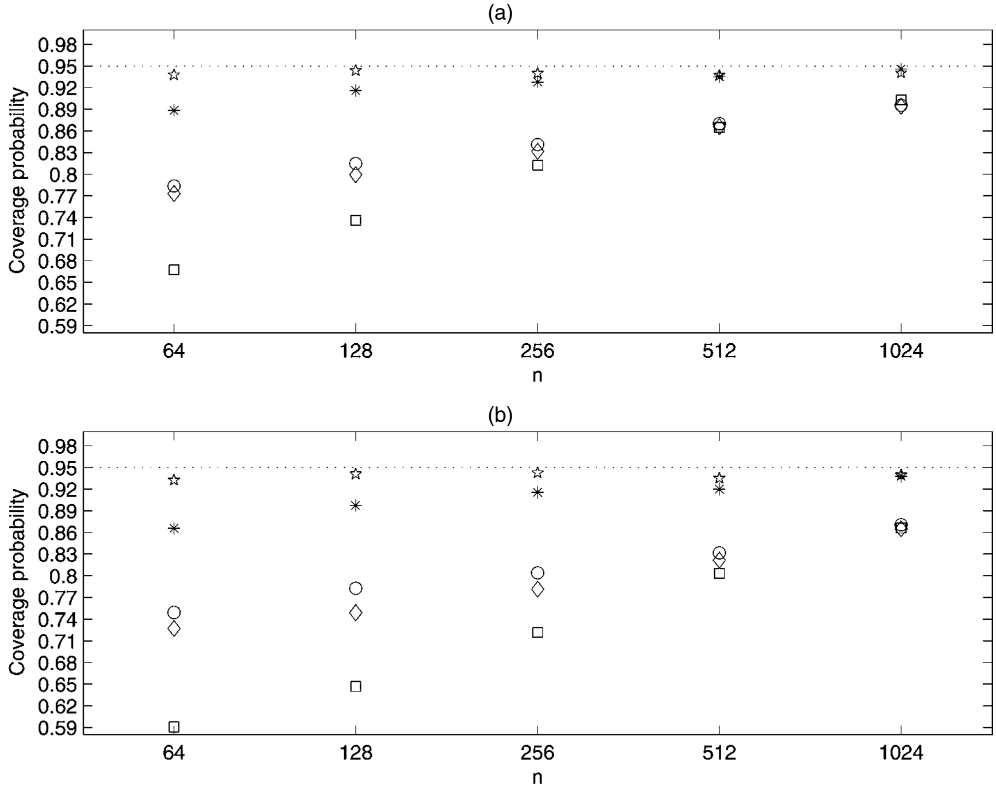

Figure 4. Coverage Probabilities of 95% Nominal Symmetric Confidence Intervals. Regression errors are heteroscedastic AR(1) with autoregression coefficient equal to ρ. For (a), $\rho = .90$; for (b), $\rho = .95$ (□ CI_{QS}; ◇ CI_{var^*}; ○ CI_{per}; ✳ $CI_{per\text{-}t,QS}$; ☆ $CI_{per\text{-}t,var^*}$).

data is important. Using the bootstrap standard error estimate instead of the QS kernel-based standard error estimate results in better finite-sample performance, especially under strong autocorrelation in the errors. The presence of heteroscedasticity [i.e., for AR(1)–HET2] leads to smaller coverage rates for both bootstrap percentile-t confidence intervals, which results in worse undercoverage for $CI_{\text{per-}t,\text{QS}}$ and some undercoverage for $CI_{\text{per-}t,\text{var}^*}$. Nevertheless, here too replacing the QS-kernel standard error used to studentize the original t-statistic by the bootstrap standard error estimate helps reduce the coverage error of bootstrap percentile-t confidence intervals.

4. CONCLUSIONS

This article gives conditions under which the MBB of Künsch (1989) and Liu and Singh (1992a) provides consistent estimators of the asymptotic variance of the LS estimator in (possibly misspecified) linear regression models. Although we have focused on the MBB, similar results hold for the stationary bootstrap of Politis and Romano (1994) (also see Gonçalves 2000). The Monte Carlo results obtained in this article indicate that bootstrap variance-based percentile-t confidence intervals have coverage rates closer to the desired levels in the context of a particular linear regression model. This is an interesting finding. An important topic for future research would be to obtain formal conditions under which bootstrap standard error estimates have better higher-order asymptotic accuracy than conventional first-order asymptotic theory-based standard error estimates. This could help explain the improved accuracy of bootstrap standard error-based confidence intervals found in our Monte Carlo experiments.

APPENDIX: PROOFS

Throughout this appendix, C denotes a generic constant that does not depend on n, and $\mathbb{1}(A)$ denotes the indicator function of any set A. In obtaining our results, we use the mixingale property of processes NED on a mixing process. The concept of L_2-mixingales was introduced by McLeish (1975) and generalized to L_q-mixingales (for $q > 1$) by Andrews (1988). Let $(\Omega, \mathcal{G}, P)$ be a probability space on which a sequence of random variables $\{\mathbf{Z}_t\}_{t=1}^{\infty}$ is defined, and let $\{\mathcal{G}^t\}_{t=1}^{\infty}$ be a nondecreasing sequence of sub–σ-fields of $\mathcal{G}$. We say that $\{\mathbf{Z}_t, \mathcal{G}^t\}_{t=1}^{\infty}$ is an L_q-mixingale (for some $q > 1$) if there exist nonnegative constants $\{c_t\}_{t=1}^{\infty}$ and $\{\psi_m\}_{m=0}^{\infty}$ such that $\psi_m \to 0$ as $m \to \infty$, and for all $t \ge 1$ and $m \ge 0$ we have $\|E(\mathbf{Z}_t|\mathcal{G}^{t-m})\|_q \le c_t\psi_m$ and $\|\mathbf{Z}_t - E(\mathbf{Z}_t|\mathcal{G}^{t+m})\|_q \le c_t\psi_{m+1}$. We make use of the following result.

Lemma A.1. For $q \ge 2$, let $\{Z_t, \mathcal{G}^t\}$ be an L_q-mixingale with bounded mixingale constants $\{c_t\}$ and mixingale coefficients $\{\psi_m\}$ satisfying $\sum_{m=1}^{\infty}\psi_m < \infty$. Let $\{Z_{nt}^* : t = 1, \dots, n\}$ denote a bootstrap resample of $\{\mathbf{Z}_t : t = 1, \dots, n\}$ obtained with the MBB. If $\ell_n = o(n)$ with $\ell_n \to \infty$, then $E(E^*|\sum_{t=1}^n Z_{nt}^*|^q) = O(n^{q/2}) + O(\ell_n^q)$.

Proof. We follow Künsch (1989) and write $\sum_{t=1}^n Z_{nt}^* = \sum_{i=1}^k Y_{n,i}$, where $\{Y_{n,i}\}$ are iid with $P^*(Y_{n,i} = Z_{j+1} + \cdots + Z_{j+\ell_n}) = \frac{1}{n-\ell_n+1}$, $j = 0, \dots, n - \ell_n$. Hence,

$$E^*\left|\sum_{t=1}^n Z_{nt}^*\right|^q$$

$$= E^*\left|\sum_{i=1}^k (Y_{n,i} - E^*(Y_{n,1})) + kE^*(Y_{n,1})\right|^q$$

$$\leq 2^{q-1}\left\{E^*\left|\sum_{i=1}^{k}(Y_{n,i}-E^*(Y_{n,1}))\right|^q + E^*|kE^*(Y_{n,1})|^q\right\}$$
$$\equiv 2^{q-1}(A_n+B_n),$$

by an application of the c_r-inequality (see, e.g., Davidson 1994, p. 140). First, consider A_n. By an extension of Burkholder's inequality to martingale difference arrays, $A_n \leq CE^*|\sum_{i=1}^{k}|Y_{n,i}-E^*(Y_{n,1})|^2|^{q/2}$, and by Hölder's inequality,

$$E^*\left|\sum_{i=1}^{k}|Y_{n,i}-E^*(Y_{n,1})|^2\right|^{q/2}$$
$$\leq E^*\left|\left(\sum_{i=1}^{k}|Y_{n,i}-E^*(Y_{n,1})|^q\right)^{2/q}k^{1-2/q}\right|^{q/2}$$
$$\leq 2^q k^{q/2}E^*|Y_{n,1}|^q, \tag{A.1}$$

where the last inequality follows by a simultaneous application of the c_r-inequality and Jensen's inequality. Because $E^*|Y_{n,1}|^q = (n-\ell_n+1)^{-1}\sum_{j=0}^{n-\ell_n}|\sum_{t=1}^{\ell_n}Z_{t+j}|^q$, we have

$$E(E^*|Y_{n,1}|^q) = (n-\ell_n+1)^{-1}\sum_{j=0}^{n-\ell_n}E\left|\sum_{t=1}^{\ell_n}Z_{t+j}\right|^q$$
$$\leq (n-\ell_n+1)^{-1}\sum_{j=0}^{n-\ell_n}\left(C\ell_n^{1/2}\right)^q$$
$$= (C)^q\ell_n^{q/2}, \tag{A.2}$$

by a maximal inequality for mixingales (Hansen 1991, lemma 2; 1992), and by the boundedness assumption on the mixingale constants $\{c_t\}$. Because $k = n/\ell_n$, it follows from (A.1) and (A.2) that $E(A_n) = O(n^{q/2})$. Next, consider B_n. Noting that $E^*(\bar{Z}_n^*) = \ell_n^{-1}E^*(Y_{n,1})$, we can write

$$E(B_n) = E\left(E^*|kE^*(Y_{n,1})|^q\right)$$
$$= (k\ell_n)^q E\left(|\ell_n^{-1}E^*(Y_{n,1})|^q\right)$$
$$= n^q E\left(|E^*(\bar{Z}_n^*)|^q\right).$$

By the properties of the MBB, we can write

$$E^*(\bar{Z}_n^*) = \frac{1}{n-\ell_n+1}\left(\sum_{t=1}^{\ell_n-1}\left(\frac{t}{\ell_n}\right)Z_t + \sum_{t=\ell_n}^{n-\ell_n+1}\left(\frac{1}{\ell_n}\right)Z_t + \sum_{t=n-\ell_n+2}^{n}\left(\frac{n-t+1}{\ell_n}\right)Z_t\right)$$
$$= \frac{1}{n-\ell_n+1}\sum_{t=1}^{n}Z_t - \frac{1}{n-\ell_n+1}\sum_{t=1}^{\ell_n-1}\left(1-\frac{t}{\ell_n}\right)Z_t - \frac{1}{n-\ell_n+1}\sum_{t=n-\ell_n+2}^{n}\left(1-\frac{n-t+1}{\ell_n}\right)Z_t$$
$$\equiv A_{n1} - A_{n2} - A_{n3},$$

which implies that $E(B_n) = n^q E[|A_{n1}-A_{n2}-A_{n3}|^q] \leq 3^{q-1}\times n^q(E|A_{n1}|^q + E|A_{n2}|^q + E|A_{n3}|^q)$. By the maximal inequality for mixingales, $E|A_{n1}|^q = O(n^{-q/2})$ if $\ell_n = o(n)$. Similarly, by the c_r-inequality and the fact that $E|Z_t|^q \leq \Delta < \infty$ (given the L^q-mixingale assumption), we have that

$$E|A_{n2}|^q \leq (n-\ell_n+1)^{-q}(\ell_n-1)^{q-1}\sum_{t=1}^{\ell_n-1}\left|1-\frac{t}{\ell_n}\right|^q E|Z_t|^q$$
$$= O\left(\frac{\ell_n^q}{(n-\ell_n+1)^q}\right).$$

By a similar argument, $E|A_{n3}|^q = O(\frac{\ell_n^q}{(n-\ell_n+1)^q})$. Hence, because $\ell_n = o(n)$, $E(B_n) \leq O(n^{q/2}) + O(\ell_n^q)$, completing the proof.

Proof of Theorem 1

The proof proceeds in three steps.

Step 1. We first show that for any $\xi > 0$ and for all n sufficiently large, there exists $\eta > 0$ such that

$$P\left[\omega : P_\omega^*\left(\lambda : \lambda_{\min}\left(\frac{\mathbf{X}^{*\prime}(\lambda,\omega)\mathbf{X}^*(\lambda,\omega)}{n}\right) < \eta/2\right) > \xi\right] < \xi. \tag{A.3}$$

For clarity in the argument that follows, it is important to make explicit the dependence of the bootstrap probability measure P^* on $\omega \in \Omega$, as was done by Gonçalves and White (2004). Similarly, we write $\mathbf{X}^*(\lambda,\omega)$ to emphasize the fact that for each $\omega \in \Omega$ and for $t = 1, 2, \ldots, n$, we let $\mathbf{X}_t^* = \mathbf{X}_{\tau_t(\lambda)}(\omega)$, where $\tau_t(\lambda)$ is a realization of the random index chosen by the MBB. Fix $\xi > 0$ arbitrarily. For $\epsilon > 0$ (to be chosen shortly), define $A_{n,\epsilon} \equiv \{\omega : |\lambda_{\min}(\frac{\mathbf{X}'(\omega)\mathbf{X}(\omega)}{n}) - \lambda_{\min}(\mathbf{A}^o)| \leq \epsilon\}$. Note that for any $\omega \in A_{n,\epsilon}$, $\lambda_{\min}(\frac{\mathbf{X}'\mathbf{X}}{n}) \geq \eta - \epsilon$, given that $\lambda_{\min}(\mathbf{A}^o) > \eta > 0$, under Assumption 1.e. Next, for any ω, define $C_{n,\omega,\epsilon} \equiv \{\lambda : |\lambda_{\min}(\frac{\mathbf{X}^{*\prime}(\lambda,\omega)\mathbf{X}^*(\lambda,\omega)}{n}) - \lambda_{\min}(\frac{\mathbf{X}'(\omega)\mathbf{X}(\omega)}{n})| \leq \epsilon\}$ and note that for $\omega \in A_{n,\epsilon}$, $C_{n,\omega,\epsilon}$ implies $B_{n,\omega,\epsilon} \equiv \{\lambda : \lambda_{\min}(\frac{\mathbf{X}^{*\prime}(\lambda,\omega)\mathbf{X}^*(\lambda,\omega)}{n}) \geq \eta - 2\epsilon\}$. Thus $A_{n,\epsilon} \cap C_{n,\omega,\epsilon} \subseteq B_{n,\omega,\epsilon}$, which implies that $P_\omega^*(B_{n,\omega,\epsilon}^c) \leq P_\omega^*(A_{n,\epsilon}^c) + P_\omega^*(C_{n,\omega,\epsilon}^c)$. Choosing $\epsilon = \eta/4$, it follows that

$$P\left(P_\omega^*\left(B_{n,\omega,\eta/4}^c\right) > \xi\right)$$
$$\leq P\left(P_\omega^*\left(A_{n,\eta/4}^c\right) > \xi/2\right) + P\left(P_\omega^*\left(C_{n,\omega,\eta/4}^c\right) > \xi/2\right)$$
$$< \frac{\xi}{2} + \frac{\xi}{2} = \xi,$$

where the last inequality holds because $P(P_\omega^*(A_{n,\eta/4}^c) > \xi/2) = P(A_{n,\eta/4}^c) < \xi/2$ for all n sufficiently large (by convergence of $\frac{\mathbf{X}'\mathbf{X}}{n}$ to $\mathbf{A}^o$) and because $P(P_\omega^*(C_{n,\omega,\eta/4}^c) > \xi/2) < \frac{\xi}{2}$ for all n sufficiently large [by lemma A.5 of Gonçalves and White 2004, given that $\ell_n = o(n)$]. This proves (A.3).

Step 2. $\mathbf{B}^{o-1/2}\mathbf{A}^o\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \Rightarrow^{d_{P^*}} \mathrm{N}(\mathbf{0}, \mathbf{I}_p)$ in probability. We can write $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) = \sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) + \mathbf{R}_n^*$, with $\mathbf{R}_n^* = -\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)\mathbb{1}\{\lambda_{\min}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) < \eta/2\}$, given the definition of $\tilde{\boldsymbol{\beta}}_n^*$ [with $\delta = \eta/2 > 0$ and η such that $\lambda_{\min}(\mathbf{A}^o) > \eta > 0$]. Because under our assumptions, by an application of theorem 2.2 of Gonçalves and White (2004), $\mathbf{B}^{o-1/2}\mathbf{A}^o\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) \Rightarrow^{d_{P^*}} \mathrm{N}(\mathbf{0}, \mathbf{I}_p)$ in probability, it suffices to show that $\mathbf{R}_n^* = o_{P^*}(1)$ in probability. For this, note that $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n) = O_{P^*}(1)$, except in a set with probability tending to 0. Moreover, $E^*(\mathbb{1}\{\lambda_{\min}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) < \eta/2\}) = P^*(\lambda_{\min}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) < \eta/2) \xrightarrow{P} 0$, as we showed in Step 1. This implies that $\mathbb{1}\{\lambda_{\min}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) < \eta/2\} \xrightarrow{P^*} 0$ in probability, proving Step 2.

Step 3. For some $\delta > 0$, $E^*(|\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)|^{2+\delta}) = O_P(1)$. Given the definition of $\tilde{\boldsymbol{\beta}}_n^*$, we can write

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_n^* - \hat{\boldsymbol{\beta}}_n)$$
$$= \left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)^{-1}\mathbb{1}\left(\lambda_{\min}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right) \geq \eta/2\right)n^{-1/2}\sum_{t=1}^{n}\mathbf{X}_{nt}^*\hat{\varepsilon}_{nt}^*.$$

By a well-known inequality for matrix norms (see, e.g., Strang 1988, p. 369, ex. 7.2.3), it follows that

$$\begin{aligned}&|\sqrt{n}(\tilde{\boldsymbol\beta}_n^* - \hat{\boldsymbol\beta}_n)|^{2+\delta}\\ &\le \left\|\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)^{-1}\right\|_1^{2+\delta} \mathbb{1}\left(\lambda_{\min}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\ge \eta/2\right)\\ &\times \left|n^{-1/2}\sum_{t=1}^n \mathbf{X}_{nt}^*\hat\varepsilon_{nt}^*\right|^{2+\delta}.\end{aligned}$$

Here, for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_1$ denotes the matrix norm defined by $\|\mathbf{A}\|_1^2 = \max_{\mathbf{x}\neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}}$. For $\mathbf{A}$ symmetric, $\|\mathbf{A}\|_1$ is equal to the largest eigenvalue of $\mathbf{A}$, that is, $\|\mathbf{A}\|_1 = \lambda_{\max}(\mathbf{A})$. When $\lambda_{\min}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) \ge \eta/2$, $\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}$ is symmetric and positive definite and we have that

$$\begin{aligned}\left\|\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)^{-1}\right\|_1 &= \lambda_{\max}\left(\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)^{-1}\right)\\ &= \lambda_{\min}^{-1}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right) \le \left(\frac{\eta}{2}\right)^{-1} = C.\end{aligned}$$

Thus

$$|\sqrt{n}(\tilde{\boldsymbol\beta}_n^* - \hat{\boldsymbol\beta}_n)|^{2+\delta} \le C\left|n^{-1/2}\sum_{t=1}^n \mathbf{X}_{nt}^*\hat\varepsilon_{nt}^*\right|^{2+\delta},$$

and it suffices to show that $E^*(|n^{-1/2}\sum_{t=1}^n \mathbf{X}_{nt}^*\hat\varepsilon_{nt}^*|^{2+\delta}) = O_P(1)$. Let $\hat\varepsilon_{nt}^* \equiv Y_{nt}^* - \mathbf{X}_{nt}^{*\prime}\hat{\boldsymbol\beta}_n$ and $\varepsilon_{nt}^* = Y_{nt}^* - \mathbf{X}_{nt}^{*\prime}\boldsymbol\beta^o$. Using these definitions and applying the c_r-inequality yields

$$\begin{aligned}&\left|n^{-1/2}\sum_{t=1}^n \mathbf{X}_{nt}^*\hat\varepsilon_{nt}^*\right|^{2+\delta}\\ &\le 2^{1+\delta}\left(\left|n^{-1/2}\sum_{t=1}^n \mathbf{X}_{nt}^*\varepsilon_{nt}^*\right|^{2+\delta} + \left|\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta^o)\right|^{2+\delta}\right)\\ &\equiv C(A_1^* + A_2^*).\end{aligned}$$

By Lemma A.1, we can show that

$$\begin{aligned}E(E^*(A_1^*)) &\le Cn^{-(2+\delta)/2}E\left[E^*\left(\left|\sum_{t=1}^n \mathbf{X}_{nt}^*\varepsilon_{nt}^*\right|^{2+\delta}\right)\right]\\ &= O(1) + O\left(\left(\frac{\ell_n^2}{n}\right)^{(2+\delta)/2}\right),\end{aligned}$$

which is $O(1)$ because $\ell_n^2/n \to 0$. To apply Lemma A.1, we need $\{\mathbf{X}_t\varepsilon_t\}$ to be a mean-0 $L_{2+\delta}$-mixingale with bounded mixingale constants $\{c_t\}$ and absolutely summable mixingale coefficients $\{\psi_m\}$, which holds under our assumptions. Thus, by Markov's inequality, $E^*(A_1^*) = O_P(1)$. For A_2^*, note that

$$\begin{aligned}A_2^* &\le \left\|\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\right\|_1^{2+\delta}|\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta^o)|^{2+\delta}\\ &= \lambda_{\max}^{2+\delta}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)|\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta^o)|^{2+\delta},\end{aligned}$$

implying that

$$E^*(A_2^*) \le E^*\left(\lambda_{\max}^{2+\delta}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\right)|\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta^o)|^{2+\delta}.$$

Because $\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta^o)$ converges in distribution, it follows that $|\sqrt{n}(\hat{\boldsymbol\beta}_n - \boldsymbol\beta^o)|^{2+\delta} = O_P(1)$. Thus, to prove that $E^*(A_2^*) = O_P(1)$, it suffices that $E^*(\lambda_{\max}^{2+\delta}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n})) = O_P(1)$. For this, note that $0 < \lambda_{\max}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) \le \mathrm{tr}(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}) = \sum_{i=1}^p (n^{-1}\sum_{t=1}^n X_{ti}^{*2})$, where, for any matrix $\mathbf{A}$, $\lambda_i(\mathbf{A})$ denotes its ith eigenvalue and $\mathrm{tr}(\mathbf{A})$ denotes its trace. Thus

$$\begin{aligned}E^*\left(\lambda_{\max}^{2+\delta}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\right) &\le E^*\left[\left(\mathrm{tr}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\right)^{2+\delta}\right]\\ &\le C\sum_{i=1}^p n^{-(2+\delta)}E^*\left(\left|\sum_{t=1}^n X_{ti}^{*2}\right|^{2+\delta}\right), \qquad \text{(A.4)}\end{aligned}$$

by an application of the c_r-inequality. Adding and subtracting appropriately yields

$$\begin{aligned}E^*\left(\left|\sum_{t=1}^n X_{ti}^{*2}\right|^{2+\delta}\right) &= E^*\left(\left|\sum_{t=1}^n (X_{ti}^{*2} - \mu_{2i}) + n\mu_{2i}\right|^{2+\delta}\right)\\ &\le CE^*\left(\left|\sum_{t=1}^n W_{ti}^*\right|^{2+\delta}\right) + \mu_{2i}^{2+\delta}n^{2+\delta}, \qquad \text{(A.5)}\end{aligned}$$

where we let $W_{ti}^* \equiv X_{ti}^{*2} - \mu_{2i}$ be the resampled version of $W_{ti} = X_{ti}^2 - \mu_{2i}$, with $\mu_{2i} \equiv E(X_{ti}^2)$. Under Assumption 1, we can show that $\{W_{ti}, \mathcal{F}^t\}$ is an $L_{2+\delta}$-mixingale with bounded mixingale constants $\{c_t\}$ and absolutely summable coefficients $\{\psi_m\}$. Thus, by Lemma A.1, we have that

$$E\left[E^*\left(\left|\sum_{t=1}^n W_{ti}^*\right|^{2+\delta}\right)\right] = O\left(n^{(2+\delta)/2}\right) + O(\ell_n^{2+\delta}). \qquad \text{(A.6)}$$

From (A.4), (A.5), and (A.6), it follows that

$$\begin{aligned}&E\left[E^*\left(\lambda_{\max}^{2+\delta}\left(\frac{\mathbf{X}^{*\prime}\mathbf{X}^*}{n}\right)\right)\right]\\ &\qquad = O\left(n^{-(2+\delta)/2}\right) + O\left(\left(\frac{\ell_n}{n}\right)^{2+\delta}\right) + O(1),\end{aligned}$$

which is $O(1)$ because $\frac{\ell_n}{n} \to 0$. This completes the proof of Step 3.

[Received August 2003. Revised November 2004.]

REFERENCES

Andrews, D. W. K. (1984), "Non-Strong Mixing Autoregressive Processes," *Journal of Applied Probability*, 21, 930–934.

——— (1988), "Laws of Large Numbers for Dependent Nonidentically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

——— (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

——— (1995), *Probability and Measure*, New York: Wiley.

Buchinsky, M. (1995), "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford, U.K.: Oxford University Press.

——— (2002), "Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes," *Journal of Econometrics*, 106, 243–269.

Davidson, R., and MacKinnon, J. G. (1999), "Bootstrap Testing in Nonlinear Models," *International Economic Review*, 40, 487–508.

Davison, A. C., and Hall, P. (1993), "On Studentizing and Blocking Methods for Implementing the Bootstrap With Dependent Data," *Australian Journal of Statistics*, 35, 215–224.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Efron, B., and Tibshirani, R. J. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy" (with discussion), *Statistica Sinica*, 1, 54–77.

Fitzenberger, B. (1997), "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions," *Journal of Econometrics*, 82, 235–287.

Freedman, D. A., and Peters, S. C. (1984), "Bootstrapping a Regression Equation: Some Empirical Results," *Journal of the American Statistical Association*, 79, 97–106.

Gallant, A. R., and White, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York: Basil Blackwell.

Ghosh, M., Parr, W. C., Singh, K., and Babu, G. J. (1984), "A Note on Bootstrapping the Sample Median," *The Annals of Statistics*, 12, 1130–1135.

Gonçalves, S. (2000), "The Bootstrap for Dependent Heterogeneous Processes," unpublished doctoral dissertation, University of California, San Diego.

Gonçalves, S., and White, H. (2002), "The Bootstrap of the Mean for Dependent Heterogeneous Arrays," *Econometric Theory*, 18, 1367–1384.

——— (2004), "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," *Journal of Econometrics*, 119, 199–219.

Götze, F., and Künsch, H. R. (1996), "Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations," *The Annals of Statistics*, 24, 1914–1933.

Hahn, J. (1995), "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11, 105–121.

Hansen, B. E. (1991), "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221.

——— (1992), "Erratum," *Econometric Theory*, 8, 421–422.

Hodrick, R. J. (1992), "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357–386.

Kilian, L. (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long Horizon Regressions?" *Journal of Applied Econometrics*, 14, 491–510.

Kirby, C. (1998), "The Restrictions on Predictability Implied by Rational Asset Pricing Models," *Review of Financial Studies*, 11, 343–382.

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

Li, H., and Maddala, G. S. (1999), "Bootstrap Variance Estimation of Nonlinear Functions of Parameters: An Application to Long-Run Elasticities of Energy Demand," *Review of Economics and Statistics*, 81, 728–733.

Liu, R. Y., and Singh, K. (1992a), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of the Bootstrap*, eds. R. LePage and L. Billiard, New York: Wiley, pp. 225–248.

——— (1992b), "Efficiency and Robustness in Resampling," *The Annals of Statistics*, 20, 370–384.

Mark, N. (1995), "Exchange Rates and Fundamentals: Evidence on Long-Run Predictability," *American Economic Review*, 85, 201–218.

McLeish, D. L. (1975), "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 5, 616–621.

Politis, D., and Romano, J. (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Politis, D., Romano, J., and Wolf, M. (1997), "Subsampling for Heteroskedastic Time Series," *Journal of Econometrics*, 81, 281–317.

Romano, J., and Wolf, M. (2003), "Improved Nonparametric Confidence Intervals in Time Series Regressions," mimeo, University of Pompeu Fabra, Dept. of Economics and Business.

Shao, J. (1992), "Bootstrap Variance Estimators With Truncation," *Statistics and Probability Letters*, 15, 95–101.

Shao, J., and Tu, D. (1995), *The Jackknife and Bootstrap*, New York: Springer-Verlag.

Strang, G. (1988), *Linear Algebra and Its Applications* (3rd ed.), Orlando, FL: Harcourt.