

Bootstrap Standard Error Estimates for Linear Regression

Sílvia Gonçalves
C.R.D.E, CIRANO and Université de Montréal* Halbert White
University of California, San Diego†

August 2, 2004

Abstract

Standard errors of parameter estimates are widely used in empirical work. The bootstrap can often provide a convenient means of estimating standard errors. The conditions under which bootstrap standard error estimates are theoretically justified, however, have not received much attention. This paper establishes conditions for the consistency of the moving blocks bootstrap (Künsch, 1989, and Liu and Singh, 1992) 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, we discuss 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.

Keywords: Bootstrap standard errors; 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, Theorem 25.12). For the sample median and smooth functions of sample means, examples of the inconsistency of bootstrap variance estimators in the i.i.d. context can be found in Ghosh et al. (1984) and Shao (1992), respectively.

*C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. Tel: (514) 343 6556. Email: silvia.goncalves@umontreal.ca. Gonçalves acknowledges financial support from Praxis XXI and a Sloan Dissertation fellowship.

†9500 Gilman Drive, La Jolla, California, 92093-0508. USA . Tel: (858) 534-3502. Email: hwhite@weber.ucsd.edu. White’s participation was supported by NSF grants SBR-9811562 and SES-0111238.

For time series observations, the moving blocks bootstrap (MBB) introduced by Künsch (1989) and Liu and Singh (1992a) has been shown to estimate consistently 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 paper 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 i.i.d. observations as a special case.¹ Although the consistency of the MBB distribution of the LS estimator is well established in the literature (see e.g. Fitzenberger 1997 and Politis et al. 1997), the consistency of the bootstrap variance of the LS estimator has not received much attention. As we remarked above, the former does not necessarily imply the latter, so that results currently available 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 a normalization by the standard error of the estimator. Our results formally justify the use of 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

¹In related work, Liu and Singh (1992b) show the consistency of the i.i.d. bootstrap variance estimator for regressions with fixed regressors and i.i.d. errors. Our results allow for stochastic regressors and autocorrelated errors.

regression model with autocorrelated (and heteroskedastic) 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 paper is organized as follows. Section 2 presents the theoretical results. In section 3, we compare the accuracy of the bootstrap estimator to that of closed form estimators of the variance. Concluding remarks are contained in section 4. All proofs are in the Appendix.

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

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

Gallant and White (1988) studied the asymptotic properties of quasi-maximum likelihood es-

timators (QMLE) for heterogeneous near epoch dependent 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) show that the MBB consistently estimates the asymptotic distribution of the QMLE. As Gonçalves and White (2004) remark, 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 least squares estimator for linear dynamic models. In particular, we give explicit conditions that justify bootstrapping the variance of the least squares (LS) estimator in possibly misspecified linear dynamic models when the data generating process is NED on a mixing process.

Assumption 1 below 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

- i) Let (Ω, \mathcal{F}, P) be a complete probability space. The observed data are a realization of a strictly stationary stochastic process $\{Z_t = (Y_t, X_t')' : \Omega \rightarrow \mathbb{R}^{p+1}, t = 1, 2, \dots\}$, $p \in \mathbb{N}$; $Z_t(\omega) = W_t(\dots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \dots)$, $\omega \in \Omega$, where $V_t : \Omega \rightarrow \mathbb{R}^v$, $v \in \mathbb{N}$, and $W_t : \times_{\tau=-\infty}^{\infty} \mathbb{R}^v \rightarrow \mathbb{R}^{p+1}$ are such that Z_t is measurable, $t = 1, 2, \dots$.
- ii) $Y_t = X_t' \beta^o + \varepsilon_t$, $t = 1, 2, \dots$, for some $\beta^o \in \mathbb{R}^p$, where $X_t' = (X_{t1}, \dots, X_{tp})$ and $E(X_t \varepsilon_t) = 0$.
- iii) For some $r > 2$, $\|Y_t\|_{6r} \leq \Delta < \infty$, $\|X_{ti}\|_{6r} \leq \Delta < \infty$, for $i = 1, \dots, p$, $t = 1, 2, \dots$.
- iv) For some small $\delta > 0$, the elements of $\{Z_t\}$ are $L_{2+\delta}$ -NED on $\{V_t\}$ with NED coefficients v_k of size $-\frac{4(r-1)^2}{(r-2)^2}$; $\{V_t\}$ is an α -mixing sequence with α_k of size $-\frac{(2+\delta)r}{r-2}$.
- v) $A^o \equiv E(X_t X_t')$ is nonsingular, i.e. $\lambda_{\min}(A^o) \geq \eta > 0$ for some $\eta > 0$, where $\lambda_{\min}(A^o)$ denotes the smallest eigenvalue of A^o .
- vi) $B^o \equiv \lim_{n \rightarrow \infty} B_n^o$ is positive definite, where $B_n^o = \text{var}(n^{-1/2} \sum_{t=1}^n X_t \varepsilon_t)$.

According to Assumption 1.i), we observe data on $(p + 1) \times 1$ random vectors $Z_t = (Y_t, X_t')'$, which are each viewed as a transformation of some underlying process $\{V_t\}$. Here, Y_t denotes the

observation t on the dependent variable and $X_t \equiv (X_{t1}, \dots, X_{tp})'$ is the $p \times 1$ vector of regressors for observation t ; X_t may include lagged dependent variables. For simplicity, we assume that the data generating process for 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.ii) specifies a linear dynamic model which may be misspecified in the sense that for all $\beta \in \mathbb{R}^p$ it is true that $P(E(Y_t|X_t) \neq X_t'\beta) > 0$. Such models are relevant for forecasting, as in this misspecified context, β^o is the parameter that minimizes the mean squared error of the linear approximation to the unknown $E(Y_t|X_t)$. In particular, under Assumption 1.v), β^o is uniquely defined by $\beta^o = (E(X'X))^{-1} E(X'Y)$, where we let $Y = (Y_1, \dots, Y_n)'$ and $X = (X_1, \dots, X_n)'$.

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

Let $\hat{\beta}_n^* = (X^{*'}X^*)^{-1} X^{*'}Y^*$ be the LS estimator of β^o based on the bootstrap data $\{Z_{nt}^* = (Y_{nt}^*, X_{nt}^{*'})'\}$ 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 $B_{t,\ell} = \{Z_t, Z_{t+1}, \dots, Z_{t+\ell-1}\}$ be the block of ℓ consecutive observations starting at Z_t ; $\ell = 1$ corresponds to the standard i.i.d. bootstrap. The MBB resamples $k = n/\ell$ blocks randomly with replacement from the set of $n - \ell + 1$ overlapping blocks $\{B_{1,\ell}, \dots, B_{n-\ell+1,\ell}\}$, where we assume for simplicity that $n = k\ell$.

One bootstrap variance-covariance matrix estimator of $\hat{\beta}_n$ is given by the bootstrap population variance-covariance matrix of $\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n)$, conditional on the original data, $\hat{C}_n^* = Var^*(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n))$. Since in general there is no closed form expression for \hat{C}_n^* , we compute an approximation to \hat{C}_n^* by Monte Carlo simulation, i.e. $\hat{C}_n^* = \lim_{B \rightarrow \infty} \frac{n}{B} \sum_{i=1}^B (\hat{\beta}_n^{*(i)} - \overline{\hat{\beta}_n^*}) (\hat{\beta}_n^{*(i)} - \overline{\hat{\beta}_n^*})'$, where $\overline{\hat{\beta}_n^*} = \frac{1}{B} \sum_{i=1}^B \hat{\beta}_n^{*(i)}$, with $\hat{\beta}_n^{*(i)}$ the bootstrap LSE evaluated on the i^{th} bootstrap replication and B the total number of bootstrap replications.

In this paper, we focus on an alternative bootstrap variance-covariance matrix estimator of $\hat{\beta}_n$,

namely the bootstrap population variance-covariance matrix of $\sqrt{n}(\tilde{\beta}_n^* - \hat{\beta}_n)$. Following Liu and Singh (1992b) and Shao and Tu (1995, Chapter 7, Section 7.2.2), we define $\tilde{\beta}_n^*$ as follows:

$$(1) \quad \tilde{\beta}_n^* = \begin{cases} (X^{*'}X^*)^{-1}X^{*'}Y^* & \text{if } \lambda_{\min}\left(\frac{X^{*'}X^*}{n}\right) \geq \frac{\delta}{2}, \\ \hat{\beta}_n & \text{otherwise,} \end{cases}$$

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

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

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

An important intermediate step towards proving the consistency of \tilde{C}_n^* for C^o is to establish the first-order asymptotic validity of the bootstrap distribution of $\sqrt{n}(\tilde{\beta}_n^* - \hat{\beta}_n)$. 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 $\{X_t\varepsilon_t\}$ 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 since $E(X_t\varepsilon_t) = 0$ for all t given the stationarity assumption and the definition of β^o . Alternatively, for Assumption 2.1 we could have assumed: (i) $\|X_{ti\varepsilon_t}\|_{3r} \leq \Delta < \infty$ for some $r > 2$, and (ii) for some $\delta > 0$, the elements of $\{X_t\varepsilon_t\}$ are $L_{2+\delta}$ -NED on $\{V_t\}$ with NED coefficients v_k of size $-\frac{2(r-1)}{r-2}$.

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

Theorem 2.1 *Under Assumption 1, if $\ell_n = o(n^{1/2})$ and $\ell_n \rightarrow \infty$, then $\tilde{C}_n^* \xrightarrow{P} C^o$, where $\tilde{C}_n^* \equiv Var^* \left(\sqrt{n}\tilde{\beta}_n^* \right)$ and $C^o \equiv A^{o-1}B^oA^{o-1}$.*

Theorem 2.1 justifies the use of \tilde{C}_n^* as an heteroskedasticity and autocorrelation consistent (HAC) variance estimator of C^o . Given the first-order asymptotic validity of the bootstrap distribution of $\sqrt{n}(\tilde{\beta}_n^* - \hat{\beta}_n)$, we show that $E^* \left(\left| \sqrt{n}(\tilde{\beta}_n^* - \hat{\beta}_n) \right|^{2+\delta} \right) = O_P(1)$, which is a sufficient

condition for the uniform integrability of the sequence $\left\{ \left| \sqrt{n} \left(\tilde{\beta}_n^* - \hat{\beta}_n \right) \right|^2 \right\}$.

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 QMLE estimators as proposed by Davidson and MacKinnon (1999). To illustrate, consider the one-step QMLE as defined in Gonçalves and White (2004):

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

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

Another useful extension of the results presented here is to quantile regression. Since 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, as this avoids nonparametric density estimation. For cross section quantile regression, Buchinsky (1995) investigates 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 is provided. Also for the cross section context, Hahn (1995) proves the first-order asymptotic validity of the bootstrap distribution of the quantile regression estimator. As

Hahn (1995, p. 107) remarks, 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) proves that the MBB estimates consistently 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) treats the quantile regression estimator in a setting analogous to the LS case. Therefore, we conjecture that the 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 2.1. 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 heteroskedastic errors. An important example of linear regression models in the applied econometrics literature are long-horizon regressions. Such regression models have been applied for example in 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 by 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 $Y_t = X_t' \beta^o + \varepsilon_t$, where $X_t' = (X_{t1}, X_{t2})$ contains five regressors, the first one being a constant ($X_{t1} \equiv 1$). We consider two of the data

generating processes (DGP) 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 :

$$\begin{aligned} X_{ti} &= \rho X_{t-1,i} + \sqrt{1 - \rho^2} v_{ti}, \quad i = 2, \dots, 5; \\ \tilde{\varepsilon}_t &= \rho \tilde{\varepsilon}_{t-1} + \sqrt{1 - \rho^2} u_t, \end{aligned}$$

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 β^o equal to zero (without loss of generality) and consider the following values for the AR parameter ρ : 0.3, 0.5 and 0.9 and 0.95. The DGP AR(1)-HET2 is obtained from the AR(1)-HOMO model by introducing conditional heteroskedasticity in the errors ε_t . In particular, we let $\varepsilon_t = |X'_{t2} \gamma| \tilde{\varepsilon}_t$ with $\gamma = (0.5, 0.5, 0.5, 0.5)$. In the simulations, 5000 random samples are generated for the sample sizes $n \in \{64, 128, 256, 512, 1024\}$. 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{C}_n , a consistent estimator of the asymptotic variance-covariance matrix $C^o = A^{o-1} B^o A^{o-1}$. We consider two different choices of \hat{C}_n . Our first choice exploits the sandwich form of C^o and is given by $\hat{C}_{n,QS} = \hat{A}_n^{-1} \hat{B}_{n,QS} \hat{A}_n^{-1}$, where $\hat{A}_n = \frac{X'X}{n}$ and $\hat{B}_{n,QS}$ 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 :

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

A second choice of \hat{C}_n is $\tilde{C}_n^* = Var^* \left(\sqrt{n} \left(\tilde{\beta}_n^* - \hat{\beta}_n \right) \right)$, the bootstrap covariance matrix of the distribution of $\sqrt{n} \left(\tilde{\beta}_n^* - \hat{\beta}_n \right)$. Our Theorem 2.1 provides a formal justification for this choice. Here, $\tilde{\beta}_n^*$ is the truncated version of the LS estimator $\hat{\beta}_n^*$ which replaces $\hat{\beta}_n^*$ with $\hat{\beta}_n$ whenever $(X^{*'} X^*)^{-1}$ does not exist. As it turned out, for our Monte Carlo design we never encountered any singularity problems. Thus, in our simulations $\tilde{\beta}_n^* = \hat{\beta}_n^*$ and \tilde{C}_n^* coincides with $\hat{C}_n^* = Var^* \left(\sqrt{n} \left(\hat{\beta}_n^* - \hat{\beta}_n \right) \right)$. Notice that \tilde{C}_n^* does not rely on the sandwich form of 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 i^{th} bootstrap replication and $\overline{\tilde{\beta}_{2n}^*} = \frac{1}{B} \sum_{i=1}^B \tilde{\beta}_{2n}^{*(i)}$. When $B \rightarrow \infty$, $\tilde{C}_{n,22,B}^*$ 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

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

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

$$(3) \quad CI_{per-t} = \hat{\beta}_{2n} \pm q_{stud,0.95}^* \sqrt{\hat{C}_{n,22}},$$

where $q_{stud,0.95}^*$ is the 95% bootstrap percentile of the absolute value of the studentized bootstrap statistic

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

Here, $\overline{\tilde{C}_{n,22}^*}$ is a consistent estimator of the bootstrap population variance of $\sqrt{n}\tilde{\beta}_{2n}^*$. (Note the use of $\overline{\tilde{C}_{n,22}^*}$ rather than $\tilde{C}_{n,22}^*$ for reasons elaborated below.) 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 another for studentizing the t -statistic evaluated on the bootstrap data (cf. $\sqrt{\tilde{C}_{n,22}^*}$ 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 stu-

dentize the bootstrap statistic, Götze and Künsch (1996) suggest 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) use a kernel variance estimator with rectangular weights and warn that triangular weights will 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 above, namely $\hat{C}_{n,QS,22}$, which relies on the sandwich form of C^o and uses the QS-kernel to estimate 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 $\bar{C}_{n,22}^*$ be the element (2, 2) of $\bar{C}_n^* = \tilde{A}_n^{*-1} \tilde{B}_n^* \tilde{A}_n^{*-1}$, where $\tilde{A}_n^* = \frac{X^{*'} X^*}{n}$ and

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

where $\{I_i\}$ are i.i.d. random variables uniformly distributed on $\{0, 1, \dots, n - \ell\}$. 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:

$$CI_{per-t,QS} = \hat{\beta}_{2n} \pm q_{stud,0.95}^* \sqrt{\hat{C}_{n,QS,22}},$$

and

$$CI_{per-t,var^*} = \hat{\beta}_{2n} \pm q_{stud,0.95}^* \sqrt{\tilde{C}_{n,22,B}^*}.$$

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

by

$$CI_{per} = \hat{\beta}_{2n} \pm q_{0.95}^*,$$

where $q_{0.95}^*$ is the 95% bootstrap percentile of the absolute value of $\sqrt{n}(\tilde{\beta}_{2n}^* - \hat{\beta}_{2n})$. Contrary to the bootstrap percentile- t confidence interval, the bootstrap percentile confidence interval does not require any standard error estimate. However, since it is not based on an asymptotically pivotal statistic, this bootstrap method does not promise any asymptotic refinements.

Choice of the bandwidth for the QS-based confidence interval and of the block size for the MBB confidence intervals is critical. We use Andrews' (1991) automatic procedure based on approximating AR(1) models for the elements of $X_t\hat{\varepsilon}_t$ 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' automatic procedure for the Bartlett kernel. This choice is easy to implement and affords meaningful comparison of our results.

Figures 1-2 contain results for the DGP AR(1)-HOMO, whereas Figures 3-4 contain 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 and the undercoverage is worse the larger is ρ . One exception is CI_{per-t,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 $\rho = 0.9$, the coverage rate of CI_{QS} is 67.34% whereas that of CI_{var^*} is 79.06%. Since 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 improve significantly 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 as 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 $\rho = 0.3$ and $\rho = 0.5$), CI_{per-t,var^*} tends to overcover for the smaller sample sizes, whereas $CI_{per-t,QS}$ 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. $\rho = 0.9$ and $\rho = 0.95$), the undercoverage of $CI_{per-t,QS}$ worsens. In contrast, CI_{per-t,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 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 heteroskedasticity, i.e. for AR(1)-HET2, leads to smaller coverage rates for both bootstrap percentile- t confidence intervals, which results in worse undercoverage for $CI_{per-t,QS}$ and some undercoverage for CI_{per-t,var^*} . Nevertheless, here too replacing the QS-kernel standard error used to studentize the original t -statistic by the bootstrap standard error estimate helps to reduce the coverage error of bootstrap percentile- t confidence intervals.

4 Conclusions

This paper gives conditions under which the moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) provides consistent estimators of the asymptotic variance of the least squares estimator in (possibly misspecified) linear regression models. Although we have focused on the moving blocks bootstrap, similar results hold for the stationary bootstrap of Politis and Romano (1994) (see Gonçalves (2000)). The Monte Carlo results obtained in the paper 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

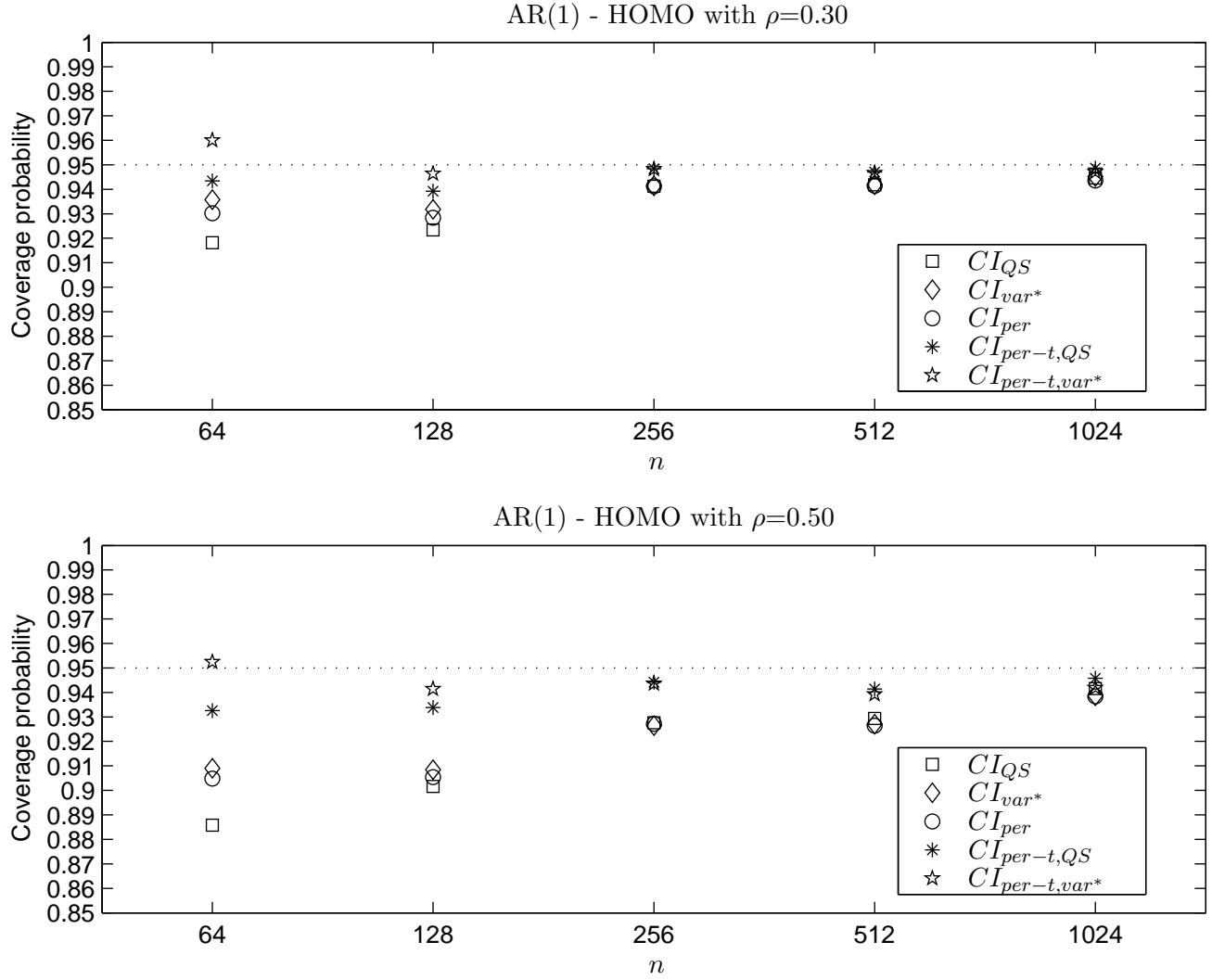


Figure 1: Coverage probabilities of 95% nominal symmetric confidence intervals

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.

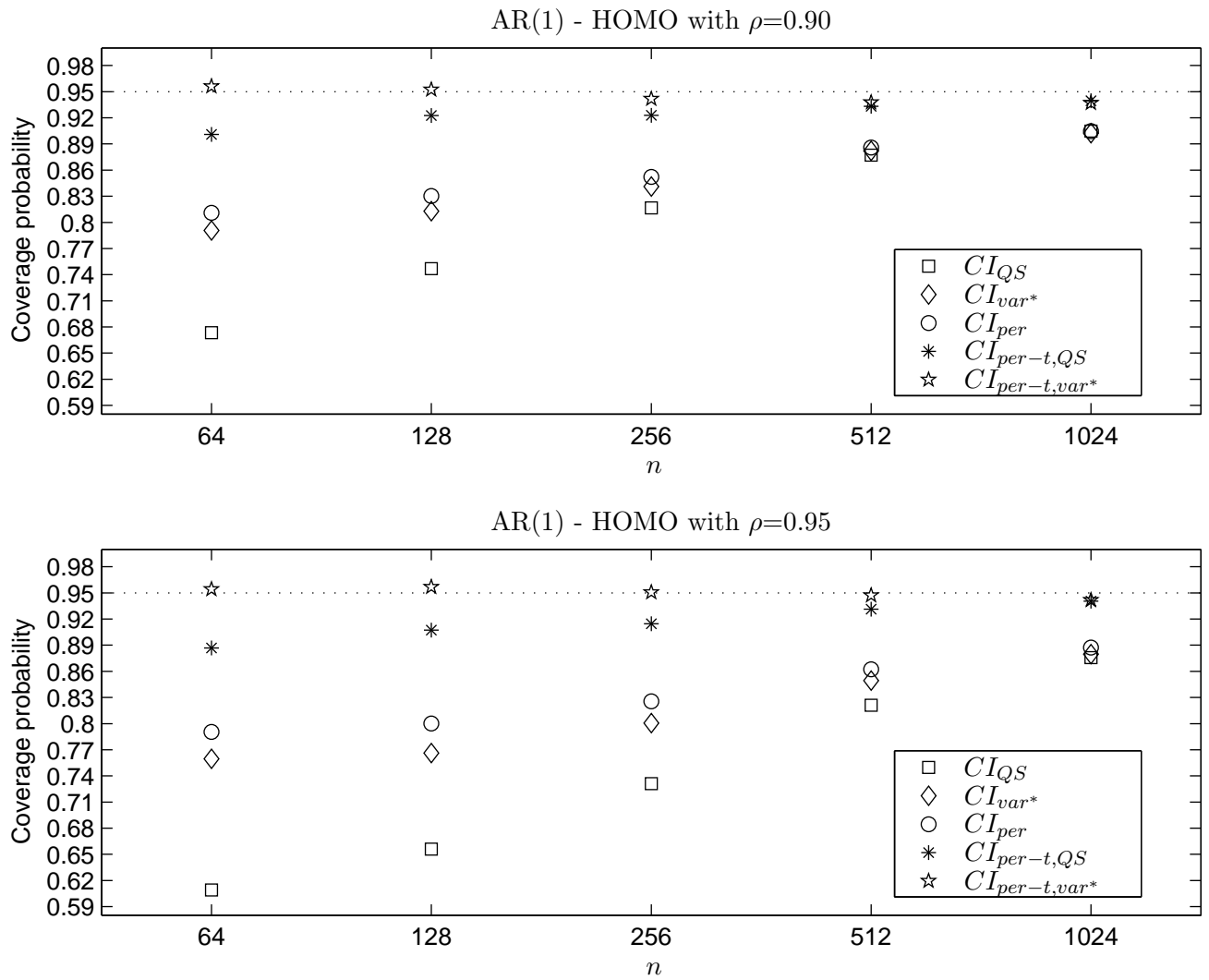


Figure 2: Coverage probabilities of 95% nominal symmetric confidence intervals

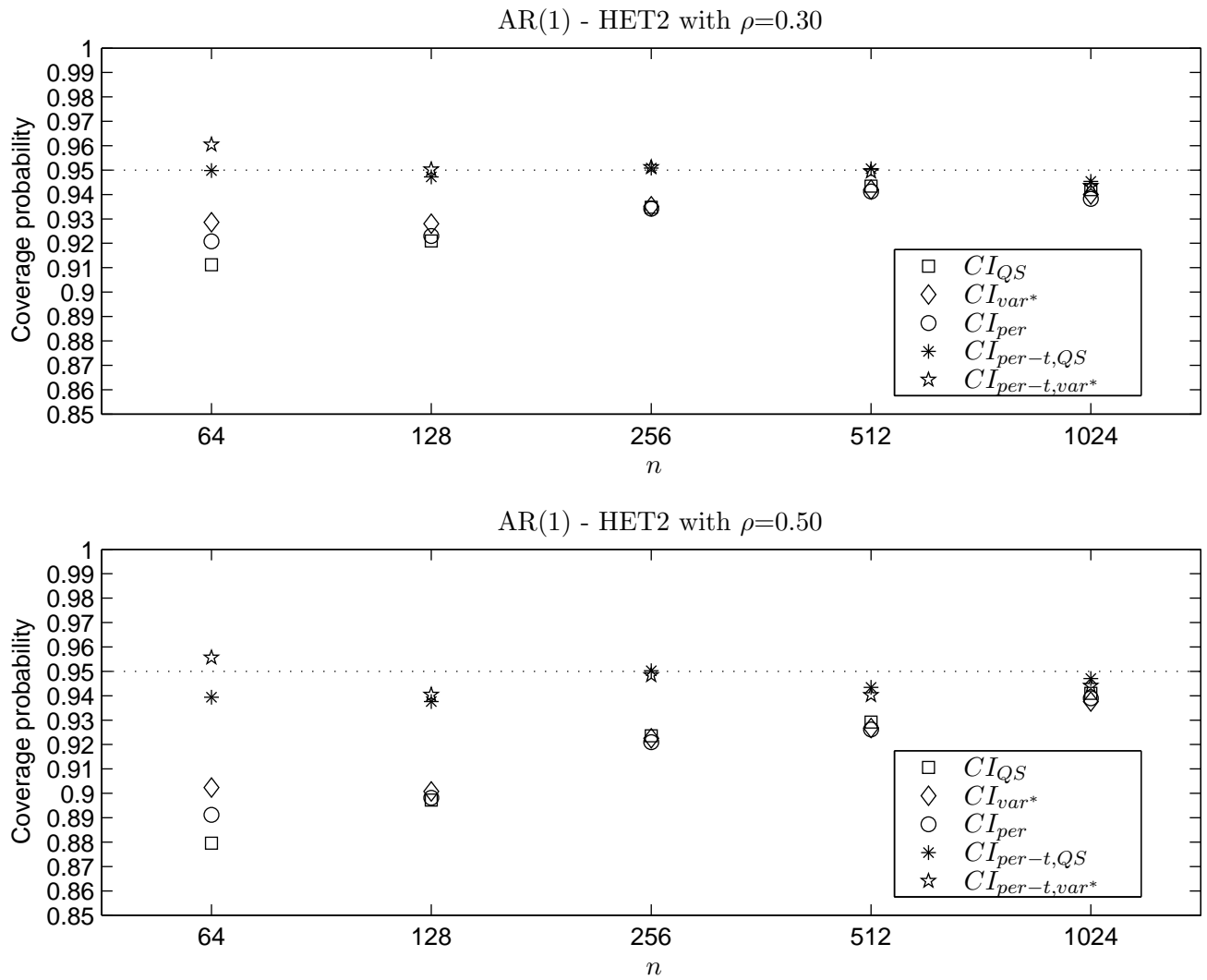


Figure 3: Coverage probabilities of 95% nominal symmetric confidence intervals

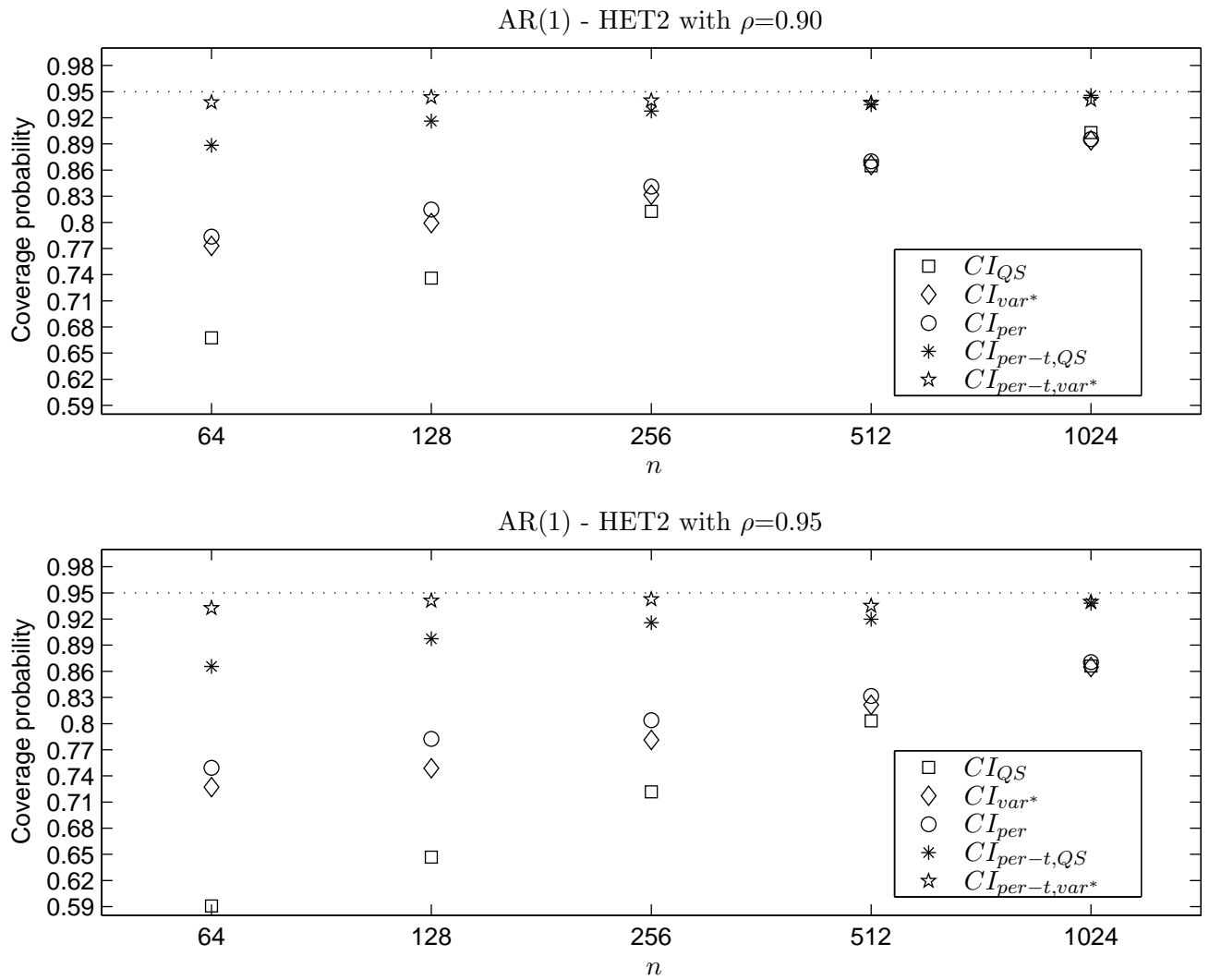


Figure 4: Coverage probabilities of 95% nominal symmetric confidence intervals

A Appendix

Throughout this appendix, C denotes a generic constant that does not depend on n . $1(A)$ 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 (Ω, \mathcal{G}, P) be a probability space on which a sequence of random variables $\{Z_t\}_{t=1}^\infty$ is defined, and let $\{\mathcal{G}^t\}_{t=1}^\infty$ be a nondecreasing sequence of sub σ -fields of \mathcal{G} . We say $\{Z_t, \mathcal{G}^t\}_{t=1}^\infty$ 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 $\psi_m \rightarrow 0$ as $m \rightarrow \infty$, and for all $t \geq 1$ and $m \geq 0$ we have: (a) $\|E(Z_t | \mathcal{G}^{t-m})\|_q \leq c_t \psi_m$, and (b) $\|Z_t - E(Z_t | \mathcal{G}^{t+m})\|_q \leq c_t \psi_{m+1}$. We will make use of the following result.

Lemma A.1 *For $q \geq 2$, 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 $\{Z_t : t = 1, \dots, n\}$ obtained with the MBB. If $\ell_n = o(n)$ with $\ell_n \rightarrow \infty$, 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 i.i.d. with $P^*(Y_{n,i} = Z_{j+1} + \dots + Z_{j+\ell_n}) = \frac{1}{n-\ell_n+1}$, $j = 0, \dots, n - \ell_n$. Hence,

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

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, $A_n \leq CE^* \left| \sum_{i=1}^k |Y_{n,i} - E^*(Y_{n,1})|^2 \right|^{q/2}$, and by Hölder's inequality,

$$(5) \quad E^* \left| \sum_{i=1}^k |Y_{n,i} - E^*(Y_{n,1})|^2 \right|^{q/2} \leq E^* \left| \left(\sum_{i=1}^k |Y_{n,i} - E^*(Y_{n,1})|^q \right)^{2/q} k^{1-2/q} \right|^{q/2} \leq 2^q k^{q/2} E^* |Y_{n,1}|^q,$$

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

$$(6) \quad E(E^* |Y_{n,1}|^q) = (n - \ell_n + 1)^{-1} \sum_{j=0}^{n-\ell_n} E \left| \sum_{t=1}^{\ell_n} Z_{t+j} \right|^q \leq (n - \ell_n + 1)^{-1} \sum_{j=0}^{n-\ell_n} (C \ell_n^{1/2})^q = (C)^q \ell_n^{q/2},$$

by a maximal inequality for mixingales (Hansen, Lemma 2, 1991; see also Hansen (1992)), and by the boundedness assumption on the mixingale constants $\{c_t\}$. Since $k = \frac{n}{\ell_n}$, it follows from (5) and

(6) 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

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

By the properties of the MBB, we can write

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

which implies that $E(B_n) = n^q E[|A_{n1} - A_{n2} - A_{n3}|^q] \leq 3^{q-1} n^q (E|A_{n1}|^q + E|A_{n2}|^q + E|A_{n3}|^q)$. 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, $E|A_{n3}|^q = O\left(\frac{\ell_n^q}{(n - \ell_n + 1)^q}\right)$. Hence, since $\ell_n = o(n)$, $E(B_n) \leq O(n^{q/2}) + O(\ell_n^q)$, completing the proof. ■

Proof of Theorem 2.1. 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

$$(7) \quad P \left[\omega : P_\omega^* \left(\lambda : \lambda_{\min} \left(\frac{X^{*'}(\lambda, \omega) X^*(\lambda, \omega)}{n} \right) < \eta/2 \right) > \xi \right] < \xi.$$

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 in Gonçalves and White (2004). Similarly, we write $X^*(\lambda, \omega)$ to emphasize the fact that for each $\omega \in \Omega$ and for $t = 1, 2, \dots, n$, we let $X_t^* = X_{\tau_t(\lambda)}(\omega)$, 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 $A_{n,\epsilon} \equiv \left\{ \omega : \left| \lambda_{\min} \left(\frac{X'(\omega)X(\omega)}{n} \right) - \lambda_{\min}(A^o) \right| \leq \epsilon \right\}$. Note that for any $\omega \in A_{n,\epsilon}$, $\lambda_{\min} \left(\frac{X'X}{n} \right) \geq \eta - \epsilon$, given that $\lambda_{\min}(A^o) > \eta > 0$, under Assumption 1.(v). Next, for any ω , define $C_{n,\omega,\epsilon} \equiv \left\{ \lambda : \left| \lambda_{\min} \left(\frac{X^{*'}(\lambda, \omega) X^*(\lambda, \omega)}{n} \right) - \lambda_{\min} \left(\frac{X'(\omega)X(\omega)}{n} \right) \right| \leq \epsilon \right\}$ and note that for $\omega \in A_{n,\epsilon}$, $C_{n,\omega,\epsilon}$ implies $B_{n,\omega,\epsilon} \equiv \left\{ \lambda : \lambda_{\min} \left(\frac{X^{*'}(\lambda, \omega) X^*(\lambda, \omega)}{n} \right) \geq \eta - 2\epsilon \right\}$. Thus, $A_{n,\epsilon} \cap C_{n,\omega,\epsilon} \subseteq B_{n,\omega,\epsilon}$, which implies $P_\omega^*(B_{n,\omega,\epsilon}^c) \leq P_\omega^*(A_{n,\epsilon}^c) + P_\omega^*(C_{n,\omega,\epsilon}^c)$. Choosing $\epsilon = \frac{\eta}{4}$, it follows that

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

where the last inequality holds because $P\left(P_\omega^*\left(A_{n,\frac{\eta}{4}}^c\right) > \xi/2\right) = P\left(A_{n,\frac{\eta}{4}}^c\right) < \xi/2$ for all n sufficiently large (by convergence of $\frac{X'X}{n}$ to A^o), and because $P\left(P_\omega^*\left(C_{n,\omega,\frac{\eta}{4}}^c\right) > \xi/2\right) < \frac{\xi}{2}$, for all n sufficiently large (by Lemma A.5 of Gonçalves and White (2004), given that $\ell_n = o(n)$). This proves (7).

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

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

$$\sqrt{n}\left(\tilde{\beta}_n^* - \hat{\beta}_n\right) = \left(\frac{X^{*'}X^*}{n}\right)^{-1}1\left(\lambda_{\min}\left(\frac{X^{*'}X^*}{n}\right) \geq \eta/2\right)n^{-1/2}\sum_{t=1}^n X_{nt}^*\hat{\varepsilon}_{nt}^*.$$

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

$$\left|\sqrt{n}\left(\tilde{\beta}_n^* - \hat{\beta}_n\right)\right|^{2+\delta} \leq \left\|\left(\frac{X^{*'}X^*}{n}\right)^{-1}\right\|_1^{2+\delta}1\left(\lambda_{\min}\left(\frac{X^{*'}X^*}{n}\right) \geq \eta/2\right)\left|n^{-1/2}\sum_{t=1}^n X_{nt}^*\hat{\varepsilon}_{nt}^*\right|^{2+\delta}.$$

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

$$\left\|\left(\frac{X^{*'}X^*}{n}\right)^{-1}\right\|_1 = \lambda_{\max}\left(\left(\frac{X^{*'}X^*}{n}\right)^{-1}\right) = \lambda_{\min}^{-1}\left(\frac{X^{*'}X^*}{n}\right) \leq (\eta/2)^{-1} = C.$$

Thus,

$$\left|\sqrt{n}\left(\tilde{\beta}_n^* - \hat{\beta}_n\right)\right|^{2+\delta} \leq C\left|n^{-1/2}\sum_{t=1}^n X_{nt}^*\hat{\varepsilon}_{nt}^*\right|^{2+\delta},$$

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

$$\left| n^{-1/2} \sum_{t=1}^n X_{nt}^* \hat{\varepsilon}_{nt}^* \right|^{2+\delta} \leq 2^{1+\delta} \left(\left| n^{-1/2} \sum_{t=1}^n X_{nt}^* \varepsilon_{nt}^* \right|^{2+\delta} + \left| \left(\frac{X^{*'} X^*}{n} \right) \sqrt{n} (\hat{\beta}_n - \beta^o) \right|^{2+\delta} \right) \equiv C (A_1^* + A_2^*).$$

By Lemma A.1, we can show that

$$E(E^*(A_1^*)) \leq C n^{-\frac{2+\delta}{2}} E \left[E^* \left(\left| \sum_{t=1}^n X_{nt}^* \varepsilon_{nt}^* \right|^{2+\delta} \right) \right] = O(1) + O \left(\left(\frac{\ell_n^2}{n} \right)^{\frac{2+\delta}{2}} \right),$$

which is $O(1)$ since $\ell_n^2/n \rightarrow 0$. To apply Lemma A.1, we need $\{X_t \varepsilon_t\}$ to be a zero-mean $L_{2+\delta}$ -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

$$A_2^* \leq \left\| \left(\frac{X^{*'} X^*}{n} \right) \right\|_1^{2+\delta} \left| \sqrt{n} (\hat{\beta}_n - \beta^o) \right|^{2+\delta} = \lambda_{\max}^{2+\delta} \left(\frac{X^{*'} X^*}{n} \right) \left| \sqrt{n} (\hat{\beta}_n - \beta^o) \right|^{2+\delta},$$

implying that

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

Since $\sqrt{n} (\hat{\beta}_n - \beta^o)$ converges in distribution it follows that $\left| \sqrt{n} (\hat{\beta}_n - \beta^o) \right|^{2+\delta} = O_P(1)$. Thus, to prove that $E^*(A_2^*) = O_P(1)$, it suffices that $E^* \left(\lambda_{\max}^{2+\delta} \left(\frac{X^{*'} X^*}{n} \right) \right) = O_P(1)$. For this, note that $0 < \lambda_{\max} \left(\frac{X^{*'} X^*}{n} \right) \leq \text{tr} \left(\frac{X^{*'} X^*}{n} \right) = \sum_{i=1}^p (n^{-1} \sum_{t=1}^n X_{ti}^{*2})$, where, for any matrix A , $\lambda_i(A)$ denotes its i^{th} eigenvalue and $\text{tr}(A)$ denotes its trace. Thus,

$$(8) \quad E^* \left(\lambda_{\max}^{2+\delta} \left(\frac{X^{*'} X^*}{n} \right) \right) \leq E^* \left[\left(\text{tr} \left(\frac{X^{*'} X^*}{n} \right) \right)^{2+\delta} \right] \leq C \sum_{i=1}^p n^{-(2+\delta)} E^* \left(\left| \sum_{t=1}^n X_{ti}^{*2} \right|^{2+\delta} \right),$$

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

$$(9) \quad E^* \left(\left| \sum_{t=1}^n X_{ti}^{*2} \right|^{2+\delta} \right) = E^* \left(\left| \sum_{t=1}^n (X_{ti}^{*2} - \mu_{2i}) + n\mu_{2i} \right|^{2+\delta} \right) \leq C E^* \left(\left| \sum_{t=1}^n W_{ti}^* \right|^{2+\delta} \right) + \mu_{2i}^{2+\delta} n^{2+\delta},$$

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

$$(10) \quad E \left[E^* \left(\left| \sum_{t=1}^n W_{ti}^* \right|^{2+\delta} \right) \right] = O \left(n^{\frac{2+\delta}{2}} \right) + O \left(\ell_n^{2+\delta} \right).$$

From (8), (9) and (10), it follows that

$$E \left[E^* \left(\lambda_{\max}^{2+\delta} \left(\frac{X^{*'} X^*}{n} \right) \right) \right] = O \left(n^{-\frac{2+\delta}{2}} \right) + O \left(\left(\frac{\ell_n}{n} \right)^{2+\delta} \right) + O(1),$$

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

References

- [1] Andrews, D.W.K., 1984. "Non-strong mixing autoregressive processes," *Journal of Applied Probability* 21, 930-934.
- [2] Andrews, D.W.K. 1988. "Laws of large numbers for dependent nonidentically distributed random variables," *Econometric Theory* 4, 458-467.
- [3] Andrews, D.W.K., 1991. "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* 59, 817-858.
- [4] Billingsley, P., 1968. *Convergence of probability measures*. Wiley, New York.
- [5] Billingsley, P., 1995. *Probability and measure*, Wiley, New York.
- [6] Buchinsky, M., 1995. "Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study," *Journal of Econometrics*, 68, 303-338.
- [7] Davidson, J. , 1994. *Stochastic limit theory*, Oxford University Press, Oxford.
- [8] Davidson, J., 2002. "Establishing conditions for the functional central limit theorem in non-linear and semiparametric time series processes," *Journal of Econometrics* 106, 243-269.
- [9] Davidson, R, and MacKinnon, J.G., 1999. "Bootstrap testing in nonlinear models," *International Economic Review* 40, 487-508.
- [10] 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.
- [11] Efron, B., 1979. "Bootstrap methods: another look at the jackknife," *Annals of Statistics* 7, 1-26.
- [12] 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.
- [13] Fitzenberger, B., 1997. "The moving blocks bootstrap and robust inference for linear least squares and quantile regressions," *Journal of Econometrics* 82, 235-287.
- [14] Freedman, D.A. and Peters, S.C., 1984. "Bootstrapping a regression equation: some empirical results," *Journal of the American Statistical Association*, 79, 97-106.
- [15] Gallant, A.R., and White, H., 1988. *A unified theory of estimation and inference for nonlinear dynamic models*, Basil Blackwell, New York.
- [16] Ghosh, M., Parr, W.C., Singh, K., and Babu, G.J., 1984. "A note on bootstrapping the sample median," *Annals of Statistics* 12, 1130-1135.

- [17] Gonçalves, S., 2000. The bootstrap for dependent heterogeneous processes, Ph.D. dissertation, UCSD.
- [18] Gonçalves, S., and White, H., 2002. "The bootstrap of the mean for dependent heterogeneous arrays," *Econometric Theory*, 18, 1367-1384.
- [19] Gonçalves, S., and White, H., 2004. "Maximum likelihood and the bootstrap for nonlinear dynamic models," *Journal of Econometrics*, 2004, 119, 199-219.
- [20] Götze, F. and Künsch, H.R., 1996. "Second-order correctness of the blockwise bootstrap for stationary observations," *Annals of Statistics*, 24, 1914-1933.
- [21] Hahn, J., 1995. "Bootstrapping quantile regression estimators," *Econometric Theory*, 11, 105-121.
- [22] Hansen, B.E., 1991. "Strong laws for dependent heterogeneous processes," *Econometric Theory* 7, 213-221.
- [23] Hansen, B.E., 1992. "Erratum", *Econometric Theory* 8, 421-422.
- [24] Kilian, L., 1999. "Exchange rates and monetary fundamentals: what do we learn from long horizon regressions?," *Journal of Applied Econometrics*, 14, 491-510.
- [25] Kirby, C., 1998. "The restrictions on predictability implied by rational asset pricing models," *Review of Financial Studies*, 11, 343-82.
- [26] Künsch, H.R., 1989. "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics* 17, 1217-1241.
- [27] Liu, R.Y., Singh, K., 1992a. "Moving blocks jackknife and bootstrap capture weak dependence," in: LePage and Billiard (Eds), *Exploring the limits of the bootstrap*, Wiley, New York.
- [28] Liu, R.Y., Singh, K., 1992b. "Efficiency and robustness in resampling," *Annals of Statistics* 20, 370-384.
- [29] 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.
- [30] Mark, N., 1995. "Exchange rates and fundamentals: evidence on long run predictability," *American Economic Review* 85, 201-218.
- [31] McLeish, D.L., 1975. "A maximal inequality and dependent strong laws," *Annals of Probability* 5, 616-621.
- [32] Politis, D. and J. Romano, 1994. The stationary bootstrap, *Journal of American Statistical Association* 89, 1303-1313.
- [33] Politis, D., Romano, J. and Wolf, M., 1997. "Subsampling for heteroskedastic time series," *Journal of Econometrics* 81, 281-317.
- [34] Romano, J. and Wolf, M., 2003. "Improved nonparametric confidence intervals in time series regressions," mimeo, Department of Economics and Business, University of Pompeu Fabra.
- [35] Shao, J., 1992. "Bootstrap variance estimators with truncation," *Statistics and Probability Letters* 15, 95-101.

- [36] Shao, J., and Tu, D., 1995. *The jackknife and bootstrap*, Springer-Verlag, New York.
- [37] Strang, G., 1988. *Linear algebra and its applications*. Third Edition, Harcourt, Orlando.