



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Econometrics 133 (2006) 531–555

JOURNAL OF
Econometrics

www.elsevier.com/locate/jeconom

Bootstrapping GMM estimators for time series

Atsushi Inoue^{a,*}, Mototsugu Shintani^b

^a*Department of Agricultural and Resource Economics, North Carolina State University,
Raleigh, NC 27695-8109, USA*

^b*Department of Economics, Vanderbilt University, Nashville, TN 37235, USA*

Available online 10 August 2005

Abstract

This paper considers the bootstrap for the GMM estimator of overidentified linear models when autocorrelation structures of moment functions are unknown. When moment functions are uncorrelated after finite lags, Hall and Horowitz, [1996. Bootstrap critical values for tests based on generalized method of moments estimators. *Econometrica* 64, 891–916] showed that errors in the rejection probabilities of the bootstrap tests are $o(T^{-1})$. However, this rate cannot be obtained with the HAC covariance matrix estimator since it converges at a nonparametric rate. By incorporating the HAC covariance matrix estimator in the Edgeworth expansion of the distribution, we show that the bootstrap provides asymptotic refinements when the characteristic exponent of the kernel function is greater than two.

© 2005 Elsevier B.V. All rights reserved.

JEL classification: C12; C22; C32

Keywords: Asymptotic refinements; Block bootstrap; Dependent data; Edgeworth expansions; Instrumental variables

1. Introduction

In this paper we establish that the bootstrap provides asymptotic refinements for the two-step generalized method of moments (GMM) estimator of possibly

*Corresponding author. Tel.: +1 919 515 5969; fax: +1 919 515 1824.

E-mail addresses: atsushi@unity.ncsu.edu (A. Inoue), mototsugu.shintani@vanderbilt.edu (M. Shintani).

overidentified linear models. Our analysis differs from earlier work by Hall and Horowitz (1996) and Andrews (2002a) in that we allow correlation of moment functions beyond finitely many lags. In typical empirical applications, the autocovariance structure of moment functions is unknown and the inverse of the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator is often used as a weighting matrix in GMM estimation. In finite samples, however, it is well-known that coverage probabilities based on the HAC covariance estimator are often too low, and that the t test tends to reject too frequently (see Andrews, 1991). In this paper, we consider a bootstrap method for the GMM estimator for the purpose of improving the finite sample performance of the t test and Hansen's (1982) test of overidentifying restrictions (J test).

We use the block bootstrap originally proposed by Künsch (1989) for weakly dependent data (see also Carlstein, 1986). When the block length increases at a suitable rate with the sample size, such block bootstrap procedures eventually will capture the unknown structure of dependence. The block bootstrap has been applied to the GMM estimation by Hall and Horowitz (1996) and Andrews (2002a) and thus our analysis is largely related to their work. Hall and Horowitz (1996) show that the nonoverlapping block bootstrap provides asymptotic refinements for GMM for a certain class of dependent data.¹ To be more specific, for a series with the moment function that is uncorrelated after finitely many lags, errors in the symmetrical distribution function are $o(T^{-1})$ and errors in the rejection probabilities (ERPs) of symmetrical tests are $o(T^{-1})$. Thus, when the moment function has such a dependence structure, asymptotic refinements can be obtained with a parametric rate. Andrews (2002a) establishes the higher-order equivalence of the k -step bootstrap, which is computationally attractive, and the standard bootstrap for statistics based on nonlinear extremum estimators including GMM. While Andrews' (2002a) results generalize those of Hall and Horowitz (1996) in the sense that they cover both nonoverlapping and overlapping bootstraps, the series are still required to be uncorrelated after some finite lags.

Economic theory often provides information about the specification of moment conditions, but not necessarily about the dependence structure of the moment conditions. Therefore, it is of practical interest to know whether the asymptotic refinement can be provided with the bootstrap procedure based on the HAC covariance matrix estimator designed for a series with more general forms of autocorrelation. To prove an asymptotic refinement, we follow a conventional approach and rely on Edgeworth expansions (see Hall, 1992, for the Edgeworth expansion view of the bootstrap). However, because the HAC covariance matrix estimator cannot be written as a function of sample moments and converges at a nonparametric rate that is slower than $T^{-1/2}$, the widely used result of Götze and Hipp (1983) cannot be directly applied to establish the existence of Edgeworth expansions.

¹For independent series, Hahn (1996) shows the first-order validity of the bootstrap for GMM. Brown and Newey (1995) also propose an alternative efficient bootstrap method based on the empirical likelihood.

Recent studies by Götze and Künsch (1996) and Lahiri (1996) contain some results on Edgeworth expansions that can be applied to HAC covariance matrix estimation. Using such Edgeworth expansions, Götze and Künsch (1996) and Lahiri (1996) show that the block bootstrap can provide asymptotic refinements for a smooth function of sample means and for estimators in a linear regression model, respectively. However, their results can only cover the refinement in the coverage probabilities of the one-sided confidence intervals or the ERPs of the one-sided test.

Since typical analyses based on GMM estimator employ the two-sided confidence intervals and the J test, we prove that the bootstrap provides asymptotic refinements for these statistics in overidentified linear models estimated by GMM. To our knowledge, the higher-order properties of the block bootstrap for GMM with unknown autocovariance structures have not been formally investigated.²

We show that the order of both the first-order and bootstrap approximation error depends on the asymptotic bias of the HAC covariance estimator. For the bootstrap to provide asymptotic refinements for symmetric confidence intervals and for the J test statistic, one must therefore use kernels, such as the truncated kernel (White, 1984), the trapezoidal kernel (Politis and Romano, 1995) and the Parzen (b) kernel, of which the asymptotic bias vanishes quickly enough. Hall and Horowitz (1996) and Andrews (2002a) consider the truncated kernel only. This is also in contrast to the results of Götze and Künsch (1996) who show that, for all kernels but the Bartlett kernel, the bootstrap provides asymptotic refinements for one-sided confidence intervals. For the symmetric confidence interval and the J test statistic, $O(T^{-1/2})$ terms do not exist in the Edgeworth expansion because these terms are even functions and are canceled out. However, the asymptotic bias of the HAC covariance matrix estimator is deterministic and is not an even function. Thus, the bias becomes important in our higher-order asymptotics.

While we restrict our attention to linear models, they are of particular interest in empirical macroeconomics. GMM estimation of linear models has been applied to the expectation hypothesis of the term structure (Campbell and Shiller, 1991), the monetary policy reaction function (Clarida et al., 2000), the permanent-income hypothesis (Runkle, 1991), and the present value model of stock prices (West, 1988). To evaluate the performance of our bootstrap procedure in finite samples, we conduct a small Monte Carlo experiment. As an empirical example, we apply our bootstrap procedure to the monetary policy reaction function estimated by Clarida et al. (2000). Since the GMM estimates often have policy implications in structural econometric models, it is important for researchers to obtain accurate confidence intervals.

The remainder of the paper is organized as follows. Section 2 introduces the model and describes the proposed bootstrap procedure. Section 3 presents the assumptions and theoretical results. Section 4 provides some Monte Carlo results. Section 5

²Recent related studies include Hansen (2000) and Gonçalves and White (2004). Hansen (2000) considers the Edgeworth expansion of test statistics for nonlinear restrictions in the GMM framework but rules out the use of HAC covariance matrix estimators. Gonçalves and White (2004) establish the first-order validity of the bootstrap for nonlinear dynamic models under very general conditions.

presents an empirical illustration. Section 6 concludes the analysis. Proofs of the theorems are relegated to an appendix.

2. Model and bootstrap procedure

Consider a stationary time series $(x'_t, y_t, z'_t)'$ which satisfies

$$E[z_t u_t] = 0, \tag{1}$$

where $u_t = y_t - \beta'_0 x_t$, β_0 is a p -dimensional parameter, x_t is a p -dimensional vector, z_t is a k -dimensional vector which may contain lagged values of y_t , and $p < k$. Given T_0 observations $\{(x'_t, y_t, z'_t)'\}_{t=1}^{T_0}$, we are interested in two-step GMM estimation of β_0 based on the moment condition (1). Let ℓ denote the lag truncation parameter used in HAC covariance matrix estimation and $T = T_0 - \ell + 1$.³ We first obtain the first-step GMM estimator $\tilde{\beta}_T$ by minimizing

$$\left[\frac{1}{T_0} \sum_{t=1}^{T_0} z_t (y_t - \beta' x_t) \right]' V_T \left[\frac{1}{T_0} \sum_{t=1}^{T_0} z_t (y_t - \beta' x_t) \right]$$

with respect to β , where V_T is some $k \times k$ positive semidefinite matrix. Then we obtain the second-step GMM estimator $\hat{\beta}_T$ by minimizing

$$\left[\frac{1}{T} \sum_{t=1}^T z_t (y_t - \beta' x_t) \right]' \hat{S}_T^{-1} \left[\frac{1}{T} \sum_{t=1}^T z_t (y_t - \beta' x_t) \right],$$

where

$$\begin{aligned} \hat{S}_T &= \frac{1}{T} \sum_{t=1}^T \left[z_t \hat{u}_t^2 z'_t + \sum_{j=1}^{\ell} \omega\left(\frac{j}{\ell}\right) (z_{t+j} \hat{u}_{t+j} \hat{u}_t z'_t + z_t \hat{u}_t \hat{u}_{t+j} z'_{t+j}) \right] \\ &= \sum_{j=-\ell+1}^{\ell-1} \omega(j/\ell) \hat{\Gamma}_j \end{aligned}$$

is the HAC covariance matrix estimator for the moment function (1), $\hat{u}_t = y_t - \tilde{\beta}'_T x_t$, and $\omega(\cdot)$ is a kernel. Following [Rothenberg \(1984\)](#), we consider the distribution of the studentized statistic of a linear combination of the parameter, $T^{1/2}(c' \hat{\Sigma}_T c)^{-1/2} c' (\hat{\beta}_T - \beta_0)$ where $\hat{\Sigma}_T = ((1/T) \sum_{t=1}^T x_t z'_t \hat{S}_T^{-1} (1/T) \sum_{t=1}^T z_t x'_t)^{-1}$ and c is an arbitrary nonzero p -dimensional vector. We also consider the distribution of the J test statistic

$$J_T = \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t (y_t - \tilde{\beta}'_T x_t) \right]' \hat{S}_T^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t (y_t - \tilde{\beta}'_T x_t) \right].$$

³We use T observations and the modified HAC covariance matrix estimator \hat{S}_T to obtain asymptotic refinements for symmetric confidence intervals and the J test statistic. This modification is not necessary for obtaining asymptotic refinements of one-sided confidence intervals. See also [Hall and Horowitz \(1996, p. 895\)](#).

We use the following overlapping block bootstrap procedure, which is originally attributed to [Künsch \(1989\)](#). Suppose that $T = b\ell$ for some integer b .

Step 1: Let N_1, N_2, \dots, N_b be iid uniform random variables on $\{0, 1, \dots, T - \ell\}$ and let

$$(x_{(j-1)\ell+i}^*, y_{(j-1)\ell+i}^*, z_{(j-1)\ell+i}^*)' = (x'_{N_j+i}, y_{N_j+i}, z'_{N_j+i})'$$

for $1 \leq i \leq \ell$ and $1 \leq j \leq b$.

Step 2: Calculate the first-step bootstrap GMM estimator $\tilde{\beta}_T^*$ by minimizing

$$\left[\frac{1}{T} \sum_{t=1}^T z_t^*(y_t^* - \beta' x_t^*) - \mu_T^* \right]' V_T \left[\frac{1}{T} \sum_{t=1}^T z_t^*(y_t^* - \beta' x_t^*) - \mu_T^* \right],$$

where

$$\mu_T^* = \frac{1}{T - \ell + 1} \sum_{t=0}^{T-\ell} \frac{1}{\ell} \sum_{i=1}^{\ell} z_{t+i}(y_{t+i} - \hat{\beta}'_T x_{t+i}).$$

Step 3: Compute the second-step bootstrap GMM estimator $\hat{\beta}_T^*$ by minimizing

$$\left[\frac{1}{T} \sum_{t=1}^T z_t^*(y_t^* - \beta' x_t^*) - \mu_T^* \right]' \hat{S}_T^{*-1} \left[\frac{1}{T} \sum_{t=1}^T z_t^*(y_t^* - \beta' x_t^*) - \mu_T^* \right],$$

where

$$\hat{S}_T^* = \frac{1}{T} \sum_{k=1}^b \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (z_{N_k+i} \hat{u}_{N_k+i}^* - \mu_T^*) (z_{N_k+j} \hat{u}_{N_k+j}^* - \mu_T^*)',$$

$$\hat{u}_t^* = y_t - \tilde{\beta}_T^{*'} x_t.$$

Step 4: Obtain the bootstrap version of the studentized statistic $T^{1/2} (c' \hat{\Sigma}_T^* c)^{-1/2} c' (\hat{\beta}_T^* - \beta_T)$, where $\hat{\Sigma}_T^* = ((1/T) \sum_{t=1}^T x_t^* z_t^{*'} \hat{S}_T^{*-1} (1/T) \sum_{t=1}^T z_t^* x_t^{*'})^{-1}$ and the J test statistic

$$J_T^* = \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^T [z_t^*(y_t^* - \hat{\beta}_T^{*'} x_t^*) - \mu_T^*] \right\}' \hat{S}_T^{*-1} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^T [z_t^*(y_t^* - \hat{\beta}_T^{*'} x_t^*) - \mu_T^*] \right\}.$$

By repeating Steps 1–4 sufficiently many times, one can approximate the finite-sample distributions of the studentized statistic and the J test statistic by the empirical distributions of their bootstrap version.

Remarks. 1. [Davison and Hall \(1993\)](#) show that naïve applications of the block bootstrap do not provide asymptotic refinements for studentized statistics involving the long-run variance estimator. Specifically, they show that the error of the naïve bootstrap is of order $O(b^{-1}) + O(\ell^{-1})$ and thus is greater than or equal to the error of the first-order asymptotic approximation. We therefore modify the bootstrap version of the HAC covariance matrix estimator in Step 3 (see [Götze and Künsch, 1996](#), for the just-identified case). Our treatment differs from that of [Hall and Horowitz \(1996\)](#)

and Andrews (2002a) who modify the test statistics instead of the GMM criterion function. The expression \hat{S}_T^* given in Step 3 is a consistent estimator for the variance of the bootstrapped moment function with the bootstrap probability measure.

2. One can consider choosing the HAC truncation parameter and the block length differently. Let ℓ denote the HAC truncation parameter and ℓ^* denote the block length. If $\ell^* = c\ell$ for some $c > 0$, our asymptotic refinement results will carry through. Otherwise, the rate of the bootstrap approximation error will be dominated by the faster rate of ℓ and ℓ^* , i.e., $o(\max(\ell, \ell^*)/T)$, as implied by our theory in Section 3 when the truncated kernel is used. In this case, it is thus optimal to set $\ell^* = c\ell$ for some $c > 0$ in terms of asymptotic refinements.⁴ However, the optimal value of c is indeterminate because our asymptotic refinement result is given only in terms of the rate. Following Götze and Künsch (1996), we simply set the block length equal to the HAC truncation parameter.

3. Asymptotic theory

In this section, we present our main theoretical results. Unless otherwise noted, we shall denote the Euclidean norm of a vector x by $\|x\|$. First, we provide the following set of assumptions.

Assumption 1.

- (a) $\{(x'_t, y_t, z'_t)'\}$ is strictly stationary and strong mixing with mixing coefficients satisfying $\alpha_m \leq (1/d)\exp(-dm)$ for some $d > 0$.
- (b) There is a unique $\beta_0 \in \mathfrak{R}^p$ such that $E[z_t u_t] = E[z_t(y_t - \beta'_0 x_t)] = 0$.
- (c) Let $R_t = ((z_t u_t)', \text{vec}(z_t x'_t)')'$. Then $E\|R_t\|^{r+\eta} < \infty$ for $r \geq 12$ and some $\eta > 0$.
- (d) Let \mathcal{F}_a^b denote the sigma-algebra generated by R_a, R_{a+1}, \dots, R_b . For all $m, s, t = 1, 2, \dots$ and $A \in \mathcal{F}_{t-s}^{t+s}$,

$$E\{P(A|\mathcal{F}_{-\infty}^{t-1} \cup \mathcal{F}_{t+1}^\infty) - P(A|\mathcal{F}_{t-s-m}^{t-1} \cup \mathcal{F}_{t+1}^{t+s+m})\} \leq (1/d)\exp(-dm).$$

- (e) For all $m, t = 1, 2, \dots$ and $\theta \in \mathfrak{R}^{p+k+1}$ such that $1/d < m < t$ and $|\theta| \geq d$,

$$E\left\{E\left[\exp\left[i\theta' \sum_{s=t-m}^{t+m} (R_s - E(R_s))\right] \middle| \mathcal{F}_{-\infty}^{t-1} \cup \mathcal{F}_{t+1}^\infty\right]\right\} \leq \exp(-d).$$

- (f) $\omega : \mathfrak{R} \rightarrow [-1, 1]$ satisfies (i) $\omega(0) = 1$, (ii) $\omega(x) = \omega(-x) \forall x \in \mathfrak{R}$, (iii) $\omega(x) = 0 \forall |x| \geq 1$, (iv) $\omega(\cdot)$ is continuous at 0 and at all but a finite number of other points.
- (g) $\ell T^{-2/r} \rightarrow \infty$ and $\ell = O(T^{1/3})$ as $T \rightarrow \infty$.

⁴Suppose $\ell/\ell^* \rightarrow \infty$ [$\ell^*/\ell \rightarrow \infty$]. Then choosing a smaller divergence rate of ℓ^* [resp. ℓ] will improve the order of the approximation error.

- (h) \hat{S}_T is a positive semidefinite matrix that converges in probability to a positive definite matrix $S_0 \equiv \sum_{j=-\infty}^{\infty} E(z_0 u_0 u_j z_j')$.
- (i) The first-step estimator $\tilde{\beta}_T$ satisfies $E|T^{1/2}(\tilde{\beta}_T - \beta_0)|^r = O(1)$, and V_T is a positive semidefinite matrix that converges to a positive definite matrix V at rate $O(\ell^{1/2} T^{-1/2})$.

Remarks. Assumption 1(c) requires that at least the 12th moment of the moment function be finite, and we will later require that at least the 36th moment be finite. Although this condition is strong, it is not atypical in the literature on higher-order asymptotic theory. For example, a sufficient (but not necessary) condition for Assumptions 3(f) and 4 of Hall and Horowitz (1996) is the finiteness of the 33rd moment of the moment functions and of their derivatives. Assumptions 1(d) and 1(e) are from Götze and Künsch (1996). Hall and Horowitz (1996, Assumptions 1 and 6) impose similar assumptions. Assumption 1(f) is a subset of Andrews' (1991) class of kernels \mathcal{K}_1 . For example, the truncated kernel (White, 1984), Bartlett kernel (Newey and West, 1987), and Parzen kernel (Gallant, 1987) satisfy Assumption 1(f).⁵ The range of divergence rates of ℓ allowed in Assumption 1(g) is more restrictive than the one typically assumed in the literature on HAC covariance matrix estimation (e.g., Theorem 1 of Andrews, 1991) but is less restrictive than the one Hall and Horowitz (1996) assumed for the divergence rate of the block length. The lower bound reflects the trade-off between dependence and moment conditions. The weaker the moment conditions are, the more dependent the process can be, and thus the block length must be larger. Andrews (2002a) does not require such a lower bound because he assumes that all moments exist in his Assumption 2(b). For further comparison with Hall and Horowitz (1996), see Remark 3 following Corollary 1. While the \sqrt{T} -consistency of the first-step estimator is sufficient for the first-order asymptotic theory (e.g., Assumption B(i) of Andrews, 1991), further conditions provided in Assumption 1(i) are required for the higher-order analysis.

We next present our main results. Let q denote the characteristic exponent of the kernel ω . That is, q is the largest real number such that $\lim_{x \rightarrow 0} (1 - \omega(x))/|x|^q \in [0, \infty)$.

Theorem 1. *Suppose that Assumption 1 holds. Let*

$$\Psi_T(x) = \Phi(x) + T^{-1/2} p_1(x) \phi(x) + \ell T^{-1} p_2(x) \phi(x),$$

$$\Psi_{J,T}(x) = F_{\gamma_{k-p}^2}(x) + \ell T^{-1} p_J(x) f_{\gamma_{k-p}^2}(x),$$

denote the Edgeworth expansions of $P(T^{1/2}(c' \hat{\Sigma}_T c)^{-1/2} c'(\hat{\beta}_T - \beta_0) \leq x)$ and $P(J_T \leq x)$, respectively, where $\Phi(x)$ and $\phi(x)$ denote the standard normal distribution and density

⁵Our proofs depend on the assumption that the lag order greater than or equal to ℓ receive zero weight. We do not know whether the bootstrap provides asymptotic refinements for one-sided confidence intervals when the quadratic spectral kernel (Andrews, 1991) is used. The bootstrap does not provide asymptotic refinements for symmetric confidence intervals and the J test statistic when this kernel is used as its characteristic exponent is two.

functions, $F_{\chi^2_{k-p}}$ and $f_{\chi^2_{k-p}}$ are the distribution and density functions of a χ^2 random variable with degree of freedom $k - p$, p_1 is even, and p_2 and p_J are odd. Then

$$\sup_{x \in \mathfrak{R}} |\mathbf{P}(T^{1/2}(c' \hat{\Sigma}_T c)^{-1/2} c' (\hat{\beta}_T - \beta_0) \leq x) - \Psi_T(x)| = o(\ell T^{-1}) + O(\ell^{-q}), \tag{2}$$

$$\sup_{x \geq 0} |\mathbf{P}(J_T \leq x) - \Psi_{J,T}(x)| = o(\ell T^{-1}) + O(\ell^{-q}). \tag{3}$$

Theorem 2. Suppose that Assumption 1 holds with $r \geq 12$ replaced by $r \geq 36$.⁶ Let

$$\Psi_T^*(x) = \Phi(x) + p_1^*(x)\phi(x) + \ell T^{-1} p_2^*(x)\phi(x),$$

$$\Psi_{J,T}^*(x) = F_{\chi^2_{q-p}}(x) + \ell T^{-1} p_J^*(x) f_{\chi^2_{k-p}}(x),$$

denote the Edgeworth expansions of $\mathbf{P}^*(T^{1/2}(c' \hat{\Sigma}_T c)^{-1/2} c' (\hat{\beta}_T^* - \hat{\beta}_T) \leq x)$ and $\mathbf{P}^*(J_T^* \leq x)$, respectively, where p_1^* is even, and p_2^* and p_J^* are odd. Then

$$\sup_{x \in \mathfrak{R}} |\mathbf{P}^*(T^{1/2}(c' \hat{\Sigma}_T c)^{-1/2} c' (\hat{\beta}_T^* - \hat{\beta}_T) \leq x) - \Psi_T^*(x)| = o_p(\ell T^{-1}), \tag{4}$$

$$\sup_{x \geq 0} |\mathbf{P}^*(J_T^* \leq x) - \Psi_{J,T}^*(x)| = o_p(\ell T^{-1}), \tag{5}$$

where \mathbf{P}^* is the probability measure induced by the bootstrap conditional on the data.

Theorems 1 and 2 show that the distributions of the studentized statistic and the J test statistic and their bootstrap versions can be approximated by their Edgeworth expansions. The following corollary shows the order of the bootstrap approximation error for the symmetric confidence interval and the J test statistic.

Corollary 1. Suppose that Assumption 1 holds with $r \geq 12$ replaced by $r \geq 36$. Then

$$\begin{aligned} \sup_x & |\mathbf{P}^*(T^{1/2}(|(c' \hat{\Sigma}_T^* c)^{-1/2} c' (\hat{\beta}_T^* - \hat{\beta}_T)| \leq x) \\ & - \mathbf{P}(T^{1/2}(|(c' \hat{\Sigma}_T c)^{-1/2} c' (\hat{\beta}_T - \beta_0)| \leq x))| = o_p(\ell T^{-1}) + O_p(\ell^{-q}) \end{aligned} \tag{6}$$

$$\sup_x |\mathbf{P}(J_T^* > x) - \mathbf{P}(J_T > x)| = o_p(\ell T^{-1}) + O_p(\ell^{-q}). \tag{7}$$

Remarks. 1. For the symmetric confidence interval and the J test, the approximation errors made by the first-order asymptotic theory are of order

$$O(\ell T^{-1}) + O(\ell^{-q}), \tag{8}$$

whereas the bootstrap approximation errors are of order

$$o_p(\ell T^{-1}) + O_p(\ell^{-q}). \tag{9}$$

⁶Theorem 6 requires strengthening the moment conditions in Theorem 1. To show that Assumption 1(c) is satisfied for the block bootstrap version of the data, we prove that the moments with respect to the bootstrap probability measure and the corresponding population moments are “close”, using a moment inequality that requires stronger moment conditions.

In (8) with $q < \infty$, the approximation error $O(\ell^{-q})$ is due to the bias of HAC estimation and is present not only for one-sided confidence intervals but also for two-sided confidence intervals and the J test. This error will remain present even if the bias term is included and higher-order terms are considered in asymptotic expansions.⁷

2. It follows from (8) and (9) that the bootstrap provides asymptotic refinements if the bias of the HAC covariance matrix estimator vanishes quickly enough, i.e., $O(\ell^{-q}) = o(\ell T^{-1})$. For the one-sided confidence interval, Götze and Künsch (1996) show that the bootstrap does not provide asymptotic refinements for the Bartlett kernel but for kernels of which the characteristic exponent equals two (see the last paragraph of pp. 1920–1921). However, because the $O(T^{-1/2})$ term is even and is canceled out in the Edgeworth expansion for the symmetric confidence interval and the J test statistic, bias becomes important and our results differ from theirs. Specifically, the conditions $O(\ell^{-q}) = o(\ell T^{-1})$ and $\ell = O(T^{1/3})$ imply that the bootstrap can provide asymptotic refinements only for kernels whose characteristic exponent is greater than two. For example, the truncated kernel and the trapezoidal kernel (Politis and Romano, 1995) have no asymptotic bias, and a class of kernels sometimes referred to as the Parzen (b) kernel can also satisfy the condition on the characteristic exponent. Under the assumption of exponentially decaying mixing coefficients, these kernels satisfy $O(\ell^{-q}) = o(\ell T^{-1})$.

3. If a kernel with a characteristic exponent $q > 2$ is used and if $O(\ell^{-q}) = o(\ell T^{-1})$ is satisfied, the order of the bootstrap approximation error is $o(\ell/T)$. This outcome is smaller than the order of the first-order asymptotic approximation error $O(\ell/T)$, and thus the bootstrap provides asymptotic refinements. As the divergence rate of ℓ becomes slower, the orders of the asymptotic and bootstrap approximation errors become smaller. However, because ℓ must diverge so that $\ell / \max(T^{1/(q+1)}, T^{2/r}) \rightarrow \infty$, there is no optimal divergence rate of ℓ . In an extreme case, if all the moments exist ($r \rightarrow \infty$) as in the case of bounded random variables and if a kernel with $q = \infty$ is used, the approximation error can be made arbitrarily close to those in Hall and Horowitz (1996), i.e., $o(T^{-1})$ for the bootstrap case and $O(T^{-1})$ for the asymptotic case.

4. The problem with these kernels is that the resulting HAC covariance matrix estimator is not necessarily positive semidefinite. Although we use an empirical procedure originally proposed by Andrews (2002b), this is a very troubling issue which is difficult to resolve.

4. Monte Carlo results

In this section, we conduct a small simulation study to examine the accuracy of the proposed bootstrap procedure. We consider the following stylized linear regression

⁷In the notation used in the appendix, the cumulants of g_T do not converge at a rate necessary for asymptotic refinements even if the bias term is considered in the Edgeworth expansion.

model with an intercept and a regressor, x_t :

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad \text{for } t = 1, \dots, T. \quad (10)$$

The disturbance and the regressors are generated from the following AR(1) processes with common ρ ,

$$u_t = \rho u_{t-1} + \varepsilon_{1t}, \quad (11)$$

$$x_t = \rho x_{t-1} + \varepsilon_{2t}, \quad (12)$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim N(0, I_2)$. In the simulation, we use $\beta = (\beta_1, \beta_2)' = (0, 0)'$ for the regression parameter and $\rho \in \{0.5, 0.9, 0.95\}$ for the AR parameters. For instruments, we use x_t, x_{t-1} and x_{t-2} in addition to an intercept. This choice of instruments implies an overidentified model with 2 degrees of freedom for the J test.

The choice of the block length for the bootstrap is important in practice. Ideally, one would choose a longer block length for more persistent processes and a shorter block length for less persistent processes.

In the literature on HAC estimation, this is typically accomplished by selecting the lag truncation parameter that minimizes the mean squared error of the HAC covariance matrix estimator (see Andrews, 1991; Newey and West, 1994).

Because the truncated kernel and trapezoidal kernel have no asymptotic bias, however, one cannot take advantage of the usual bias–variance trade-off and thus no optimal block length can be defined for these kernels. Thus, we propose the following procedure, which is similar to the general-to-specific modeling strategy for selecting the lag order of autoregressions in the literature on unit root testing (see Hall, 1994; Ng and Perron, 1995).

According to the Wold representation theorem, the moment function has a moving average (MA) representation of possibly infinite order. The idea is to approximate this MA representation by a sequence of finite-order MA processes. Because the block bootstrap is originally designed to capture the dependence of m -dependent-type processes when ℓ is fixed, it makes sense to approximate the process by an MA process that is m -dependent.

The proposed procedure takes the following steps:

Step 1: Let $\ell_1 < \ell_2 < \dots < \ell_{\max}$ be candidate block lengths and set $k = \max - 1$.

Step 2: Test the null that every element of the moment function is MA(ℓ_k) against the alternative that at least one of the elements is MA(ℓ_{k+1}).

Step 3: If the null is not rejected and if $k > 1$, then let $k = k - 1$ and go to Step 2. If the null is not rejected and if $k = 1$, then let $\ell = \ell_1$. If the null is rejected, then set $\ell = \ell_{k+1}$.

Because there is parameter uncertainty due to first-step estimation and because we apply a univariate testing procedure to each element of the moment function, it is difficult to control the size of this procedure. In this Monte Carlo experiment, therefore, we use the 1% level critical value to be conservative.

Our primary interest is to evaluate the errors in the coverage probabilities of symmetric confidence intervals for the regression slope parameter β_2 and the ERPs of the J test for overidentification. The kernel function employed for the bootstrap is the truncated kernel, the trapezoidal kernel with parameter $\frac{1}{2}$ and the Parzen (b) kernel with $q = 3$. The block length is selected using the procedure described above.

When kernels of which the characteristic exponent is greater than two, the resulting HAC covariance matrix estimator is not necessarily positive semidefinite although its limit is positive definite. When it is not positive definite we replace the block length by a smaller one for which the HAC covariance matrix estimator is positive definite.⁸

In addition to the results based on the bootstrap critical values, we report the asymptotic results based on the truncated, Bartlett, and QS kernels for the purpose of comparison. The bandwidth for the truncated kernel is selected based on the procedure also used for the bootstrap. Andrews' (1991) data-dependent bandwidth selection procedure is used for the Bartlett and QS kernels. Since Andrews and Monahan's (1992) prewhitening procedure based on VAR(1) model is frequently used in practice, we also report the results based on Bartlett and QS kernels applied to the prewhitened series.

Table 1 summarizes the result of the simulation study using the 90% coverage probability and the 10% nominal significance level. Results for two sample sizes (T), 64 and 128, are reported. Each bootstrap critical value is constructed from 499 replications of the bootstrap sampling process. In all experiments, the number of Monte Carlo trials is 5000.

For most cases, the coverage probabilities of the bootstrap confidence interval are more accurate than those of the asymptotic confidence interval. The degree of the reduction in the errors depends on the value of the AR parameters as well as the sample size. The bootstrap works quite well with persistent processes. Indeed, the empirical coverage probabilities of the bootstrap confidence intervals are much closer to the nominal ones compared to the asymptotic confidence intervals when T is 128 and ρ is 0.95. Because the moment functions have an AR(1) autocovariance structure, the prewhitening procedure has a considerable advantage in our simulation design. However, the bootstrap performs as well as the conventional prewhitened HAC procedure with asymptotic critical values. In contrast, the advantage of the bootstrap for the J test is not clear because the J test performs quite well even with asymptotic critical values.⁹ Judging from the results of this experiment, we recommend our bootstrap procedure especially for the confidence interval of regression parameters and for persistent series.

5. Empirical illustration

To illustrate the proposed bootstrap approach, we conduct bootstrap inference about the parameters in the monetary policy reaction function estimated by Clarida,

⁸Andrews (2002b, footnote 4) suggested the procedure for the extremum estimator to achieve higher-order asymptotic efficiency.

⁹See Tauchen (1986) and Hall and Horowitz (1996) for similar findings.

Table 1
Coverage probabilities of confidence intervals and empirical size of the J test

Kernel	$T_0 + 1$	Coverage probabilities of 90% confidence intervals				Rejection frequencies of nominal 10% J test		
		$\rho = 0.5$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0.95$	
(1) Bootstrap								
Truncated	64	81.5 (2.99) [6.4]	78.1 (3.92) [7.2]	75.0 (4.01) [6.2]	9.6	8.4	7.3	
	128	84.5 (4.32) [5.1]	87.3 (6.79) [6.8]	87.2 (7.51) [5.4]	10.7	10.3	9.7	
Trapezoidal	64	81.6 [5.0]	77.0 [6.2]	73.4 [6.1]	10.0	9.2	9.2	
	128	84.9 [1.8]	87.6 [3.3]	86.5 [4.4]	10.9	12.4	11.6	
Parzen (b)	64	83.7 [3.0]	76.8 [5.3]	73.5 [5.5]	9.2	8.9	9.0	
	128	85.5 [1.1]	87.8 [2.7]	86.7 [4.1]	10.3	11.6	11.8	
(2) First-order asymptotics								
Truncated	64	73.0	55.7	51.0	11.2	10.9	10.7	
	128	78.0	63.1	57.8	10.8	11.4	11.3	
Trapezoidal	64	73.1	54.6	48.7	11.2	13.2	13.2	
	128	78.4	62.6	55.2	10.6	13.7	14.5	

Parzen (b)	64	75.3	54.2	48.6	10.2	12.9	13.8
	128	79.1	62.8	55.5	9.5	13.1	14.7
Bartlett	64	78.4 (4.18)	56.0 (11.22)	49.5 (14.63)	8.0	4.3	3.5
	128	83.1 (5.38)	66.0 (15.29)	57.5 (21.54)	7.8	5.2	3.9
QS	64	75.7 (4.22)	46.2 (13.31)	36.5 (18.32)	16.8	35.1	41.8
	128	82.2 (4.96)	59.9 (16.36)	47.2 (25.17)	13.1	27.1	34.5
Bartlett (Prewhitened)	64	79.8 (1.02)	65.9 (1.52)	59.8 (1.58)	10.8	13.8	15.1
	128	84.7 (1.02)	74.6 (1.58)	68.2 (1.65)	10.2	12.9	14.6
QS (Prewhitened)	64	79.9 (1.26)	65.3 (1.63)	59.1 (1.66)	11.4	15.1	16.9
	128	84.6 (1.24)	74.2 (1.63)	67.7 (1.67)	10.3	13.7	15.2

Note: Numbers in parenthesis are (average) block length and bandwidths selected using automatic procedures. Numbers in brackets are the frequencies of psd corrections.

Galí and Gertler (2000, hereafter, CGG). In CGG, the target for the federal funds rate r_t^* is given by

$$r_t^* = r^* + \beta(E[\pi_{t+1}|\Omega_t] - \pi^*) + \gamma E[x_t|\Omega_t], \quad (13)$$

where π_t is the inflation rate, π^* is the target for inflation, Ω_t is the information set at time t , x_t is the output gap, and r^* is the target with zero inflation and output gap. Policy rules (13) with $\beta > 1$ and $\gamma > 0$ are stabilizing and those with $\beta \leq 1$ and $\gamma \leq 0$ are destabilizing. CGG obtain the GMM estimates of β and γ based on the set of unconditional moment conditions

$$E\{[r_t - (1 - \rho_1 - \rho_2)[rr^* - (\beta - 1)\pi^* + \beta\pi_{t+1} + \gamma x_t] + \rho_1 r_{t-1} + \rho_2 r_{t-2}]z_t\} = 0, \quad (14)$$

where r_t is the actual federal fund rate, rr^* is the equilibrium real rate, and z_t is a vector of instruments. They find that the GMM estimate of β is significantly less than unity during the pre-Volcker era, while the estimate is significantly greater than unity during the Volcker–Greenspan era.

We reexamine these findings by applying our bootstrap procedure as well as the bootstrap procedure of the standard HAC asymptotics. We obtain GMM estimates of β and γ based on the linear moment conditions

$$E\{[r_t - c - \theta_1\pi_{t+1} - \theta_2x_t - \rho_1r_{t-1} - \rho_2r_{t-2}]z_t\} = 0, \quad (15)$$

where $c = (1 - \rho_1 - \rho_2)[rr^* - (\beta - 1)\pi^*]$. Then $\hat{\beta}_T = \hat{\theta}_{1T}/(1 - \hat{\rho}_{1T} - \hat{\rho}_{2T})$ and $\hat{\gamma}_T = \hat{\theta}_{2T}/(1 - \hat{\rho}_{1T} - \hat{\rho}_{2T})$, where $\hat{\theta}_{1T}$, $\hat{\theta}_{2T}$, $\hat{\rho}_{1T}$ and $\hat{\rho}_{2T}$ are the GMM estimates of θ_1 , θ_2 , ρ_1 and ρ_2 , respectively. We use CGG's baseline data set and two sample periods, the pre-Volcker period (1960:1–1979:2) and the Volcker–Greenspan period (1979:3–1996:3) (see CGG for the description of the data source). In addition to their baseline specification, we construct the optimal weighting matrix using the inverse of the HAC covariance matrix estimator to allow for more general dynamic specifications in the determination of the actual funds rate. For the asymptotic confidence intervals, we use the conventional prewhitened and recolored estimates based on the Bartlett and QS kernels with the automatic bandwidth selection method (Andrews, 1991; Andrews and Monahan, 1992). For the confidence intervals constructed from our bootstrap, we use the truncated, trapezoidal, and Parzen (b) kernels. We use the data-dependent procedure described in the previous section to select the block length for the bootstrap. The number of bootstrap replications is set to 499.

Table 2 presents GMM estimates of these parameters. Asymptotic standard errors are reported in parentheses. The first two rows of each of Tables 2(a) and (b) replicate CGG's results. These findings are robust to whether or not the HAC covariance matrix estimator is used.

Table 3 shows 90% two-sided confidence intervals of these parameters. Consistent with CGG's findings, the upper bound of the asymptotic confidence interval for β is less than unity during the pre-Volcker period, and the lower bound is far greater than unity during the Volcker–Greenspan period. Using these estimates, CGG suggest that the Fed was accommodating inflation before 1979, but not after 1979.

Table 2
GMM Estimates of the policy rule parameters

Kernel	β	γ	J
(a) <i>Pre-Volcker period: 1960:1–1979:2</i>			
None	0.834 (0.067)	0.274 (0.087)	13.075 (0.126)
Bartlett	0.871 (0.030)	0.392 (0.073)	22.206 (0.671)
QS	0.871 (0.030)	0.388 (0.073)	22.242 (0.673)
(b) <i>Volcker–Greenspan period: 1979:3–1996:3</i>			
None	2.153 (0.379)	0.933 (0.454)	21.376 (0.625)
Bartlett	2.258 (0.148)	0.854 (0.224)	23.314 (0.726)
QS	2.280 (0.148)	0.803 (0.216)	34.607 (0.978)

Note: Asymptotic standard errors for the estimates of β and γ , and asymptotic p values for the J statistics are in parentheses. For the asymptotic confidence interval based on the Bartlett and QS kernels, the data-dependent bandwidth estimator of Andrews (1991) and the prewhitening procedure of Andrews and Monahan (1992) are used. The estimated bandwidths are reported in Table 3. “None” indicates that the inverse of the variance–covariance matrix is used as the weighting matrix.

The bootstrap confidence interval, however, indicates that β may be greater than unity even during the pre-Volcker period, consistent with the view that the Fed has always been combating inflation. Moreover, unlike the asymptotic confidence interval, the bootstrap confidence interval does not rule out that γ is negative during the Volcker–Greenspan period.

6. Concluding remarks

In this paper we establish that the bootstrap provides asymptotic refinements for the GMM estimator of overidentified linear models when autocorrelation structures of moment functions are unknown. Because of the nonparametric nature of the HAC covariance matrix estimator, the order of the bootstrap approximation error is larger than $O(T^{-1})$, the typical order of the bootstrap approximation error for parametric estimators.

By taking into account the HAC covariance matrix estimator in the Edgeworth expansion, we find that kernels with a characteristic exponent strictly greater than two are required for the refinement for the symmetric confidence intervals and the J test statistic. This contrasts with the result for the one-sided confidence interval that can be obtained for widely used kernels with a characteristic exponent equal to two. Our finding shows the importance of the choice of kernels in the bootstrap compared to the conventional first-order asymptotic theory. Nevertheless, the bootstrap provides improved approximations relative to the first-order approximation.

Table 3
90% Confidence intervals of the policy rule parameters

	Kernel	ℓ	β	γ
(a) <i>Pre-Volcker period: 1960:1–1979:2</i>				
Asymptotic	None	0	(0.724, 0.945)	(0.131, 0.416)
	Bartlett	0.640	(0.822, 0.921)	(0.272, 0.512)
	QS	0.944	(0.823, 0.920)	(0.268, 0.507)
Bootstrap	Truncated	4	(0.019, 4.199)	(−0.901, 0.203)
	Trapezoidal	3	(0.755, 1.075)	(0.031, 0.145)
	Parzen (b)	3	(0.797, 1.100)	(0.096, 0.181)
(b) <i>Volcker–Greenspan period: 1979:3–1996:3</i>				
Asymptotic	None	0	(1.530, 2.776)	(0.187, 1.680)
	Bartlett	1.227	(2.015, 2.502)	(0.485, 1.222)
	QS	1.460	(2.038, 2.523)	(0.449, 1.158)
Bootstrap	Truncated	4	(0.842, 1.638)	(−0.541, −0.301)
	Trapezoidal	3	(0.177, 3.289)	(−0.820, 0.719)
	Parzen (b)	3	(0.230, 3.255)	(−0.798, 0.782)

Note: “None” indicates that the inverse of the variance–covariance matrix is used as the weighting matrix. ℓ denotes the bandwidth for the asymptotic confidence interval and the block length for the bootstrap confidence interval. For the asymptotic confidence interval based on the Bartlett and QS kernels, the data-dependent bandwidth estimator of Andrews (1991) and the prewhitening procedure of Andrews and Monahan (1992) are used. For the bootstrap confidence interval, the data-dependent procedure described in Section 4 is used to select the block length.

We note that an extension of the present results to nonlinear dynamic models as well as further investigation of data-dependent methods for selecting the optimal block length would be useful.

Acknowledgements

We thank Jordi Galí for providing us with the data and program used in Clarida, Galí, and Gertler (2000). We also thank the two anonymous referees, Valentina Corradi, Alastair Hall, Lutz Kilian, Yoshihiko Nishiyama, and seminar participants at Brown University, the Colloque C.R.D.E. Conference: Resampling Methods in Econometrics, University of Michigan, the 2001 NBER Summer Institute, the 2001 North American Summer Meeting of the Econometric Society and the 2000 Triangle Econometrics Conference for helpful comments.

Appendix A

A.1. Notation and Lemmas

\otimes denotes the Kronecker product operator. $\|\cdot\|$ denotes the Euclidean norm, i.e., $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$, where x is an n -dimensional vector. We will write $\omega(j/\ell)$ as ω_j for

notational simplicity. $\kappa_j(x)$ denotes the j th cumulant of a random variable x . $\text{vec}(\cdot)$ is the column-by-column vectorization function. $\text{vech}(\cdot)$ denotes the column stacking operator that stacks the elements on and below the leading diagonal.

Let $u_t = y_t - \beta'_0 x_t$, $\hat{u}_t = y_t - \hat{\beta}'_T x_t$, $v_t = z_t u_t$, $\hat{v}_t = z_t \hat{u}_t$, $w_t = z_t x'_t$,

$$\hat{\Gamma}_j = \begin{cases} \frac{1}{T} \sum_{t=1}^T \hat{v}_{t+j} \hat{v}'_t & j \geq 0, \\ \frac{1}{T} \sum_{t=1}^T \hat{v}_t \hat{v}'_{t-j} & j < 0, \end{cases} \quad \nabla \tilde{\Gamma}_j \delta = \begin{cases} \frac{1}{T} \sum_{t=1}^T (v_{t+j} z'_t x'_t + z_{t+j} v'_t x'_{t+j}) \delta & j \geq 0, \\ \frac{1}{T} \sum_{t=1}^T (v_t z'_{t-j} x'_{t-j} + z_t v'_{t-j} x'_t) \delta & j < 0, \end{cases}$$

$$\tilde{\Gamma}_j = \begin{cases} \frac{1}{T} \sum_{t=1}^T v_{t+j} v'_t & j \geq 0, \\ \frac{1}{T} \sum_{t=1}^T v_t v'_{t-j} & j < 0, \end{cases} \quad \nabla \Gamma_j \delta = \begin{cases} \text{E}(v_{t+j} z'_t x'_t + z_{t+j} v'_t x'_{t+j}) \delta & j \geq 0, \\ \text{E}(v_t z'_{t-j} x'_{t-j} + z_t v'_{t-j} x'_t) \delta & j < 0, \end{cases}$$

$$\Gamma_j = \begin{cases} \text{E}(v_{t+j} v'_t) & j \geq 0, \\ \text{E}(v_t v'_{t-j}) & j < 0, \end{cases} \quad \delta' \nabla^2 \tilde{\Gamma}_j \delta = \begin{cases} \frac{1}{T} \sum_{t=1}^T \delta' x_{t+j} z_{t+j} z'_t x'_t \delta & j \geq 0, \\ \frac{1}{T} \sum_{t=1}^T \delta' x_t z_t z'_{t-j} x'_{t-j} \delta & j < 0. \end{cases}$$

$$\hat{S}_T = \sum_{j=-\ell}^{\ell} \omega_j \hat{\Gamma}_j, \quad \tilde{S}_T = \sum_{j=-\ell}^{\ell} \omega_j \tilde{\Gamma}_j, \quad \bar{S}_T = \sum_{j=-\ell}^{\ell} \omega_j \Gamma_j,$$

$$S_T = \sum_{j=-T+1}^{T-1} \left(1 - \frac{|j|}{T}\right) \Gamma_j, \quad \nabla \tilde{S}_T \delta = \sum_{j=-\ell}^{\ell} \omega_j \nabla \tilde{\Gamma}_j \delta, \quad \nabla \bar{S}_T \delta = \sum_{j=-\ell}^{\ell} \omega_j \nabla \Gamma_j \delta,$$

$$\nabla S \delta = \sum_{j=-\infty}^{\infty} \nabla \Gamma_j \delta, \quad \delta' \nabla^2 \tilde{S}_T \delta = \sum_{j=-\ell}^{\ell} \omega_j \delta' \nabla^2 \tilde{\Gamma}_j \delta, \quad S_0 = \sum_{j=-\infty}^{\infty} \Gamma_j,$$

where δ is a p -dimensional vector. Let $G_T = (1/T) \sum_{t=1}^T z_t x'_t$, $G_0 = \text{E}(z_t x_t)$, and $m_T = T^{-1/2} \sum_{t=1}^T v_t$. Then the studentized statistic can be written as

$$f_T = \sqrt{T} (c' \hat{\Sigma} c)^{-1/2} c' (\hat{\beta}_T - \beta_0) = (c' (G'_T \hat{S}_T^{-1} G_T)^{-1} c)^{-1/2} c' (G'_T \hat{S}_T^{-1} G_T)^{-1} G'_T \hat{S}_T^{-1} m_T.$$

We use the following notation for the bootstrap. Let

$$m_T^* = \frac{1}{\sqrt{T}} \sum_{t=1}^T (z_t^* u_t^* - \mu_T^*) = \frac{1}{\sqrt{b}} \sum_{k=1}^b B_{N_k},$$

$$B_{N_k} = \frac{1}{\sqrt{\ell}} \sum_{i=1}^{\ell} (z_{N_k+i} \hat{u}_{N_k+i} - \mu_T^*) = \frac{1}{\sqrt{\ell}} \sum_{i=1}^{\ell} (\hat{v}_{N_k+i}^* - \mu_T^*),$$

$$\widehat{B}_{N_k} = \frac{1}{\sqrt{\ell}} \sum_{i=1}^{\ell} (z_{N_{k+i}} \widehat{u}_{N_{k+i}}^* - \mu_{N_k}^*), \quad \widehat{u}_i^* = y_i - \widetilde{\beta}^{*'} x_i,$$

$$G_T^* = \frac{1}{T} \sum_{i=1}^T z_i^* x_i^{*'} = \frac{1}{b} \sum_{k=1}^b F_{N_k},$$

$$F_{N_k} = \frac{1}{\ell} \sum_{i=1}^{\ell} z_{N_{k+i}} x_{N_{k+i}}' = \frac{1}{\ell} \sum_{i=1}^{\ell} w_{N_{k+i}},$$

$$\widehat{S}_T^* = \frac{1}{b} \sum_{k=1}^b \widehat{B}_{N_k} \widehat{B}_{N_k}', \quad \widetilde{S}_T^* = \frac{1}{b} \sum_{k=1}^b B_{N_k} B_{N_k}', \quad S_T^* = \text{Var}^*(m_T^*).$$

Then the bootstrap version of the first-step and the second-step GMM estimators can be written as

$$\begin{aligned} \widetilde{\beta}^* &= \widehat{\beta} + \left[\frac{1}{b} \sum_{k=1}^b F'_{N_k} V_T \frac{1}{b} \sum_{k=1}^b F_{N_k} \right]^{-1} \frac{1}{b} \sum_{k=1}^b F'_{N_k} V_T \frac{1}{\sqrt{Tb}} \sum_{k=1}^b B_{N_k}, \\ &= \widehat{\beta} + [G_T^* V_T G_T^*]^{-1} G_T^* V_T \frac{1}{\sqrt{T}} m_T^*, \end{aligned}$$

$$\begin{aligned} \widehat{\beta}^* &= \widehat{\beta} + \left[\frac{1}{b} \sum_{k=1}^b F'_{N_k} \widehat{S}_T^{*-1} \frac{1}{b} \sum_{k=1}^b F_{N_k} \right]^{-1} \frac{1}{b} \sum_{k=1}^b F'_{N_k} \widehat{S}_T^{*-1} \frac{1}{\sqrt{Tb}} \sum_{k=1}^b B_{N_k}, \\ &= \widehat{\beta} + [G_T^* \widehat{S}_T^{*-1} G_T^*]^{-1} G_T^* \widehat{S}_T^{*-1} \frac{1}{\sqrt{T}} m_T^*, \end{aligned}$$

respectively.

Next, we will present the lemmas used in the proofs of the theorems. The proof of the lemmas are in the technical appendix which is available upon request from the authors. Lemma A.1 produces a Taylor series expansion of the studentized statistic f_T . Lemma A.2 provides bounds on the moments and will be used in the proofs of Lemmas A.3–A.6. Lemma A.3 shows the limits and the convergence rates of the first three cumulants of g_T in (A.1), that will be used to derive the formal Edgeworth expansion. Lemmas A.5 and A.6 provide bounds on the approximation error. For convenience, we present Lemma A.8 that will be used in the proofs of Lemmas A.9 and A.10. Lemma A.9 shows the consistency and convergence rate of the bootstrap version of the moments. Lemma A.10 shows the limits and the convergence rates of the first three cumulants of the bootstrap version.

Lemma A.1.

$$\begin{aligned} f_T &= \mathbf{a}' m_T + \mathbf{b}' [\text{vec}(G_T - G_0) \otimes m_T] + \mathbf{c}' [\text{vech}(\widehat{S}_T - S_0) \otimes m_T] \\ &\quad + \mathbf{d}' [\text{vec}(G_T - G_0) \otimes \text{vech}(\widehat{S}_T - S_0) \otimes m_T] \\ &\quad + \mathbf{e}' [\text{vech}(\widehat{S}_T - S_0) \otimes \text{vech}(\widehat{S}_T - S_0) \otimes m_T] + O_p((\ell/T)^{3/2}) \end{aligned}$$

$$\begin{aligned}
 &= \mathbf{a}'m_T + \mathbf{b}'[\text{vec}(G_T - G_0) \otimes m_T] + \mathbf{c}'[\text{vech}(\hat{S}_T - \bar{S}_T) \otimes m_T] \\
 &\quad + \mathbf{c}'[\text{vech}(\bar{S}_T - S_0) \otimes m_T] + \mathbf{d}'[\text{vec}(G_T - G_0) \otimes \text{vech}(\hat{S}_T - \bar{S}_T) \otimes m_T] \\
 &\quad + \mathbf{e}'[\text{vech}(\hat{S}_T - \bar{S}_T) \otimes \text{vech}(\hat{S}_T - \bar{S}_T) \otimes m_T] \\
 &\quad + \mathbf{d}'[\text{vec}(G_T - G_0) \otimes \text{vech}(\bar{S}_T - S_0) \otimes m_T] \\
 &\quad + \mathbf{e}'[\text{vech}(\hat{S}_T - \bar{S}_T) \otimes \text{vech}(\bar{S}_T - S_0) \otimes m_T] \\
 &\quad + \mathbf{e}'[\text{vech}(\bar{S}_T - S_0) \otimes \text{vech}(\hat{S}_T - \bar{S}_T) \otimes m_T] \\
 &\quad + \mathbf{e}'[\text{vech}(\bar{S}_T - S_0) \otimes \text{vech}(\bar{S}_T - S_0) \otimes m_T] + O_p((\ell/T)^{3/2}) \\
 &\equiv g_T + \mathbf{c}'[\text{vech}(\bar{S}_T - S_0) \otimes m_T] + \mathbf{d}'[\text{vec}(G_T - G_0) \otimes \text{vech}(\bar{S}_T - S_0) \otimes m_T] \\
 &\quad + \mathbf{e}'[\text{vech}(\hat{S}_T - \bar{S}_T) \otimes \text{vech}(\bar{S}_T - S_0) \otimes m_T] + \mathbf{e}'[\text{vech}(\bar{S}_T - S_0) \\
 &\quad \otimes \text{vech}(\hat{S}_T - \bar{S}_T) \otimes m_T] + \mathbf{e}'[\text{vech}(\bar{S}_T - S_0) \otimes \text{vech}(\bar{S}_T - S_0) \otimes m_T] \\
 &\quad + O_p((\ell/T)^{3/2}), \tag{A.1}
 \end{aligned}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ and \mathbf{e} are $k, k^2p, k(k^2 + k)/2, k^2(k^2 + k)p/2$ and $k((k^2 + k)/2)^2$ -dimensional vectors of smooth functions of G_0 and S_0 , respectively, and $g_T = \mathbf{a}'m_T$.

Lemma A.2.

$$E\|m_T\|^{r+\eta} = O(1), \tag{A.2}$$

$$E\|T^{1/2}\text{vec}(G_T - G_0)\|^{r+\eta} = O(1), \tag{A.3}$$

$$E\|(T/\ell)^{1/2}\text{vech}(\tilde{S}_T - \bar{S}_T)\|^{r/2} = O(1), \tag{A.4}$$

$$E\|(T/\ell)^{1/2}\text{vech}(\nabla\tilde{S}_T - \nabla\bar{S}_T)\|^{r/2} = O(1), \tag{A.5}$$

$$E\|T^{1/2}\text{vech}(\hat{S}_T - \tilde{S}_T)\|^{r/2} = O(1). \tag{A.6}$$

Lemma A.3.

$$T^{1/2}\kappa_1(g_T) = \alpha_\infty + O(\ell^{-q}) + o(\ell T^{-1/2}), \tag{A.7}$$

$$(T/\ell)(\kappa_2(g_T) - 1) = \gamma_\infty + O(\ell^{-1/2}), \tag{A.8}$$

$$T^{1/2}\kappa_3(g_T) = \kappa_\infty - 3\alpha_\infty + O(\ell^{-q}) + o(\ell T^{-1/2}), \tag{A.9}$$

$$(T/\ell)(\kappa_4(g_T) - 3) = \zeta_\infty + O(\ell^{-1/2}), \tag{A.10}$$

where

$$\begin{aligned}
 \alpha_\infty &= \mathbf{b}' \sum_{i=-\infty}^{\infty} E[w_0 \otimes v_i] + \mathbf{c}' \sum_{i,j=-\infty}^{\infty} E[\text{vech}(v_0 v_i') \otimes v_j] \\
 &\quad + \mathbf{c}' \sum_{i=-\infty}^{\infty} E\{\text{vech}[\nabla\bar{S}(E(w_0)' V E(w_0))^{-1} E(w_0)' V v_0] \otimes v_i\},
 \end{aligned}$$

$$\begin{aligned} \gamma_\infty &= 2 \lim_{T \rightarrow \infty} \frac{1}{\ell} \sum_{j=-\ell}^{\ell} \sum_{i,k=-T}^T \mathbb{E}\{\mathbf{a}'v_0\mathbf{c}'[\text{vech}(v_i v'_{i-j} - \Gamma_j) \otimes v_k]\} \\ &+ 2 \lim_{T \rightarrow \infty} \frac{1}{\ell T} \sum_{i,l=-\ell}^{\ell} \sum_{i,k,m=-T}^T \mathbb{E}\{\mathbf{a}'v_0\mathbf{e}'[\text{vech}(v_i v'_{i-j} - \Gamma_j) \otimes \text{vech}(v_k v'_{k-l} - \Gamma_l) \otimes v_m]\} \\ &+ \lim_{T \rightarrow \infty} \frac{1}{\ell T} \sum_{j,k,m=-T}^T \sum_{i,l=-\ell}^{\ell} \mathbb{E}\{\mathbf{c}'[\text{vech}(v_0 v'_{-i} - \Gamma_i) \otimes v_j]\mathbf{c}'[\text{vech}(v_k v'_{k-l} - \Gamma_k) \otimes v_m]\}, \end{aligned}$$

$$\begin{aligned} \kappa_\infty &= \sum_{i,j=-\infty}^{\infty} \mathbb{E}(\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{a}'v_j) \\ &+ 3 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i,j,k=-T+1}^{T-1} \mathbb{E}\{\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{b}'[\text{vech}(w_j - \mathbb{E}(w_j)) \otimes v_k]\} \\ &+ 3 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i,j,k,l=-T}^T \mathbb{E}\{\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{c}'[\text{vech}(v_j v'_{j-k} - \Gamma_k) \otimes v_l]\} \\ &+ 3 \lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{i,j,k=-T}^T \mathbb{E}\{\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{c}'\text{vech}[\nabla\bar{S}(\mathbb{E}(w_0)'V\mathbb{E}(w_0))^{-1}\mathbb{E}(w_0)'Vv_j] \otimes v_k\}, \end{aligned}$$

$$\begin{aligned} \zeta_\infty &= \frac{4}{\ell T} \sum_{i,j,k,m=-T}^T \sum_{l=-\ell}^{\ell} \mathbb{E}\{\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{a}'v_j\mathbf{c}'[\text{vech}(v_k v'_{k-l} - \Gamma_l) \otimes v_m]\} \\ &+ \lim_{T \rightarrow \infty} \frac{4}{\ell T^2} \sum_{i,j,k,m,o=-T}^T \sum_{l,n=-\ell}^{\ell} \mathbb{E}\{\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{a}'v_j\mathbf{e}'[\text{vech}(v_k v'_{k-l} - \Gamma_l) \otimes \text{vech}(v_m v'_{m-n} - \Gamma_n) \otimes v_o]\} \\ &+ \lim_{T \rightarrow \infty} \frac{6}{\ell T^2} \sum_{i,j,l,m,o=-T}^T \sum_{k,n=-\ell}^{\ell} \mathbb{E}\{\mathbf{a}'v_0\mathbf{a}'v_i\mathbf{c}'[\text{vech}(v_j v'_{j-k} - \Gamma_k) \otimes v_l] \\ &\times \mathbf{c}'[\text{vech}(v_m v'_{m-n} - \Gamma_n) \otimes v_o]\} \\ &- 12 \lim_{T \rightarrow \infty} \frac{1}{\ell} \sum_{j,l=-T}^T \sum_{k=-\ell}^{\ell} \mathbb{E}\{\mathbf{a}'v_0\mathbf{c}'[\text{vech}(v_j v'_{j-k} - \Gamma_k) \otimes v_l]\} \\ &- 12 \lim_{T \rightarrow \infty} \frac{1}{\ell T} \sum_{j,l,n=-T}^T \sum_{k,m=-\ell}^{\ell} \mathbb{E}\{\mathbf{a}'v_0\mathbf{e}'[\text{vech}(v_j v'_{j-k} - \Gamma_k) \otimes (v_l v'_{l-m} - \Gamma_m) \otimes v_n]\} \\ &- 6 \lim_{T \rightarrow \infty} \frac{1}{\ell T^2} \sum_{j,k,m=-T}^T \sum_{i,l=-\ell}^{\ell} \mathbb{E}\{\mathbf{c}'[(v_0 v'_{-i} - \Gamma_i) \otimes v_j]\mathbf{c}'[(v_k v'_{k-l} - \Gamma_l) \otimes v_m]\}. \end{aligned}$$

Lemma A.4. Let $\psi_{g,T}(\theta)$ denote the characteristic function of g_T . Then

$$\begin{aligned} \psi_{g,T}(\theta) &= \exp\left(-\frac{\theta^2}{2}\right) \left[1 + T^{-(1/2)}(\alpha_\infty(i\theta) - \frac{i\theta^3}{6}(\kappa_\infty - 3\alpha_\infty)) \right. \\ &\quad \left. - \frac{\ell}{T} \left(\frac{\theta^2}{2} \gamma_\infty - \frac{\theta^4}{24} \zeta_\infty \right) + o\left(\frac{\ell}{T}\right) \right], \end{aligned} \tag{A.11}$$

$$P(g_T \leq x) = \Psi(x) + T^{-1/2}p_1(x) + (\ell/T)p_2(x) + o(\ell/T). \tag{A.12}$$

Lemma A.5. *Following Götze and Künsch (1996), define a truncation function by*

$$\tau(x) = T^\gamma x f(T^{-\gamma} \|x\|) / \|x\|,$$

where $\gamma \in (2/r, 1/2)$ and $f \in C^\infty(0, \infty)$ satisfies (i) $f(x) = x$ for $x \leq 1$; (ii) f is increasing; and (iii) $f(x) = 2$ for $x \geq 2$. Let f_T^\dagger denote f_T with $\bar{R}_t \equiv (v_t', \tilde{v}_t', \text{vec}(w_t)')$ replaced by

$$\bar{R}_t^\dagger = (v_t^\dagger', \tilde{v}_t^\dagger', \text{vec}(w_t^\dagger)')' = \tau((v_t', \tilde{v}_t', \text{vec}(w_t)')').$$

Let Ψ_T^\dagger and $\Psi_{g,T}^\dagger$ denote the Edgeworth expansions of f_T^\dagger and g_T^\dagger , respectively. Let $\psi_{g,T}^\dagger(x)$ and $\tilde{\psi}_{g,T}^\dagger(x)$ denote the characteristic functions of g_T^\dagger and $\Psi_{g,T}^\dagger$, respectively. Then

$$\begin{aligned} \sup_x |P(f_T \leq x) - \Psi_T(x)| &\leq C \int_{|\theta| < T^{1-2/r}} |\psi_{g,T}^\dagger(\theta) - \tilde{\psi}_{g,T}^\dagger(\theta)| |\theta|^{-1} d\theta \\ &+ O(\ell^{-q}) + o(\ell T^{-1}). \end{aligned} \tag{A.13}$$

Lemma A.6. *For $0 < \varepsilon < 1/6$,*

$$\int_{|\theta| \leq T^\varepsilon} |\psi_{g,T}^\dagger(\theta) - \tilde{\psi}_{g,T}^\dagger(\theta)| |\theta|^{-1} d\theta = o(\ell T^{-1}). \tag{A.14}$$

Lemma A.7.

$$\int_{T^\varepsilon < |\theta| < T^{1-2/r}} |\psi_{g,T}^\dagger(\theta) - \tilde{\psi}_{g,T}^\dagger(\theta)| |\theta|^{-1} d\theta = o(\ell T^{-1}). \tag{A.15}$$

Lemma A.8. *For $1 \leq s \leq r/2$,*

$$E^*[\|\text{vec}(F_{N_j})\|^s] - E\{E^*[\|\text{vec}(F_{N_j})\|^s]\} = O_p(b^{-1/2}), \tag{A.16}$$

$$E^*[\|B_{N_j}\|^s] - E\{E^*[\|B_{N_j}\|^s]\} = O_p(b^{-1/2}). \tag{A.17}$$

Lemma A.9. *Let $G_0^* = E^*(G_T^*)$ and b_T^* and c_T^* denote the bootstrap version of \mathbf{b} and \mathbf{c} in Lemma A.1 with S_0 replaced by S_T^* , respectively. Then*

$$G_0^* = G_0 + O_p(T^{-1/2}), \tag{A.18}$$

$$S_T^* = S + O(\ell^{-1}) + O_p(b^{-1/2}). \tag{A.19}$$

Lemma A.10. *Let*

$$\alpha_T^* = T^{1/2} \kappa_1^*(g_T^*),$$

$$\gamma_T^* = (T/\ell)(\kappa_2^*(g_T^*) - 1) = (T/\ell)(E^*(g_T^{*2}) - [E^*(g_T^*)]^2 - 1),$$

$$\kappa_T^* = T^{1/2} E^*(g_T^{*3}) = T^{1/2} \{\kappa_3^*(g_T^*) + 3E^*(g_T^{*2})E^*(g_T^*) - 2[E^*(g_T^*)]^3\},$$

$$\zeta_T^* = (T/\ell)(\kappa_4^*(g_T^*) - 3).$$

Then

$$\begin{aligned} \alpha_T^* &= \alpha_\infty + T^{1/2} \mathbf{b}' E^* [\text{vec}(G_T^* - G_0^*) \otimes m_T^*] + T^{1/2} \mathbf{c}' E^* [\text{vech}(\tilde{S}_T^* - S_T^*) \otimes m_T^*] \\ &\quad + T^{1/2} \mathbf{c}' E^* [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] + o_p^*(\ell T^{-1/2}) \\ &= \alpha_\infty + O_p(\ell^{-1}) + O_p(b^{-1/2}) + o_p^*(\ell T^{-1/2}), \end{aligned} \tag{A.20}$$

$$\begin{aligned} \gamma_T^* &= \gamma_\infty + 2(T/\ell) E^* \{ \mathbf{a}' m_T^* \mathbf{b}' [\text{vec}(G_T^* - G_0^*) \otimes m_T^*] \\ &\quad + 2(T/\ell) E^* \{ \mathbf{a}' m_T^* \mathbf{c}' [\text{vech}(\tilde{S}_T^* - S_T^*) \otimes m_T^*] \\ &\quad + 2(T/\ell) E^* \{ \mathbf{a}' m_T^* \mathbf{c}' [\text{vech}(\tilde{S}_T^* - S_T^*) \otimes \text{vech}(\tilde{S}_T^* - S_T^*) \otimes m_T^*] \} \\ &\quad + (T/\ell) E^* \{ \mathbf{c}' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \}^2 + o_p^*(1), \\ &= \gamma_\infty + o_p(1) + o_p^*(1), \end{aligned} \tag{A.21}$$

$$\begin{aligned} \kappa_T^* &= \kappa_\infty + T^{1/2} E^* [(\mathbf{a}' m_T^*)^3] + 3T^{1/2} E^* \{ (\mathbf{a}' m_T^*)^2 \mathbf{b}' [\text{vec}(G_T - G_0)' \otimes m_T] \} \\ &\quad + 3T^{1/2} E^* \{ (\mathbf{a}' m_T^*)^2 \mathbf{c}' [\text{vech}(\tilde{S}_T^* - S_T^*) \otimes m_T^*] \} \\ &\quad + 3T^{1/2} E^* \{ (\mathbf{a}' m_T^*)^2 \mathbf{c}' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \} + o_p^*(\ell T^{-1/2}) \\ &= \kappa_\infty + O_p(\ell^{-1/2}) + O_p(b^{-1/2}) + o_p^*(\ell T^{-1/2}), \end{aligned} \tag{A.22}$$

$$\begin{aligned} \zeta_T^* &= \zeta_\infty + 4(T/\ell) E^* \{ (\mathbf{a}' m_T^*)^3 \mathbf{c}' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \} \\ &\quad + 4(T/\ell) E^* \{ (\mathbf{a}' m_T^*)^3 e' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes \text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \} \\ &\quad + 6(T/\ell) E^* \{ (\mathbf{a}' m_T^*)^2 \{ \mathbf{c}' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \}^2 \} \\ &\quad - 12(T/\ell) E^* \{ \mathbf{a}' m_T^* \mathbf{c}' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \} \\ &\quad - 12(T/\ell) E^* \{ \mathbf{a}' m_T^* e' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes \text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \} \\ &\quad - 6(T/\ell) E^* \{ \mathbf{c}' [\text{vech}(\hat{S}_T^* - \tilde{S}_T^*) \otimes m_T^*] \}^2 + o_p^*(1), \\ &= \zeta_\infty + o_p(1) + o_p^*(1), \end{aligned} \tag{A.23}$$

where $\alpha_\infty, \gamma_\infty, \kappa_\infty,$ and ζ_∞ are defined in Lemma A.4.

A.2. Proofs of main theorems

Proof of Theorem 1. We have

$$\begin{aligned} &\sup_x |P(f_T \leq x) - \Psi_T(x)| \\ &\leq C \int_{|\theta| < T^{1-2/r}} |\psi_{g,T}^\dagger(\theta) - \tilde{\psi}_{g,T}^\dagger(\theta)| |\theta|^{-1} d\theta + O(\ell^{-q}) + o(\ell T^{-1}) \\ &= C \left[\int_{|\theta| < T^e} |\psi_{g,T}^\dagger(\theta) - \tilde{\psi}_{g,T}^\dagger(\theta)| |\theta|^{-1} d\theta \right. \end{aligned}$$

$$\begin{aligned}
 & + \int_{T^\varepsilon < |\theta| < T^{1-2/r}} |\psi_{g,T}^\dagger(\theta) - \tilde{\psi}_{g,T}^\dagger(\theta)| |\theta|^{-1} d\theta \Big] + O(\ell^{-q}) + o(\ell T^{-1}) \\
 & = o(\ell T^{-1}) + O(\ell^{-q}), \tag{A.24}
 \end{aligned}$$

where the first inequality follows from Lemma A.5 and the last equality follows from Lemmas A.6 and A.7. Thus, the result for the studentized statistic (2). Let $J_T^{1/2} = \hat{S}_T^{-1/2} \sum_{t=1}^T z_t(y_t - \hat{\beta}'_T x_t)$. Then it follows from the first-order condition that

$$J_T^{1/2} = (I_k - G_T(G'_T G_T)^{-1} G'_T) S_T^{-1/2} \sum_{t=1}^T z_t(y_t - \hat{\beta}'_T x_t). \tag{A.25}$$

By the singular value decomposition, there are $k \times k - q$ matrices A , B and $(k - p) \times (k - p)$ diagonal matrix Λ with positive diagonal elements such that $A'A = I_{k-p}$, $B'B = I_{k-p}$ and

$$I_k - G_T(G'_T G_T)^{-1} G'_T = A\Lambda^{1/2}B'.$$

Thus, we can write

$$J_T^{1/2} = A\Lambda^{1/2}B'S_T^{-1/2} \sum_{t=1}^T z_t(y_t - \hat{\beta}'_T x_t) = A\tilde{J}_T^{1/2}, \tag{A.26}$$

where $\tilde{J}_T^{1/2} = \Lambda^{1/2}B'S_T^{-1/2} \sum_{t=1}^T z_t(y_t - \hat{\beta}'_T x_t)$ is a $(k - p)$ -dimensional vector. Note that

$$J_T = \tilde{J}'_T A' A \tilde{J}_T = \tilde{J}'_T \tilde{J}_T. \tag{A.27}$$

The rest of the proof takes the following steps. First, one can show that Lemmas A.1–A.7 hold for $c'\tilde{J}_T^{1/2}$ except that **a**, **b**, **c**, **d** and **e** now take different values. Second, because the characteristic function of $\tilde{J}_T^{1/2}$ can be derived from an linear combination of $\tilde{J}_T^{1/2}$, the distribution of $\tilde{J}_T^{1/2}$ can be approximated by its Edgeworth expansion in a suitable sense. Lastly, a modification of Theorem 1 of Chandra and Ghosh (1979) with $s = 5$ completes the proof of (3). We make the following modifications: (i) the order of the third term of the Edgeworth expansion $\xi_{s-1,n}(z)$, i.e., $1/n$, is replaced by ℓ/n ; (ii) the order of such approximation errors, $o(n^{-(s-3)/2})$, by $o(\ell/n) + O(1/\ell^q)$ where $n = T$; (iii) the order of the second term of the Edgeworth expansion $\psi_{m,n}$, $1/n$, is replaced by ℓ/n . These modifications do not change the parity of $R(\alpha)$ which is the crucial element of their proof (see Remark 2.5 of Chandra and Ghosh, 1979, pp. 27–28). Thus, their proof will carry through for our version of their theorem. \square

Proof of Theorem 2. For iid observations, a modification of Theorem 1 with $\ell = 1$ yields

$$\sup_{x \in \mathcal{W}} |\mathbb{P}(T^{1/2}(c'\hat{\Sigma}_T c)^{-1/2} c'(\hat{\beta}_T - \beta_0) \leq x) - \Psi_T(x)| = o(T^{-1}), \tag{A.28}$$

$$\sup_{x \geq 0} |\mathbb{P}(J_T \leq x) - \Psi_{J,T}(x)| = o(T^{-1}), \tag{A.29}$$

under Assumptions 1(b)–(d)(i), $\ell = 1$ and Assumption 1(e) replaced by the standard Cramer condition. It suffices to show that the conditions on $R_t = (v'_t, \text{vec}(w_t))'$ required for the Edgeworth expansion of Theorem 1 are also satisfied for $Q_{N_j} = (B'_{N_j}, \text{vec}(F_{N_j}))'$ for $j = 1, \dots, b$ conditionally on the sample $\chi_T = \{(x'_t, y_t, z'_t)\}_{t=1}^T$, uniformly for all χ_T in a set of which the probability tends to 1 as $T \rightarrow \infty$. Without a loss of generality, we check the conditions using B_{N_j} . For Assumption A1(b), we have

$$E^*[B_{N_j}] = E^*[B_{N_1}] = \frac{1}{\sqrt{\ell}} \sum_{i=1}^{\ell} E^*(z_{N_1+i} u_{N_1+i} - \mu_T^*) = 0. \tag{A.30}$$

For Assumption A1(c), it follows from Lemma A.2 that

$$E[E^*|B_{N_j}|^{r+\eta}] = \frac{1}{T - \ell + 1} \sum_{t=0}^{T-\ell} E \left| \frac{1}{\sqrt{\ell}} \sum_{i=1}^{\ell} v_{t+i} \right|^{r+\eta} = E \left| \frac{1}{\sqrt{\ell}} \sum_{i=1}^{\ell} v_{t+i} \right|^{r+\eta} < \infty. \tag{A.31}$$

From the proof of Theorem 4.2 of Götze and Künsch (1996),

$$E^*|B_{N_j}|^{r+\eta} - E[E^*|B_{N_j}|^{r+\eta}] = O_p(b^{-1/2}). \tag{A.32}$$

Combining the two results implies that the probability of $E^*|B_{N_j}|^{r+\eta} < \infty$ tends to unity.

By construction, the moving block bootstrap sample is based on the independent sampling of B_{N_j} . Therefore, Assumption A1(d) is trivially satisfied (with a probability one) using a sigma-field defined by $\sigma(N_j)$ for $j = 1, \dots, b$, conditionally on the sample χ_T . For the same reason, we can replace Assumption A1(e) by the standard Cramér condition and we need only to show that the condition holds with probability tends to one. Using an argument that appeared in the proof of Theorem 4.2 of Götze and Künsch (1996), we have that

$$P \left\{ \sup_{d < |t| < b^{1/2}} |E^* \exp[itB_{N_1}]| \leq 1 - \zeta \right\} = 1 - o(T^{-1}) \tag{A.33}$$

for some $0 < \zeta < \frac{1}{2}$. \square

Proof of Corollary 1. It follows from Lemmas A.3–A.5, Lemmas A.9–A.10 and Theorems 1 and 2. \square

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
 Andrews, D.W.K., 2002a. Higher-order improvements of a computationally attractive k -step bootstrap for extremum estimators. *Econometrica* 70, 119–162.
 Andrews, D.W.K., 2002b. Equivalence of the higher-order asymptotic efficiency of k -step and extremum statistics. *Econometric Theory* 18, 1040–1085.

- Andrews, D.W.K., Monahan, J.C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.
- Brown, B.W., Newey, W.K., 1995. Bootstrapping for GMM. Rice University and MIT.
- Campbell, J.Y., Shiller, R.J., 1991. Yields spread and interest rate movements: a bird's eye view. *Review of Economic Studies* 58, 495–514.
- Carlstein, E., 1986. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics* 14, 1171–1179.
- Chandra, T.K., Ghosh, J.K., 1979. Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhyā A* 41, 22–47.
- Clarida, R., Galí, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. *Quarterly Journal of Economics* 115, 147–180.
- Davison, A.C., Hall, P., 1993. On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* 35, 215–224.
- Gallant, A.R., 1987. *Nonlinear Statistical Models*. Wiley, New York.
- Gonçalves, S., White, H., 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics* 119, 199–219.
- Götze, F., Hipp, C., 1983. Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 211–239.
- Götze, F., Künsch, H.R., 1996. Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914–1933.
- Hahn, J., 1996. A note on bootstrapping generalized method of moments estimators. *Econometric Theory* 12, 187–197.
- Hall, A.R., 1994. Testing for a unit root in time series with pretest data-based model selection. *Journal of Business & Economic Statistics* 12, 461–470.
- Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*. Springer, New York.
- Hall, P., Horowitz, J.L., 1996. Bootstrap critical values for tests based on generalized method of moments estimators. *Econometrica* 64, 891–916.
- Hansen, B.E., 2000. *Edgeworth expansions for the Wald and GMM statistics for nonlinear restrictions*. University of Wisconsin.
- Hansen, L.P., 1982. Large sample properties of generalized method of moment estimators. *Econometrica* 50, 1029–1054.
- Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Lahiri, S.N., 1996. On Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models. *Journal of Multivariate Analysis* 56, 42–59.
- Newey, W.K., West, K.D., 1987. A simple positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix. *Econometrica* 55, 703–708.
- Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.
- Ng, S., Perron, P., 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncated lag. *Journal of the American Statistical Association* 90, 268–281.
- Politis, D.N., Romano, J.P., 1995. Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* 16, 67–103.
- Rothenberg, T.J., 1984. Approximating the distributions of econometric estimators and test statistics. In: Griliches, Z., Intriligator, M.D. (Eds.), *Handbook of Econometrics*, vol. 2. North-Holland, Amsterdam, pp. 881–935.
- Runkle, D.E., 1991. Liquidity constraints and the permanent-income hypothesis. *Journal of Monetary Economics* 27, 73–98.
- Tauchen, G., 1986. Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data. *Journal of Business & Economic Statistics* 4, 397–416.
- West, K.D., 1988. Dividend innovations and stock price volatility. *Econometrica* 56, 37–61.
- White, H., 1984. *Asymptotic Theory for Econometricians*. Academic Press, New York.