Econometrica, Vol. 71, No. 4 (July, 2003), 1049–1082

BOOTSTRAP METHODS FOR MARKOV PROCESSES

BY JOEL L. HOROWITZ[1]

The block bootstrap is the best known bootstrap method for time-series data when the analyst does not have a parametric model that reduces the data generation process to simple random sampling. However, the errors made by the block bootstrap converge to zero only slightly faster than those made by first-order asymptotic approximations. This paper describes a bootstrap procedure for data that are generated by a Markov process or a process that can be approximated by a Markov process with sufficient accuracy. The procedure is based on estimating the Markov transition density nonparametrically. Bootstrap samples are obtained by sampling the process implied by the estimated transition density. Conditions are given under which the errors made by the Markov bootstrap converge to zero more rapidly than those made by the block bootstrap.

KEYWORDS: Time series, resampling, asymptotic refinement, Edgeworth expansion.

1. INTRODUCTION

THIS PAPER DESCRIBES a bootstrap procedure for data that are generated by a (possibly higher-order) Markov process. The procedure is also applicable to non-Markov processes, such as finite-order MA processes, that can be approximated with sufficient accuracy by Markov processes. Under suitable conditions, the procedure is more accurate than the block bootstrap, which is the leading nonparametric method for implementing the bootstrap with time-series data.

The bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data or a model estimated from the data. Under conditions that hold in a wide variety of econometric applications, the bootstrap provides approximations to distributions of statistics, coverage probabilities of confidence intervals, and rejection probabilities of tests that are more accurate than the approximations of first-order asymptotic distribution theory. Monte Carlo experiments have shown that the bootstrap can spectacularly reduce the difference between the true and nominal probabilities that a test rejects a correct null hypothesis (hereinafter the error in the rejection probability or ERP). See Horowitz (1994, 1997, 2001) for examples. Similarly, the bootstrap can greatly reduce the difference between the true and nominal coverage probabilities of a confidence interval (the error in the coverage probability or ECP).

The methods that are available for implementing the bootstrap and the improvements in accuracy that it achieves relative to first-order asymptotic

[1] I thank Kung-Sik Chan, Wolfgang Härdle, Bruce Hansen, Oliver Linton, Daniel McFadden, Whitney Newey, Efstathios Paparoditis, Gene Savin, and two anonymous referees for many helpful comments and suggestions. This research was supported in part by NSF Grant SES-9910925.

approximations depend on whether the data are a random sample from a distribution or a time series. If the data are a random sample, then the bootstrap can be implemented by sampling the data randomly with replacement or by sampling a parametric model of the distribution of the data. The distribution of a statistic is estimated by its empirical distribution under sampling from the data or parametric model (bootstrap sampling). To summarize important properties of the bootstrap when the data are a random sample, let n be the sample size and T_n be a statistic that is asymptotically distributed as $N(0, 1)$ (e.g., a t statistic for testing a hypothesis about a slope parameter in a linear regression model). Then the following results hold under regularity conditions that are satisfied by a wide variety of econometric models. See Hall (1992) for details.

1. The error in the bootstrap estimate of the one-sided probability $P(T_n \leq z)$ is $O_p(n^{-1})$, whereas the error made by first order asymptotic approximations is $O(n^{-1/2})$.

2. The error in the bootstrap estimate of the symmetrical probability $P(|T_n| \leq z)$ is $O_p(n^{-3/2})$, whereas the error made by first-order approximations is $O(n^{-1})$.

3. When the critical value of a one-sided hypothesis test is obtained by using the bootstrap, the ERP of the test is $O(n^{-1})$, whereas it is $O(n^{-1/2})$ when the critical value is obtained from first-order approximations. The same result applies to the ECP of a one-sided confidence interval. In some cases, the bootstrap can reduce the ERP of a one-sided test to $O(n^{-3/2})$ (Hall (1992, p. 178); Davidson and MacKinnon (1999)).

4. When the critical value of a symmetrical hypothesis test is obtained by using the bootstrap, the ERP of the test is $O(n^{-2})$, whereas it is $O(n^{-1})$ when the critical value is obtained from first-order approximations. The same result applies to the ECP of a symmetrical confidence interval.

The practical consequence of these results is that the ERP's of tests and ECP's of confidence intervals based on the bootstrap are often substantially smaller than ERP's and ECP's based on first-order asymptotic approximations. These benefits are available with samples of the sizes encountered in applications (Horowitz (1994, 1997, 2001)).

The situation is more complicated when the data are a time series. To obtain asymptotic refinements, bootstrap sampling must be carried out in a way that suitably captures the dependence structure of the data generation process (DGP). If a parametric model is available that reduces the DGP to independent random sampling (e.g., an ARMA model), then the results summarized above continue to hold under appropriate regularity conditions. See, for example, Andrews (1999) and Bose (1988, 1990). If a parametric model is not available, then the best known method for generating bootstrap samples consists of dividing the data into blocks and sampling the blocks randomly with replacement. This is called the block bootstrap. The blocks, whose lengths increase with increasing size of the estimation data set, may be nonoverlapping (Carlstein (1986), Hall (1985)) or overlapping (Hall (1985), Künsch (1989)). Regardless of the method that is used, blocking distorts the dependence structure of the data and, thereby, increases the error made by the bootstrap. The main results are that under regularity conditions and when the block length is chosen optimally:

1. The errors in the bootstrap estimates of one-sided and symmetrical probabilities are almost surely $O_p(n^{-3/4})$ and $O_p(n^{-6/5})$, respectively (Hall, Horowitz, and Jing (1995)).

2. The ECP's (ERP's) of one-sided and symmetrical confidence intervals (tests) are $O(n^{-3/4})$ and $O(n^{-5/4})$, respectively (Zvingelis (2001)).

Thus, the errors made by the block bootstrap converge to zero at rates that are slower than those of the bootstrap based on data that are a random sample. Monte Carlo results have confirmed this disappointing performance of the block bootstrap (Hall and Horowitz (1996)).

The relatively poor performance of the block bootstrap has led to a search for other ways to implement the bootstrap with dependent data. Bühlmann (1997, 1998), Choi and Hall (2000), Kreiss (1992), and Paparoditis (1996) have proposed a sieve bootstrap for linear processes (that is, AR, vector AR, or invertible MA processes of possibly infinite order). In the sieve bootstrap, the DGP is approximated by an AR(p) model in which p increases with increasing sample size. Bootstrap samples are generated by the estimated AR(p) model. Choi and Hall (2000) have shown that the ECP of a one-sided confidence interval based on the sieve bootstrap is $O(n^{-1+\varepsilon})$ for any $\varepsilon > 0$, which is only slightly larger than the ECP of $O(n^{-1})$ that is available when the data are a random sample. This result is encouraging, but its practical utility is limited. If a process has a finite-order ARMA representation, then the ARMA model can be used to reduce the DGP to random sampling from some distribution. Standard methods can be used to implement the bootstrap, and the sieve bootstrap is not needed. Sieve methods have not been developed for nonlinear processes such as nonlinear autoregressive, ARCH, and GARCH processes.

The bootstrap procedure described in this paper applies to a linear or nonlinear DGP that is a (possibly higher-order) Markov process or can be approximated by one with sufficient accuracy. The procedure is based on estimating the Markov transition density nonparametrically. Bootstrap samples are obtained by sampling the process implied by the estimated transition density. This procedure will be called the Markov conditional bootstrap (MCB). Conditions are given under which:

1. The errors in the MCB estimates of one-sided and symmetrical probabilities are almost surely $O(n^{-1+\varepsilon})$ and $O(n^{-3/2+\varepsilon})$, respectively, for any $\varepsilon > 0$.

2. The ERP's (ECP's) of one sided and symmetrical tests (confidence intervals) based on the MCB are $O(n^{-1+\varepsilon})$ and $O(n^{-3/2+\varepsilon})$, respectively, for any $\varepsilon > 0$.

Thus, under the conditions that are given here, the errors made by the MCB converge to zero more rapidly than those made by the block bootstrap. Moreover for one-sided probabilities, symmetrical probabilities, and one-sided confidence intervals and tests, the errors made by the MCB converge only slightly less rapidly than those made by the bootstrap for data that are sampled randomly from a distribution.

The conditions required to obtain these results are stronger than those required to obtain asymptotic refinements with the block bootstrap. If the required conditions are not satisfied, then the errors made by the MCB may converge more slowly than those made by the block bootstrap. Moreover, as will be explained in Section 3.2, the MCB suffers from a form of the curse of dimensionality of nonparametric estimation. A large data set (e.g., high-frequency financial data) is likely to be needed to obtain good performance if the DGP is a high-dimension vector process or a high-order Markov process. Thus, the MCB is not a replacement for the block bootstrap. The MCB is, however, an attractive alternative to the block bootstrap when the conditions needed for good performance of the MCB are satisfied.

There have been several previous investigations of the MCB. Rajarshi (1990) gave conditions under which the MCB consistently estimates the asymptotic distribution of a statistic. Datta and McCormick (1995) gave conditions under which the error in the MCB estimator of the distribution function of a normalized sample average is almost surely $o(n^{-1/2})$. Hansen (1999) proposed using an empirical likelihood estimator of the Markov transition probability but did not prove that the resulting version of the MCB is consistent or provides asymptotic refinements. Chan and Tong (1998) proposed using the MCB in a test for multimodality in the distribution of dependent data. Paparoditis and Politis (2001, 2002) proposed estimating the Markov transition probability by resampling the data in a suitable way. No previous authors have evaluated the ERP or ECP of the MCB or compared its accuracy to that of the block bootstrap. Thus, the results presented here go well beyond those of previous investigators.

The MCB is described informally in Section 2 of this paper. Section 3 presents regularity conditions and formal results for data that are generated by a Markov process. Section 4 extends the MCB to generalized method of moments (GMM) estimators and approximate Markov processes. Section 5 presents the results of a Monte Carlo investigation of the numerical performance of the MCB. Section 6 presents concluding comments. The proofs of theorems are in the Appendix.

2. INFORMAL DESCRIPTION OF THE METHOD

This section describes the MCB procedure for data that are generated by a Markov process and provides an informal summary of the main results of the paper. For any integer j, let $X_j \in \mathbb{R}^d$ $(d \geq 1)$ be a continuously distributed random variable. Let $\{X_j : j = 1, 2, \ldots, n\}$ be a realization of a strictly stationary, qth order Markov process. Thus,

$$\begin{aligned} &\boldsymbol{P}(X_j \leq x_j | X_{j-1} = x_{j-1}, X_{j-2} = x_{j-2}, \ldots) \\ &\quad = \boldsymbol{P}(X_j \leq x_j | X_{j-1} = x_{j-1}, \ldots, X_{j-q} = x_{j-q}) \end{aligned}$$

almost surely for d-vectors $x_j, x_{j-1}, x_{j-2}, \ldots$ and some finite integer $q \geq 1$. It is assumed that q is known. Cheng and Tong (1992) show how to estimate q. In addition, for technical reasons that are discussed further in Section 3.1, it is assumed that X_j has bounded support and that $\mathrm{cov}(X_j, X_{j+k}) = 0$ if $k > M$ for some $M < \infty$. Define $\mu = \boldsymbol{E}(X_1)$ and $m = n^{-1}\sum_{j=1}^{n} X_j$.

2.1. *Statement of the Problem*

The problem addressed in the remainder of this section and in Section 3 is to carry out inference based on a Studentized statistic, T_n, whose form is

$$(2.1) \qquad T_n = n^{1/2}[H(m) - H(\mu)]/s_n,$$

where H is a sufficiently smooth, scalar-valued function, s_n^2 is a consistent estimator of the variance of the asymptotic distribution of $n^{1/2}[H(m) - H(\mu)]$, and $T_n \to^d N(0, 1)$ as $n \to \infty$. The objects of interest are (i) the probabilities $\boldsymbol{P}(T_n \le t)$ and $\boldsymbol{P}(|T_n| \le t)$ for any finite, scalar t, (ii) the probability that a test based on T_n rejects the correct hypothesis $H_0 : H[\boldsymbol{E}(X)] = H(\mu)$, and (iii) the coverage probabilities of confidence intervals for $H(\mu)$ that are based on T_n. To avoid repetitive arguments, only probabilities and symmetrical hypothesis tests are treated explicitly. An α-level symmetrical test based on T_n rejects H_0 if $|T_n| > z_{n\alpha}$, where $z_{n\alpha}$ is the α-level critical value. Arguments similar to those made in this section and Section 4 can be used to obtain the results stated in the introduction for one-sided tests and for confidence intervals based on the MCB.

The focus on statistics of the form (2.1) with a continuously distributed X may appear to be restrictive, but this appearance is misleading. A wide variety of statistics that are important in applications can be approximated with negligible error by statistics of the form (2.1). In particular, as will be explained in Section 4.1, t statistics for testing hypotheses about parameters estimated by GMM can be approximated this way.[2]

2.2. *The MCB Procedure*

Consider the problem of estimating $\boldsymbol{P}(T_n \le z)$, $\boldsymbol{P}(|T_n| \le z)$, or $z_{n\alpha}$. For any integer $j > q$, define $Y_j = (X'_{j-1}, \ldots, X'_{j-q})'$. Let p_y denote the probability density function of $Y_{q+1} \equiv (X'_q, \ldots, X'_1)'$. Let f denote the probability density function of X_j conditional on Y_j. If f and p were known, then $\boldsymbol{P}(T_n \le z)$ and $\boldsymbol{P}(|T_n| \le z)$ could be estimated as follows:

1. Draw $\widetilde{Y}_{q+1} \equiv (\widetilde{X}'_q \ldots, \widetilde{X}'_1)'$ from the distribution whose density is p_y. Draw $\widetilde{X}_{q+1}$ from the distribution whose density is $f(\cdot|\widetilde{Y}_{q+1})$. Set $\widetilde{Y}_{q+2} = (\widetilde{X}'_{q+1}, \ldots, \widetilde{X}'_2)'$.

[2] Statistics with asymptotic chi-square distributions are not treated explicitly in this paper. However, the arguments made here can be extended to show that asymptotic chi-square statistics based on GMM estimators behave like symmetrical statistics. See, for example, the discussion of the GMM test of overidentifying restrictions in Hall and Horowitz (1996) and Andrews (1999). Under regularity conditions similar to those of Sections 3 and 4, the results stated here for symmetrical probabilities and tests apply to asymptotic chi-square statistics based on GMM estimators. Andrews (1999) defines a class of “minimum ρ estimators” that is closely related to GMM estimators. The results here also apply to t tests based on minimum ρ estimators.

2. Having obtained $\widetilde{Y}_j \equiv (\widetilde{X}'_{j-1}, \ldots, \widetilde{X}'_{j-q})'$ for any $j \geq q+2$, draw $\widetilde{X}_j$ from the distribution whose density is $f(\cdot|\widetilde{Y}_j)$. Set $\widetilde{Y}_{j+1} = (\widetilde{X}'_j, \ldots, \widetilde{X}'_{j-q+1})'$.

3. Repeat step 2 until a simulated data series $\{\widetilde{X}_j : j = 1, \ldots, n\}$ has been obtained. Compute μ as (say) $\int x_1 p_y(x_q, \ldots, x_1) dx_q \cdots dx_1$. Then compute a simulated test statistic $\widetilde{T}_n$ by substituting the simulated data into (2.1).

4. Estimate $\boldsymbol{P}(T_n \leq z)$ $(\boldsymbol{P}(|T_n| \leq z))$ from the empirical distribution of $\widetilde{T}_n(|\widetilde{T}_n|)$ that is obtained by repeating steps 1–3 many times. Estimate $z_{n\alpha}$ by the $1-\alpha$ quantile of the empirical distribution of $|\widetilde{T}_n|$.

This procedure cannot be implemented in an application because f and p_y are unknown. The MCB replaces f and p_y with kernel nonparametric estimators. To obtain the estimators, let K_f be a kernel function (in the sense of nonparametric density estimation) of a $d(q+1)$-dimensional argument. Let K_p be a kernel function of a dq-dimensional argument. Let $\{h_n : n = 1, 2, \ldots\}$ be a sequence of positive constants (bandwidths) such that $h_n \to 0$ as $n \to \infty$. Conditions that K_f, K_p, and $\{h_n\}$ must satisfy are given in Section 3. For $x \in \mathbb{R}^d$, $y \in \mathbb{R}^{dq}$, and $z = (x, y)$, define

$$p_{nz}(x, y) = \frac{1}{(n-q)h_n^{d(q+1)}} \sum_{j=q+1}^{n} K_f\left(\frac{x - X_j}{h_n}, \frac{y - Y_j}{h_n}\right)$$

and

$$p_{ny}(y) = \frac{1}{(n-q)h_n^{dq}} \sum_{j=q+1}^{n} K_p\left(\frac{y - Y_j}{h_n}\right).$$

The estimators of p_y and f, respectively, are p_{ny} and

$$f_n(x \mid y) = p_{nz}(x, y)/p_{ny}(y). \tag{2.2}$$

The MCB estimates $\boldsymbol{P}(T_n \leq z)$, $\boldsymbol{P}(|T_n| \leq z)$, and $z_{n\alpha}$ by repeatedly sampling the Markov process generated by the transition density $f_n(x \mid y)$. However, $f_n(x \mid y)$ is an inaccurate estimator of $f(x \mid y)$ in regions where $p_y(y)$ is close to zero. To obtain the asymptotic refinements described in Section 1, it is necessary to avoid such regions. Here, this is done by truncating the MCB sample. To carry out the truncation, let $C_n = \{y : p_{ny}(y) \geq \lambda_n\}$, where $\lambda_n > 0$ for each $n = 1, 2, \ldots$, and $\lambda_n \to 0$ as $n \to \infty$ at a rate that is specified in Section 3.1. Having obtained realizations $\widehat{X}_1, \ldots, \widehat{X}_{j-1}$ $(j \geq q+1)$ from the Markov process induced by $f_n(x \mid y)$, the MCB retains a realization of $\widehat{X}_j$ only if $(\widehat{X}_j, \ldots, \widehat{X}_{j-q+1}) \in C_n$. Thus, MCB proceeds as follows:

MCB 1. Draw $\widehat{Y}_{q+1} \equiv (\widehat{X}'_q, \ldots, \widehat{X}'_1)'$ from the distribution whose density is p_{ny}. Retain $\widehat{Y}_{q+1}$ if $\widehat{Y}_{q+1} \in C_n$. Otherwise, discard the current $\widehat{Y}_{q+1}$ and draw a new one. Continue this process until a $\widehat{Y}_{q+1} \in C_n$ is obtained.

MCB 2. Having obtained $\widehat{Y}_j \equiv (\widehat{X}'_{j-1}, \ldots, \widehat{X}'_{j-q})'$ for any $j \geq q+1$, draw $\widehat{X}_j$ from the distribution whose density is $f_n(\cdot|\widehat{Y}_j)$. Retain $\widehat{X}_j$ and set

$\widehat{Y}_{j+1} = (\widehat{X}'_j, \ldots, \widehat{X}'_{j-q+1})'$ if $(\widehat{X}_j, \ldots, \widehat{X}_{j-q+1}) \in C_n$. Otherwise, discard the current $\widehat{X}_j$ and draw a new one. Continue this process until an $\widehat{X}_j$ is obtained for which $(\widehat{X}_j, \ldots, \widehat{X}_{j-q+1}) \in C_n$.

MCB 3. Repeat step 2 until a bootstrap data series $\{\widehat{X}_j : j = 1, \ldots, n\}$ has been obtained. Compute the bootstrap test statistic $\widehat{T}_n \equiv n^{1/2}[H(\hat{m}) - H(\hat{\mu})]/\hat{s}_n$, where $\hat{m} = n^{-1}\sum_{j=1}^{n} \widehat{X}_j$, $\hat{\mu}$ is the mean of X relative to the distribution induced by the sampling procedure of steps MCB 1 and MCB 2 (bootstrap sampling), and $\hat{s}_n^2$ is an estimator of the variance of the asymptotic distribution of $n^{1/2}[H(\hat{m}) - H(\hat{\mu})]$ under bootstrap sampling.

MCB 4. Estimate $\boldsymbol{P}(T_n \le z)$ $(\boldsymbol{P}(|T_n| \le z))$ from the empirical distribution of $\widehat{T}_n$ $(|\widehat{T}_n|)$ that is obtained by repeating steps 1–3 many times. Estimate $z_{n\alpha}$ by the $1-\alpha$ quantile of the empirical distribution of $|\widehat{T}_n|$. Denote this estimator by $\hat{z}_{n\alpha}$.

A symmetrical test of H_0 based on T_n and the bootstrap critical value $\hat{z}_{n\alpha}$ rejects at the nominal α level if $|T_n| \ge \hat{z}_{n\alpha}$.

2.3. *Properties of the MCB*

This section presents an informal summary of the main results of the paper and of the arguments that lead to them. The results are stated formally in Section 3. Let $\|\cdot\|$ denote the Euclidean norm. Let $\widehat{\boldsymbol{P}}$ denote the probability measure induced by the MCB sampling procedure (steps MCB 1–MCB 2) conditional on the data $\{X_j : j = l, \ldots, n\}$. Let any $\varepsilon > 0$ be given.

The main results are that under regularity conditions stated in Section 3.1:

$$\sup_z |\widehat{\boldsymbol{P}}(\widehat{T}_n \le z) - \boldsymbol{P}(T_n \le z)| = O(n^{-1+\varepsilon}) \tag{2.3}$$

almost surely,

$$\sup_z |\widehat{\boldsymbol{P}}(|\widehat{T}_n| \le z) - \boldsymbol{P}(|T_n| \le z)| = O(n^{-3/2+\varepsilon}) \tag{2.4}$$

almost surely, and

$$\boldsymbol{P}(|T_n| > \hat{z}_{n\alpha}) = \alpha + O(n^{-3/2+\varepsilon}). \tag{2.5}$$

These results may be contrasted with the analogous ones for the block bootstrap. The block bootstrap with optimal block lengths yields $O_p(n^{-3/4})$, $O_p(n^{-6/5})$, and $O_p(n^{-5/4})$ for the right-hand sides of (2.3)–(2.5), respectively (Hall, Horowitz, and Jing (1995), Zvingelis (2001)). Therefore, the MCB is more accurate than the block bootstrap under the regularity conditions of Section 3.1.

These results are obtained by carrying out Edgeworth expansions of $\boldsymbol{P}(T_n \le z)$ and $\widehat{\boldsymbol{P}}(\widehat{T}_n \le z)$. Additional notation is needed to describe the expansions. Let $\chi \equiv \{X_j\colon j = 1, \ldots, n\}$ denote the data. Let Φ and ϕ, respectively, denote the standard normal distribution function and density. The jth cumulant of T_n $(j \le 4)$ has the form $n^{-1/2}\kappa_j + o(n^{-1/2})$ if j is odd and $I(j = 2) + n^{-1}\kappa_j + o(n^{-1})$ if j is even, where κ_j is a constant (Hall (1992 p. 46)). Define $\kappa = (\kappa_1, \ldots, \kappa_4)'$. Conditional on χ, the jth cumulant of $\widehat{T}_n$ almost surely has the form $n^{-1/2}\hat{\kappa}_j + o(n^{-1/2})$ if j is odd and $I(j=2) + n^{-1}\hat{\kappa}_j + o(n^{-1})$ if j is even. The quantities $\hat{\kappa}_j$ depend on χ. They are nonstochastic relative to bootstrap sampling but are random variables relative to the stochastic process that generates χ. Define $\hat{\kappa} = (\hat{\kappa}_1, \ldots, \hat{\kappa}_4)'$.

Under the regularity conditions of Section 3.1, $\boldsymbol{P}(T_n \le z)$ has the Edgeworth expansion

$$(2.6) \qquad \boldsymbol{P}(T_n \le z) = \Phi(z) + \sum_{j=1}^{2} n^{-j/2}\pi_j(z, \kappa)\phi(z) + O(n^{-3/2})$$

uniformly over z, where $\pi_j(z, \kappa)$ is a polynomial function of z for each κ, a continuously differentiable function of the components of κ for each z, an even function of z if $j = 1$, and an odd function of z if $j = 2$. Moreover, $\boldsymbol{P}(|T_n| \le z)$ has the expansion

$$(2.7) \qquad \boldsymbol{P}(|T_n| \le z) = 2\Phi(z) - 1 + 2n^{-1}\pi_2(z, \kappa)\phi(z) + O(n^{-3/2})$$

uniformly over z. Conditional on χ, the bootstrap probabilities $\widehat{P}(\widehat{T}_n \le z)$ and $\widehat{P}(|\widehat{T}_n| \le z)$ have the expansions

$$(2.8) \qquad \widehat{\boldsymbol{P}}(\widehat{T}_n \le z) = \Phi(z) + \sum_{j=1}^{2} n^{-j/2}\pi_j(z, \hat{\kappa})\phi(z) + O(n^{-3/2})$$

and

$$(2.9) \qquad \widehat{\boldsymbol{P}}(|\widehat{T}_n| \le z) = 2\Phi(z) - 1 + 2n^{-1}\pi_2(z, \hat{\kappa})\phi(z) + O(n^{-3/2})$$

uniformly over z almost surely. Therefore,

$$(2.10) \qquad |\widehat{\boldsymbol{P}}(\widehat{T}_n \le z) - \boldsymbol{P}(T_n \le z)| = O(n^{-1/2}\|\hat{\kappa} - \kappa\|) + O(n^{-1})$$

and

$$(2.11) \qquad |\widehat{\boldsymbol{P}}(|\widehat{T}_n| \le z) - \boldsymbol{P}(|T_n| \le z)| = O(n^{-1}\|\hat{\kappa} - \kappa\|) + O(n^{-3/2})$$

almost surely uniformly over z. Under the regularity conditions of Section 3.1,

$$(2.12) \qquad \|\hat{\kappa} - \kappa\| = O(n^{-1/2+\varepsilon})$$

almost surely for any $\varepsilon > 0$. Results (2.3)–(2.4) follow by substituting (2.12) into (2.10)–(2.11).

To obtain (2.5), observe that $\boldsymbol{P}(|T_n| \le z_{n\alpha}) = \widehat{\boldsymbol{P}}(|\widehat{T}_n| \le \hat{z}_{n\alpha}) = 1 - \alpha$. It follows from (2.7) and (2.10) that

$$(2.13) \qquad 2\Phi(z_{n\alpha}) - 1 + 2n^{-1}\pi_2(z_{n\alpha}, \kappa)\phi(z_{n\alpha}) = 1 - \alpha + O(n^{-3/2})$$

and

$$2\Phi(\hat{z}_{n\alpha}) - 1 + 2n^{-1}\pi_2(\hat{z}_{n\alpha}, \hat{\kappa})\phi(\hat{z}_{n\alpha}) = 1 - \alpha + O(n^{-3/2}) \tag{2.14}$$

almost surely. Let v_α denote the $1-\alpha/2$ quantile of the N(0,1) distribution. Then Cornish-Fisher inversions of (2.13) and (2.14) (e.g., Hall (1992, p. 88–89)) give

$$z_{n\alpha} = v_\alpha - n^{-1}\pi_2(v_\alpha, \kappa) + O(n^{-3/2}) \tag{2.15}$$

and

$$\hat{z}_{n\alpha} = v_\alpha - n^{-1}\pi_2(v_\alpha, \hat{\kappa}) + O(n^{-3/2}) \tag{2.16}$$

almost surely. Therefore,

$$\boldsymbol{P}(|T_n| \le \hat{z}_{n\alpha}) = \boldsymbol{P}\{|T_n| \le z_{n\alpha} + n^{-1}[\pi_2(v_\alpha, \hat{\kappa}) - \pi_2(v_\alpha, \kappa)] + O(n^{-3/2})\} \tag{2.17}$$

$$= \boldsymbol{P}[|T_n| \le z_{n\alpha} + O(n^{-1}\|\hat{\kappa} - \kappa\|) + O(n^{-3/2})]. \tag{2.18}$$

Result (2.5) follows by applying (2.12) to the right-hand side of (2.18).

3. MAIN RESULTS

This section presents theorems that formalize results (2.3)–(2.5).

3.1. *Assumptions*

Results (2.3)–(2.5) are established under assumptions that are stated in this section. The proof of the validity of the Edgeworth expansions (2.6)–(2.9) relies on a theorem of Götze and Hipp (1983) and requires certain restrictive assumptions. See Assumption 4 below. It is likely that the expansions are valid under weaker assumptions, but proving this conjecture is beyond the scope of this paper. The results of this section hold under weaker assumptions if the Edgeworth expansions remain valid.

The following additional notation is used. Let p_z denote the probability density function of $Z_{q+1} \equiv (X'_{q+1}, Y'_{q+1})'$. Let $\widehat{\boldsymbol{E}}$ denote the expectation with respect to the distribution induced by bootstrap sampling (steps MCB 1 and MCB 2 of Section 2.2). Define $\widetilde{\Sigma} = \widehat{\boldsymbol{E}}[n(\hat{m} - \hat{\mu})(\hat{m} - \hat{\mu})']$ and $\tilde{s}_n^2 = \nabla H(\hat{\mu})'\widetilde{\Sigma}\nabla H(\hat{\mu})$. For reasons that are explained later in this section, it is assumed that $\boldsymbol{E}(X_1-\mu)\times(X_{1+j}-\mu)'=0$ if $j>M$ for some integer $M<\infty$. Define

$$\Sigma_n=(n-M)^{-1}\sum_{i=1}^{n-M}\left\{(X_i-m)(X_i-m)'+\sum_{j=1}^{M}[(X_i-m)(X_{i+j}-m)'+(X_{i+j}-m)(X_i-m)']\right\},$$

$$\widehat{\Sigma}_n=(n-M)^{-1}\sum_{i=1}^{n-M}\left\{(\widehat{X}_i-\hat{m})(\widehat{X}_i-\hat{m})'+\sum_{j=1}^{M}[(\widehat{X}_i-\hat{m})(\widehat{X}_{i+j}-\hat{m})'+(\widehat{X}_{i+j}-\hat{m})(\widehat{X}_i-\hat{m})']\right\},$$

$$\overline{\Sigma}_n=\widehat{\boldsymbol{E}}\left\{(\widehat{X}_1-\hat{\mu})(\widehat{X}_1-\hat{\mu})'+\sum_{j=1}^{M}[(\widehat{X}_1-\hat{\mu})(\widehat{X}_{1+j}-\hat{\mu})'+(\widehat{X}_{1+j}-\hat{\mu})(\widehat{X}_1-\hat{\mu})']\right\},$$

and $\vartheta_n^2=\nabla H(\hat{\mu})'\overline{\Sigma}_n\nabla H(\hat{\mu})/\tilde{s}_n^2$. Note that ϑ_n^2 can be evaluated in an application because the bootstrap DGP is known. For any $\lambda>0$, define, $\widetilde{C}_\lambda=\{x_q\in\mathbb{R}^d: p_y(x_q,\dots,x_1)\geq\lambda$ for some $x_{q-1},\dots,x_1\}$. Let $\mathscr{B}(\widetilde{C}_\lambda)$ denote the measurable subsets of $\widetilde{C}_\lambda$. Let $\boldsymbol{P}^k(\xi,A)$ denote the k-step transition probability from a point $\xi\in\widetilde{C}_\lambda$ to a set $A\subset B(\widetilde{C}_\lambda)$.

ASSUMPTION 1: $\{X_j: j=1,2,\dots,n; X_j\in\mathbb{R}^d\}$ *is a realization of a strictly stationary, qth order Markov process that is geometrically strongly mixing* (*GSM*).[3]

ASSUMPTION 2: (*i*) *The distribution of* Z_{q+1} *is absolutely continuous with respect to Lebesgue measure*. (ii) *For* $t\in\mathbb{R}^d$ *and each* k *such that* $0<k\leq q$,

$$\lim_{\|t\|\to\infty}\sup\boldsymbol{E}|\boldsymbol{E}[\exp(\iota t'X_j)|X_{j'},|j-j'|\leq k, j\neq j']|<1. \tag{3.1}$$

(iii) *The functions* p_y, p_z, *and* f *are bounded*. (iv) *For some* $\ell\geq 2$, p_y *and* p_z *are everywhere at least* ℓ *times continuously differentiable with respect to any mixture of their arguments*.

ASSUMPTION 3: (i) *H is three times continuously differentiable in a neighborhood of* μ. (ii) *The gradient of H is nonzero in a neighborhood of* μ.

ASSUMPTION 4: (i) X_j *has bounded support*. (ii) *For all sufficiently small* $\lambda>0$, *some* $\varepsilon>0$, *and some integer* $k>0$,

$$\sup_{\xi,\zeta\in\widetilde{C}_\lambda;A\subset\mathscr{B}(\widetilde{C}_\lambda)}|\boldsymbol{P}^k(\xi,A)-\boldsymbol{P}(\zeta,A)|<1-\varepsilon.$$

[3] GSM means that the process is strongly mixing with a mixing parameter that is an exponentially decreasing function of the lag length.

(iii) *For some* $M < \infty$, $\boldsymbol{E}(X_1 - \mu)(X_{1+j} - \mu)' = 0$ *if* $j > M$. (iv) *The sample and bootstrap variance estimators are* $s_n^2 = \nabla H(m)'\Sigma_n \nabla H(m)$, *and* $\hat{s}_n^2 = \vartheta_n^2 \nabla H(\hat{m})'\widehat{\Sigma}_n \nabla H(\hat{m})$. (v) *For* λ_n *as in Assumption* 6, $\boldsymbol{P}[p_y(Y_{q+1}) < \lambda_n | Y_q = y_q] = o[(\log n)^{-1}]$ *as* $n \to \infty$ *uniformly over* y_q *such that* $p_y(y_q) \geq \lambda_n$.

Let K be a bounded, continuous function whose support is $[-1, 1]$ and that is symmetrical about 0. For each integer $j = 0, \ldots, \ell$, let K satisfy

$$\int_{-1}^{1} v^j K(v)dv = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } 1 \leq j < \ell, \\ B_K \text{ (nonzero)} & \text{if } j = \ell. \end{cases}$$

For any integer $J > 0$, let $W^{(j)}$ $(j = 1, \ldots, J)$ denote the jth component of the vector $W \in \mathbb{R}^J$.

ASSUMPTION 5: *Let* $v_f \in \mathbb{R}^{d(q+1)}$ *and* $v_p \in \mathbb{R}^{dq}$. K_f *and* K_p *have the forms*

$$K_f(v_f) = \prod_{j=1}^{d(q+1)} K\big(v_f^{(j)}\big); \quad K_p(v_p) = \prod_{j=1}^{dq} K(v_p^{(j)}).$$

ASSUMPTION 6: (i) *Let* $\eta = 1/[2\ell + d(q+1)]$. *Then* $h_n = c_h n^{-\eta}$ *for some finite constant* $c_h > 0$. (ii) $\lambda_n \propto (\log n)^2 h_n^\ell$.

Assumptions 2(i), 2(ii), and 4 are used to insure the validity of the Edgeworth expansions (2.6)–(2.9). Condition (3.1) is a dependent-data version of the Cramér condition. See Götze and Hipp (1983). Assumptions 2(iii) and 2(iv) insure sufficiently rapid convergence of $\|\hat{\kappa} - \kappa\|$.

Assumptions 4(i)–4(ii) are used to show that the bootstrap DGP is GSM. The GSM property is used to prove the validity of the Edgeworth expansions (2.8)–(2.9). The results of this paper hold when X_j has unbounded support if the expansions (2.8)–(2.9) are valid and $p_y(y)$ decreases at an exponentially fast rate as $\|y\| \to \infty$.[4] Assumption 4(iii) is used to insure the validity of the Edgeworth expansions (2.6)–(2.9). It is needed because the known conditions for the validity of these expansions apply to statistics that are functions of sample moments. Under 4(iii)–4(iv), s_n and T_n are functions of sample moments of X_j. This is not the case if T_n is Studentized with a kernel-type variance estimator (e.g., Andrews (1991), Andrews and Monahan (1992), Newey and West (1987, 1994)). However, under Assumption 1, the smoothing parameter of a kernel variance estimator can be chosen so that the estimator is $n^{1/2-\varepsilon}$-consistent for any $\varepsilon > 0$. Therefore,

[4] If Y has finitely many moments and the required Edgeworth expansions are valid, then the errors made by the MCB increase as the number of moments of Y decreases. If Y has too few moments, then the errors made by the MCB decrease more slowly than the errors made by the block bootstrap.

if the required Edgeworth expansions are valid, the results established here hold when T_n is Studentized with a kernel variance estimator.[5]

The quantity ϑ_n^2 in $\hat{s}_n^2$ is a correction factor analogous to that used by Hall and Horowitz (1996) and Andrews (1999, 2002). It is needed because $\nabla H(\hat{m})'\widehat{\Sigma}_n \nabla H(\hat{m})$ is a biased estimator of the variance of the asymptotic distribution of $n^{1/2}[H(\hat{m}) - H(\hat{\mu})]$. The bias converges to zero too slowly to enable the MCB to achieve the asymptotic refinements described in Section 2.3. The correction factor removes the bias without distorting higher-order terms of the Edgeworth expansion of the CDF of $\widehat{\boldsymbol{P}}(\widehat{T}_n \le z)$.

Assumption 4(v) restricts the tail thickness of the transition density function.[6] Assumption 5 specifies convenient forms for K_f and K_p, which may be higher-order kernels. Higher-order kernels are used to insure sufficiently rapid convergence of $\|\hat{\kappa} - \kappa\|$.

3.2. *Theorems*

This section gives theorems that establish conditions under which results (2.3)–(2.5) hold. The bootstrap is implemented by carrying out steps MCB 1–MCB 4. Let $\delta = \ell/[2\ell + d(q+1)]$.

THEOREM 3.1: *Let Assumptions* 1–6 *hold. Then for every* $\nu > 0$.

$$\sup_z |\widehat{\boldsymbol{P}}(\widehat{T}_n \le z) - \boldsymbol{P}(T_n \le z)| = O(n^{-1/2-\delta+\nu}) \tag{3.2}$$

and

$$\sup_z |\widehat{\boldsymbol{P}}(|\widehat{T}_n| \le z) - \boldsymbol{P}(|T_n| \le z)| = O(n^{-1-\delta+\nu}) \tag{3.3}$$

almost surely.

THEOREM 3.2: *Let Assumptions* 1–6 *hold. For any* $\alpha \in (0,1)$ *let* $\hat{z}_{n\alpha}$ *satisfy* $\widehat{\boldsymbol{P}}(|\widehat{T}_n| > \hat{z}_{n\alpha}) = \alpha$. *Then for every* $\nu > 0$

$$\boldsymbol{P}(|T_n| > \hat{z}_{n\alpha}) = \alpha + O(n^{-1-\delta+\nu}). \tag{3.4}$$

Theorems 3.1 and 3.2 imply that results (2.3)–(2.6) hold if $\ell > (1-2\varepsilon)d(q+1)/(4\varepsilon)$. With the block bootstrap, the right-hand sides of (3.2)–(3.4) are

[5] Götze and Künsch (1996) have given conditions under which T_n with a kernel-type variance estimator has an Edgeworth expansion up to $O(n^{-1/2})$. Analogous conditions are not yet known for expansions through $O(n^{-1})$. A recent paper by Inoue and Shintani (2000) gives expansions for statistics based on the block bootstrap for linear models with kernel-type variance estimators.

[6] As an example of a sufficient condition for 4(v), let X be scalar with support [0, 1]. Then 4(v) holds if there are finite constants $c_1 > 0$, $c_2 > 0$, $\eta > 0$, and $\varepsilon > 0$ such that $c_1 x_q^\eta \le p_y(x_q, \ldots, x_1) \le c_2 x_q^\eta$ when $x_q < \varepsilon$ and $c_1(1-x_q)^\eta \le p_y(x_q, \ldots, x_1) \le c_2(1-x_q)^\eta$ when $1 - x_q < \varepsilon$. The generalization to multivariate X is analytically straightforward though notationally cumbersome.

$O_p(n^{-3/4})$, $O_p(n^{-6/5})$, and $O(n^{-5/4})$, respectively. The errors made by the MCB converge to zero more rapidly than do those of the block bootstrap if ℓ is sufficiently large. With the MCB, the right-hand side of (3.2) is $o(n^{-3/4})$ if $\ell > d(q+1)/2$, the right-hand side of (3.3) is $o(n^{-6/5})$ if $\ell > d(q+1)/3$, and the right-hand side of (3.4) is $o(n^{-5/4})$ if $\ell > d(q+1)/2$. However, the errors of the MCB converge more slowly than do those of the block bootstrap if the distribution of Z is not sufficiently smooth (ℓ is too small). Moreover, the MCB suffers from a form of the curse of dimensionality in nonparametric estimation. That is, with a fixed value of ℓ, the accuracy of the MCB decreases as d and q increase. Thus, the MCB, like all nonparametric estimators, is likely to be most attractive in applications where d and q are not large. It is possible that this problem can be mitigated, though at the cost of imposing additional structure on the DGP, through the use of dimension reduction methods. For example, many familiar time series DGP's can be represented as single-index models or nonparametric additive models with a possibly unknown link function. However, investigation of dimension reduction methods for the MCB is beyond the scope of this paper.

4. EXTENSIONS

Section 4.1 extends the results of Section 3 to tests based on GMM estimators. Section 4.2 presents the extension to approximate Markov processes.

4.1. *Tests Based on GMM Estimators*

This section gives conditions under which (3.2)–(3.4) hold for the t statistic for testing a hypothesis about a parameter that is estimated by GMM. The main task is to show that the probability distribution of the GMM t statistic can be approximated with sufficient accuracy by the distribution of a statistic of the form (2.1). Hall and Horowitz (1996) and Andrews (1999, 2002) use similar approximations to show that the block bootstrap provides asymptotic refinements for t tests based on GMM estimators.

Denote the sample by $\{X_j : j = 1, \ldots, n\}$. In this section, some components of X_j may be discrete, but there must be at least one continuous component. See Assumption 9 below. Suppose that the GMM moment conditions depend on up to $\tau \geq 0$ lags of X_j. Define $\mathscr{X}_j = (X_j', \ldots, X_{j-\tau}')'$ for some fixed integer $\tau \geq 0$ and $j \geq \tau + 1$. Let $\mathscr{X}$ denote a random vector that is distributed as $\mathscr{X}_{1+\tau}$. Estimation of the $L_\theta \times 1$ parameter θ is based on the moment condition $\boldsymbol{E}G(\mathscr{X}, \theta) = 0$, where G is a known $L_G \times 1$ function and $L_G \geq L_\theta$. Let θ_0 denote the true but unknown value of θ. Assume that $\boldsymbol{E}G(\mathscr{X}_i, \theta_0)G(\mathscr{X}_j, \theta_0)' = 0$ if $|i - j| > M_G$ for some $M_G < \infty$.[7] As in Hall and Horowitz (1996) and Andrews (1999, 2002), two forms of the GMM estimator are considered. One uses a fixed weight matrix, and the other uses an estimator of the asymptotically optimal weight matrix.

[7] This assumption is analogous to Assumption 4(iii) in Section 3.1 and is made for the same reason. The discussion of Assumption 4(iii) applies to the current assumption.

The results stated here can be extended to other forms such as the continuous updating estimator. In the first form, the estimator of θ, θ_n, solves

$$\min_{\theta\in\Theta} J_n(\theta) \equiv \left[n^{-1}\sum_{i=1+\tau}^{n} G(\mathscr{X}_i,\theta)\right]\Omega\left[n^{-1}\sum_{i=1+\tau}^{n} G(\mathscr{X}_i,\theta)\right], \tag{4.1}$$

where Θ is the parameter set and Ω is a $L_G \times L_G$, positive-semidefinite, symmetrical matrix of constants. In the second form, θ_n solves

$$\min_{\theta\in\Theta} J_n(\theta,\tilde{\theta}_n) \equiv \left[n^{-1}\sum_{i=1+\tau}^{n} G(\mathscr{X}_i,\theta)\right]\Omega_n(\tilde{\theta})\left[n^{-1}\sum_{i=1+\tau}^{n} G(\mathscr{X}_i,\theta)\right], \tag{4.2}$$

where $\tilde{\theta}_n$ solves (4.1),

$$\Omega_n(\theta) = \left\{n^{-1}\sum_{i=1+\tau}^{n}\left[G(\mathscr{X}_i,\theta)G(\mathscr{X}_i,\theta)' + \sum_{j=1+\tau}^{M_G} H(\mathscr{X}_i,\mathscr{X}_{i+j},\theta)\right]\right\}^{-1},$$

and $H(\mathscr{X}_i,\mathscr{X}_{i+j},\theta) = G(\mathscr{X}_i,\theta)G(\mathscr{X}_{i+j},\theta)' + G(\mathscr{X}_{i+j},\theta)G(\mathscr{X}_i,\theta)'$.[8]

To obtain the t statistic, let $\Omega_0 = [E\Omega_n(\theta_0)^{-1}]^{-1}$. Define the $L_G \times L_\theta$ matrices $D = E[\partial G(\mathscr{X},\theta_0)/\partial\theta]$ and

$$D_n = n^{-1}\sum_{i=1+\tau}^{n} \partial G(\mathscr{X}_i,\theta_n)/\partial\theta.$$

Define

$$\sigma = (D'\Omega D)^{-1}D'\Omega\Omega_0^{-1}\Omega D(D'\Omega D)^{-1} \tag{4.3}$$

if θ_n solves (4.1) and

$$\sigma = (D'\Omega_0 D)^{-1} \tag{4.4}$$

if θ_n solves (4.2). Let σ_n be the consistent estimator of σ that is obtained by replacing D and Ω_0 in (4.3) and (4.4) by D_n and $\Omega_n(\theta_n)$. In addition, let $(\sigma_n)_{rr}$ be the (r,r) component of σ_n, and let θ_r and θ_{nr} be the rth components of θ and θ_n, respectively. The t statistic for testing $\mathrm{H}_0: \theta_r = \theta_0$ is $t_{nr} = n^{1/2}(\theta_{nr} - \theta_{0r}/(\sigma_n)_{rr}^{1/2}$.

To obtain the MCB version of t_{nr}, let $\{\widehat{X}_j : j = 1,\ldots,n\}$ be a bootstrap sample that is obtained by carrying out steps MCB 1 and MCB 2 but with the modified transition density estimator that is described in equations (4.5)–(4.7) below. The modified estimator allows some components of X_j to be discrete. Define $\widehat{\mathscr{X}}_j = (\widehat{X}_j,\ldots,\widehat{X}_{j-\tau})$. Let $\widehat{\mathscr{X}}$ denote a random vector that is distributed as $\widehat{\mathscr{X}}_{1+\tau}$. Let $\widehat{\boldsymbol{E}}$ denote the expectation with respect to the distribution induced by bootstrap sampling. Define $\widehat{G}(\bullet,\theta) = G(\bullet,\theta) - \widehat{\boldsymbol{E}}G(\widehat{\mathscr{X}},\theta_n)$. The bootstrap version of

[8] As in Section 3, Ω_n is used instead of a kernel-type covariance matrix estimator to insure that the test statistics of interest can be approximated with sufficient accuracy by smooth functions of sample moments. This is not possible with a kernel covariance estimator.

the moment condition $\boldsymbol{E}G(\mathscr{X},\theta)=0$ is $\widehat{\boldsymbol{E}}\widehat{G}(\widehat{\mathscr{X}},\theta)=0$. As in Hall and Horowitz (1996) and Andrews (1999, 2002), the bootstrap version is recentered relative to the population version because, except in special cases, there is no θ such that $\widehat{\boldsymbol{E}}G(\widehat{\mathscr{X}},\theta)=0$ when $L_G > L_\theta$. Brown, Newey, and May (2000) and Hansen (1999) discuss an empirical likelihood approach to recentering. Recentering is unnecessary if $L_G = L_\theta$, but it simplifies the technical analysis and, therefore, is done here.[9]

To form the bootstrap version of t_{nr}, let $\hat{\theta}_n$ denote the bootstrap estimator of θ. Let $\widehat{D}_n$ be the quantity that is obtained by replacing $\mathscr{X}_i$ with $\widehat{\mathscr{X}}_i$ and θ_n with $\hat{\theta}_n$ in the expression for D_n. Define $\overline{D}_n=\widehat{\boldsymbol{E}}\partial G(\widehat{\mathscr{X}}_{1+\tau},\theta_n)/\partial\theta$,

$$\widehat{\Omega}_n(\theta)=\left\{n^{-1}\sum_{i=1+\tau}^{n}\left[\widehat{G}(\widehat{\mathscr{X}}_i,\theta)G(\widehat{\mathscr{X}}_i,\theta)'+\sum_{j=1+\tau}^{M_G}H(\widehat{\mathscr{X}}_i,\widehat{\mathscr{X}}_{i+j},\theta)\right]\right\}^{-1},$$

$$\overline{\Omega}_n(\theta)=\left\{\widehat{\boldsymbol{E}}\left[\widehat{G}(\widehat{\mathscr{X}}_{1+\tau},\theta_n)\widehat{G}(\widehat{\mathscr{X}}_{1+\tau},\theta_n)'+\sum_{j=1+\tau}^{M_G}H(\widehat{\mathscr{X}}_{1+\tau},\widehat{\mathscr{X}}_{1+\tau+j},\theta_n)\right]\right\}^{-1},$$

and

$$\widetilde{\Omega}_n(\theta)=\left\{\widehat{\boldsymbol{E}}\left[\widehat{G}(\widehat{\mathscr{X}}_{1+\tau},\theta_n)\widehat{G}(\widehat{\mathscr{X}}_{1+\tau},\theta_n)'+\sum_{j=1+\tau}^{\infty}H(\widehat{\mathscr{X}}_{1+\tau},\widehat{\mathscr{X}}_{1+\tau+j},\theta_n)\right]\right\}^{-1}.$$

Define $\hat{\sigma}_n$, by replacing G with $\widehat{G}$, $\mathscr{X}_i$ with $\widehat{\mathscr{X}}_i$, θ_n with $\hat{\theta}_n$, D_n with $\widehat{D}_n$, and $\Omega_n(\theta_n)$ with $\widehat{\Omega}_n(\hat{\theta}_n)$ in the formula for σ_n. Let $\Omega_n^*=\Omega$ if θ_n solves (4.1) and $\Omega_n^*=\widehat{\Omega}_n(\theta_n)$ if θ_n solves (4.2). Define

$$\tilde{\sigma}_n=(\overline{D}_n'\Omega_n^*\overline{D}_n)^{-1}\overline{D}_n'\Omega_n^*\overline{\Omega}_n^{-1}\Omega_n^*\overline{D}_n(\overline{D}_n'\Omega_n^*\overline{D}_n)^{-1},$$
$$\bar{\sigma}_n=(\overline{D}_n'\Omega_n^*\overline{D}_n)^{-1}\overline{D}_n'\Omega_n^*\overline{\Omega}_n^{-1}\Omega_n^*\overline{D}_n(\overline{D}_n'\Omega_n^*\overline{D}_n)^{-1},$$

and $\vartheta_{nr}^2=(\bar{\sigma}_n)_{rr}/(\tilde{\sigma}_n)_{rr}$, where $(\tilde{\sigma}_n)_{rr}$ and $(\bar{\sigma}_n)_{rr}$ are the (r,r) components of $\tilde{\sigma}_n$ and $\bar{\sigma}_n$, respectively. Let $\hat{\theta}_{nr}$ denote the rth component of $\hat{\theta}_n$. Then the MCB version of the t statistic is $\hat{t}_{nr}=\vartheta_{nr}n^{1/2}(\hat{\theta}_{nr}-\theta_{nr})/(\hat{\sigma}_n)_{rr}^{1/2}$. The quantity ϑ_{nr} is a correction factor analogous to that used by Hall and Horowitz (1996) and Andrews (1999, 2002).

Now let $V(\mathscr{X}_j,\theta)$ $(j=1+\tau,\dots,n)$ be the vector containing the unique components of $G(\mathscr{X}_j,\theta)$, $G(\mathscr{X}_j,\theta)G(\mathscr{X}_{i+j},\theta)$ $(0\le i\le M_G)$, and the derivatives through order 6 of $G(\mathscr{X}_j,\theta)$ and $G(\mathscr{X}_j,\theta)G(\mathscr{X}_{i+j},\theta)$. Let $S_{\mathscr{X}}$ denote the support of $(X_1,\dots,X_{1+\tau})$. Define p_y, p_z, and f as in Sections 2–3 but with counting measure as the dominating measure for discrete components of X_j $(j=1,\dots,n)$. The following new assumptions are used to derive the results of this section. Assumptions 7–9 are similar to ones made by Hall and Horowitz (1996) and Andrews (1999, 2002).

[9] The formulae required when $L_G=L_\theta$ depend on whether recentering is used. The formulae given here apply only with recentering.

ASSUMPTION 7: *θ_0 is an interior point of the compact parameter set Θ and is the unique solution in Θ to the equation $\boldsymbol{E}G(\mathscr{X}, \theta) = 0$.*

ASSUMPTION 8: (i) *There are finite constants C_G and C_V such that $\|G(\mathscr{X}_{1+\tau}, \theta)\| \leq C_G$ and $\|V(\mathscr{X}_{1+\tau}, \theta)\| \leq C_V$ for all $\mathscr{X}_{1+\tau} \in S_{\mathscr{X}}$ and $\theta \in \Theta$.* (ii) *$\boldsymbol{E}G(\mathscr{X}_{1+\tau}, \theta_0)G(\mathscr{X}_{1+\tau+j}, \theta_0)' = 0$ if $j > M_G$ for some $M_G < \infty$.* (iii)

$$\boldsymbol{E}\left\{G(\mathscr{X}_{1+\tau}, \theta_0)G(\mathscr{X}_{1+\tau+j}, \theta)' + \sum_{j=1}^{M_G}[G(\mathscr{X}_{1+\tau}, \theta)G(\mathscr{X}_{1+\tau+j}, \theta)' + G(\mathscr{X}_{1+\tau+j}, \theta)G(\mathscr{X}_{1+\tau}, \theta)']\right\}$$

exists for all $\theta \in \Theta$. Its smallest eigenvalue is bounded away from 0 *uniformly over θ in an open sphere, N_0, centered on θ_0.* (iv) *There is a bounded function $C_G(\bullet)$ such that*

$$\|G(\mathscr{X}_{1+\tau}, \theta_1) - G(\mathscr{X}_{1+\tau}, \theta_2)\| \leq C_G(\mathscr{X}_{1+\tau})\|\theta_1 - \theta_2\|$$

for all $\mathscr{X}_{1+\tau} \in S_{\mathscr{X}}$ and $\theta_1, \theta_2 \in \Theta$. (v) *$G$ is* 6*-times continuously differentiable with respect to the components of θ everywhere in N_0.* (vi) *There is a bounded function $C_V(\bullet)$ such that*

$$\|V(\mathscr{X}_{1+\tau}, \theta_1) - V(\mathscr{X}_{1+\tau}, \theta_2)\| \leq C_V(\mathscr{X}_{1+\tau})\|\theta_1 - \theta_2\|$$

for all $\mathscr{X}_{1+\tau} \in S_\chi$ and $\theta_1, \theta_2 \in \Theta$.

ASSUMPTION 9: (i) *X_j $(j = 1, \ldots, n)$ can be partitioned $(X_j^{(c)'}, X_j^{(d)'})'$, where $X_j^{(c)} \in \mathbb{R}^d$ for some $d \geq 1$, the distributions of $X_j^{(c)}$ and $\partial G(\mathscr{X}, \theta_0)/\partial\theta$ are absolutely continuous with respect to Lebesgue measure, and the distribution of $X_j^{(d)}$ is discrete with finitely many mass points. There need not be any discrete components of X_j, but there must be at least one continuous component.* (ii) *The functions p_y, p_z, and f are bounded.* (iii) *For some integer $\ell > 2$, p_y and p_z are everywhere at least ℓ times continuously differentiable with respect to any mixture of their continuous arguments.*

ASSUMPTION 10: *Assumptions* 2 *and* 4 *hold with $V(\mathscr{X}_j, \theta_0)$ in place of X_j.*

As in Sections 2–3, $\{\widehat{X}_j: j = 1, \ldots, n\}$ in the MCB for GMM is a realization of the stochastic process induced by a nonparametric estimator of the Markov transition density. If X_j has no discrete components, then the density estimator is (2.2) and MCB samples are generated by carrying out steps MCB 1 and MCB 2 of Section 2.2. A modified transition density estimator is needed if X_j has one or more discrete components. Let $(Y_j^{(c)'}, Y_j^{(d)'})'$ be the partition of Y_j into continuous and discrete components. The modified density estimator is

$$(4.5) \qquad \bar{f}_n(x|y) = \bar{g}_n(x, y)/\bar{p}_{ny}(y),$$

where

$$\bar{g}_n(x,y) = \frac{1}{(n-q)h_n^{d(q+1)}} \sum_{j=q+1}^{n} K_f\left(\frac{x^{(c)} - X_j^{(c)}}{h_n}, \frac{y^{(c)} - Y_j^{(c)}}{h_n}\right) \times I\big(x^{(d)} = X_j^{(d)}\big) I\big(y^{(d)} = Y_j^{(d)}\big), \tag{4.6}$$

$d = \dim(X_j^{(c)})$, and

$$\bar{p}_{ny}(y) = \frac{1}{(n-q)h_n^{dq}} \sum_{j=q+1}^{n} K_p\left(\frac{y^{(c)} - Y_j^{(c)}}{h_n}\right) I\big(y^{(d)} - Y_j^{(d)}\big). \tag{4.7}$$

The result of this section is given by the following theorem.

THEOREM 4.1: *Let Assumptions* 1, 4(i), 4(v), *and* 5–10 *hold*. *For any* $\alpha \in (0,1)$ *let* $\hat{z}_{n\alpha}$ *satisty* $\widehat{\boldsymbol{P}}(|\hat{t}_{nr}| > \hat{z}_{n\alpha}) = \alpha$. *Then* (3.2)–(3.4) *hold with* t_{nr} *and* $\hat{t}_{nr}$ *in place of* T_n *and* $\widehat{T}_n$.

4.2. *Approximate Markov Processes*

This section extends the results of Section 3.2 to approximate Markov processes. As in Sections 2–3, the objective is to carry out inference based on the statistic T_n defined in (2.1). For an arbitrary random vector V, let $\boldsymbol{p}(x_j|v)$ denote the conditional probability density of X_j at $X_j = x_j$ and $V = v$. An approximate Markov process is defined to be a stochastic process that satisfies the following assumption.

ASSUMPTION AMP : (i) $\{X_j\colon j = 0, \pm 1, \pm 2, \ldots; X_j \in \mathbb{R}^d\}$ *is strictly stationary and GSM*. (ii) *For some finite* $b > 0$, *integer* $q_0 > 0$, *all finite* j, *and all* $q \geq q_0$,

$$\sup_{x_j, x_{j-1}, \ldots} |\boldsymbol{p}(x_j|x_{j-1}, x_{j-2}, \ldots) - \boldsymbol{p}(x_j|x_{j-1}, x_{j-2}, \ldots, x_{j-q})| < e^{-bp}.$$

Assumption AMP is satisfied, for example, by the MA(1) process $Y_j = U_j + \beta U_{j-1}$, where $|\beta| < 1$ and U_j is iid with mean zero and bounded support.[10]

The MCB for an approximate Markov process (hereinafter abbreviated AMCB) is the same as the MCB except that the order q of the estimated Markov process (2.2) increases at the rate $(\log n)^2$ as $n \to \infty$. The estimated transition density (2.2) is calculated as if the data were generated by a true Markov process of order $q \propto (\log n)^2$. The AMCB is implemented by carrying out steps MCB 1–MBC 4 with the resulting estimated transition density. Because q increases very slowly as $n \to \infty$, a large value of q is not necessarily required to obtain good finite-sample performance with the AMCB. Section 5 provides an illustration.

[10] As in Section 3, the assumption of bounded support can be relaxed at the cost of additional technical complexity if the required Edgeworth expansions are valid.

To formalize the properties of the AMCB, let $\{X_j^{(q)}\}$ be the Markov process that is induced by the relation

$$(4.8)\qquad \boldsymbol{P}\big(X_j^{(q)} \le x_j | X_{j-1}^{(q)} = x_{j-1}, \ldots, X_{j-q}^{(q)} = x_{j-q}\big) \\ = \boldsymbol{P}(X_j \le x_j | X_j = x_{j-1}, \ldots, X_{j-q} = x_{j-q}).$$

Define $Y_j^{(q)} = (X_{j-1}^{(q)\prime}, \ldots, X_{j-q}^{(q)\prime})'$ and $Z_{q+1}^{(q)} \equiv (X_{q+1}^{(q)\prime}, Y_{q+1}^{(q)\prime})'$. Let $p_y^{(q)}$ and $p_z^{(q)}$, respectively, denote the probability density functions of $Y_{q+1}^{(q)}$ and $Z_{q+1}^{(q)}$. Let $f^{(q)}$ denote the probability density function of $X_j^{(q)}$ conditional on $Y_j^{(q)}$. Let f denote the probability density of X_j conditional on $X_{j-1}, X_{j-2}, \ldots$. To accommodate a Markov process whose order increases with increasing n, Assumptions 2, 5, and 6 are modified as follows. In these assumptions, q increases as $n \to \infty$ in such a way that $q/(\log n)^2 \to c$ for some finite constant $c > 0$, and $\{\ell_n\}$ is a sequence of positive, even integers satisfying $\ell_n \to \infty$ and $d(q+1)/\ell_n \to 0$ as $n \to \infty$.

ASSUMPTION 2′: *For some finite integer* n_0 *and each* $n \ge n_0$: (i) *The distribution of* $Z_{q+1}^{(q)}$ *is absolutely continuous with respect to Lebesgue measure*. (ii) *For* $t \in \mathbb{R}^d$ *and each* k *such that* $0 < k \le q$

$$(3.1)\qquad \lim_{\|t\|\to\infty} \sup \boldsymbol{E}\big|\boldsymbol{E}\big[\exp\big(\iota t' X_j^{(q)}\big)\big| X_{j'}^{(q)}, |j-j'| \le k, j \ne j'\big]\big| < 1.$$

(iii) *The functions* $p_y^{(q)}, p_z^{(q)}, f^{(q)}$, *and* f *are bounded*. (iv) *The functions* $p_y^{(q)}$ *and* $p_z^{(q)}$, *are everywhere* ℓ_n *times continuously differentiable with respect to any mixture of their arguments*.

For each positive integer, n, let K_n be a bounded, continuous function whose support is $[-1, 1]$, that is symmetrical about 0, and that satifies

$$\int_{-1}^{1} v^j K_n(v)\, dv = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } 1 \le j < \ell_n, \\ B_K \text{ (nonzero)} & \text{if } j = \ell_n. \end{cases}$$

ASSUMPTION 5′: *Let* $v_f \in \mathbb{R}^{d(q+1)}$ *and* $v_p \in \mathbb{R}^{dq}$. *For each finite integer* n, K_f *and* K_p *have the forms*

$$K_f(v_f) = \prod_{j=1}^{d(q+1)} K_n(v_f^{(j)}); \quad K_p(v_p) = \prod_{j=1}^{dq} K_n(v_p^j).$$

ASSUMPTION 6′: (i) *Let* $\eta = 1/[2\ell_n + d(q+1)]$. *Then* $h_n = c_h n^{-\eta}$ *for some finite constant* $c_h > 0$. (ii) $\lambda_n \propto (\log n)^2 h_n^{\ell_n}$.

The main difference between these assumptions and Assumptions 2, 5, and 6 of the MCB is the strengthened smoothness Assumption 2′(iv). The MCB for an order q Markov process provides asymptotic refinements whenever p_y and p_z have derivatives of fixed order $\ell > d(q+1)$, whereas the AMCB requires the existence of derivatives of all orders as $n \to \infty$.

The ability of the AMCB to provide asymptotic refinements is established by the following theorem.

THEOREM 4.2: *Let Assumptions AMCB*, 2′, 3, 4, 5′, *and* 6′ *hold with* $q/(\log n)^2 \to c$ *for some finite* $c > 0$. *For any* $\alpha \in (0,1)$ *let* $\hat{z}_{n\alpha}$ *satisfy* $\widehat{\boldsymbol{P}}(|\widehat{T}_n| > \hat{z}_{n\alpha}) = \alpha$. *Then for every* $\nu > 0$

$$\sup_z |\widehat{\boldsymbol{P}}(\widehat{T}_n \le z) - \boldsymbol{P}(T_n \le z)| = O(n^{-1+\nu}),$$

$$\sup_z |\widehat{\boldsymbol{P}}(|\widehat{T}_n| \le z) - \boldsymbol{P}(|T_n| \le z)| = O(n^{-3/2+\nu}),$$

almost surely, and

$$\boldsymbol{P}(|T_n| > \hat{z}_{n\alpha}) = \alpha + O(n^{-3/2+\nu}).$$

5. MONTE CARLO EXPERIMENTS

This section describes four Monte Carlo experiments that illustrate the numerical performance of the MCB. The number of experiments is small because the computations are very lengthy.

Each experiment consists of testing the hypothesis H_0 that the slope coefficient is zero in the regression of X_j on X_{j-1}. The coefficient is estimated by ordinary least squares (OLS), and acceptance or rejection of H_0 is based on the OLS t statistic. The experiments evaluate the empirical rejection probabilities of one-sided and symmetrical t tests at the nominal 0.05 level. Results are reported using critical values obtained from the MCB, the block bootstrap, and first-order asymptotic distribution theory. Four DGP's are used in the experiments. Two are the ARCH(1) processes

$$X_j = U_j(1 + 0.3X_{j-1}^2)^{1/2}, \tag{5.1}$$

where $\{U_j\}$ is an iid sequence that has either the $N(0,1)$ distribution or the distribution with

$$\boldsymbol{P}(U_j \le u) = 0.5[\sin^7(\pi u/2) + 1]I(|u| \le 1). \tag{5.2}$$

The other two DGP's are the GARCH(1,1) processes

$$X_j = U_j h_j^{1/2}, \tag{5.3}$$

where

$$h_j = 1 + 0.4(h_{j-1} + X_{j-1}^2) \tag{5.4}$$

and $\{U_j\}$ is an iid sequence with either the $N(0,1)$ distribution or the distribution of (5.2). DGP (5.1) is a first-order Markov process. DGP (5.3)–(5.4) is an approximate Markov process. In the experiments reported here, this DGP is approximated by a Markov process of order $q = 2$. When U_j has the admittedly somewhat artificial distribution (5.2), X_j has bounded support as required by Assumption 4. When $U_j \sim N(0,1)$, X_j has unbounded support and X_j^2 has moments only through orders 8 and 4 for models (5.1) and (5.3)–(5.4), respectively (He and Teräsvirta (1999)). Therefore, the experiments with $U_j \sim N(0,1)$ illustrate the performance of the MCB under conditions that are considerably weaker than those of the formal theory.

The MCB was carried out using the 4th-order kernel

$$K(v) = (105/64)(1 - 5v^2 + 7v^4 - 3v^6)I(|v| \le 1). \tag{5.4}$$

Implementation of the MCB requires choosing the bandwidth parameter h_n. Preliminary experiments showed that the Monte Carlo results are not highly sensitive to the choice of h_n, so a simple method motivated by Silverman's (1986) rule-of-thumb is used. This consists of setting h_n equal to the asymptotically optimal bandwidth for estimating the $(q+1)$-variate normal density $N(0, \sigma_n^2 I_{q+1})$, where I_{q+1} is the $(q+1) \times (q+1)$ identity matrix and σ_n^2 is the estimated variance of X_1. Of course, there is no reason to believe that this h_n is optimal in any sense in the MCB setting. The preliminary experiments also indicated that the Monte Carlo results are insensitive to the choice of trimming parameter, so trimming was not carried out in the experiments reported here.

Implementation of the block bootstrap requires selecting the block length. Data-based methods for selecting block lengths in hypothesis testing are not available, so results are reported here for three block lengths, (2, 5, 10). The experiments were carried out in GAUSS using GAUSS random number generators. The sample size is $n = 50$. There are 5000 Monte Carlo replications in each experiment. MCB and block bootstrap critical values are based on 99 bootstrap samples.[11]

The results of the experiments are shown in Table I. The differences between the empirical and nominal rejection probabilities (ERP's) with first-order asymptotic critical values tend to be large. The symmetrical and lower-tail tests reject the null hypothesis too often. The upper-tail test does not reject the null hypothesis often enough when the innovations have the distribution (5.2). The ERP's with block bootstrap critical values are sensitive to the block length. With some block lengths, the ERP's are small, but with others they are comparable to or larger than the ERP's with asymptotic critical values. With the MCB, the ERP's are smaller than they are with asymptotic critical values in 10 of the 12 experiments. The MCB has relatively large ERP's with the GARCH(1, 1) model and normal innovations because this DGP lacks the higher-order moments needed to obtain good accuracy with the bootstrap even with iid data.

[11] Bootstrap samples were generated by applying the inverse-distribution method to a fine grid of points. The computations are slow because the transition probability from each sampled grid point to every other point must be estimated nonparametrically.

TABLE I
RESULTS OF MONTE CARLO EXPERIMENTS

Critical Value	Block Length	Empirical Rejection Probability[a] $U \sim$ as in (5.2) ARCH(1)	$U \sim$ as in (5.2) GARCH(1,1)	$U \sim N(0,1)$ ARCH(1)	$U \sim N(0,1)$ GARCH(1,1)
		Symmetrical Tests			
Asymptotic		0.089	0.063	0.092	0.099
Block Boot.	2	0.069	0.054	0.076	0.081
	5	0.075	0.068	0.073	0.073
	10	0.072	0.074	0.073	0.058
MCB		0.044	0.048	0.054	0.067
		One-Sided Upper Tail Tests			
Asymptotic		0.032	0.035	0.058	0.064
Block Boot.	2	0.049	0.042	0.067	0.110
	5	0.085	0.052	0.085	0.091
	10	0.087	0.074	0.092	0.056
MCB		0.038	0.050	0.064	0.073
		One-Sided Lower-Tail Tests			
Asymptotic		0.092	0.075	0.091	0.093
Block Boot.	2	0.067	0.055	0.069	0.092
	5	0.078	0.058	0.085	0.084
	10	0.082	0.059	0.088	0.062
MCB		0.046	0.040	0.055	0.068

[a]Standard errors of the empirical rejection probabilities are in the range 0.0025 to 0.0042, depending on the probability.

6. CONCLUSIONS

The block bootstrap is the best known method for implementing the bootstrap with time series data when one does not have a parametric model that reduces the DGP to simple random sampling. However, the errors made by the block bootstrap converge to zero only slightly faster than those made by first-order asymptotic approximations. This paper has shown that the errors made by the MCB converge to zero more rapidly than those made by the block bootstrap if the DGP is a Markov or approximate Markov process and certain other conditions are satisfied. These conditions are stronger than those required by the block bootstrap. Therefore, the MCB is not a substitute for the block bootstrap, but the MCB is an attractive alternative to the block bootstrap when the MCB's stronger regularity conditions are satisfied. Further research could usefully investigate the possibility of developing bootstrap methods that are more accurate than the block bootstrap but impose less a priori structure on the DGP than do the MCB or the sieve bootstrap for linear processes.

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2600, U.S.A.; joel-horowitz@northwestern.edu: http://www.faculty.econ.northwestern.edu/faculty/horowitz

Manuscript received May, 2001; final revision received July, 2002.

MATHEMATICAL APPENDIX

This Appendix presents the proofs of the theorems stated in the text. To minimize the complexity of the notation, the proofs are given only for $d = 1$. The proofs for $d \geq 2$ are similar but require notation that is more complex and lengthy.

A.1. *Preliminary Lemmas*

This section states lemmas that are used to prove Theorems 3.1 and 3.2. Lemmas 1–11 establish properties of a truncated version of the DGP. Lemmas 12–14 establish properties of the bootstrap DGP. The main result is Lemma 14, which establishes the rate of convergence of $\|\hat{\kappa} - \kappa\|$. This result is used in Section A.2 to prove Theorems 3.1 and 3.2. See Section 2.3 for an informal outline. Assumptions 1–6 hold throughout. "Almost surely" is abbreviated "a.s." Define $C_{nj}^* = \{x_j : x_j, \ldots, x_{j-q+1} \in C_n\}$. A tilde over a probability density function (e.g., $\tilde{f}$) denotes the density of a truncated random variable whose density function without truncation is f. Define $f_n(x|y)$ and $p_{ny}(y)$ as in (2.2). The transition density of the bootstrap DGP is

$$\tilde{f}_n(x_j|y_j) = \frac{f_n(x_j|y_j) I(x_j, \ldots, x_{j-q+1} \in C_n)}{\Pi_n(C_{nj}^*|x_{j-1}, \ldots, x_{j-q})},$$

where

$$\Pi_n(C_{nj}^*|y_t) = \int_{C_{nj}^*} f_n(x|y_j) dx.$$

The initial bootstrap observation is sampled from the distribution whose density is

$$\tilde{p}_{ny}(x_q, \ldots, x_1) = \frac{p_{ny}(x_q, \ldots, x_1) I(x_q, \ldots, x_1 \in C_n)}{\Pi_n(C_n)},$$

where

$$\Pi_n(C_n) = \int_{C_n} p_{ny}(x_q, \ldots, x_1)\, dx_q \cdots dx_1.$$

Conditional on the data $\{X_i\colon i = 1, \ldots, n\}$, define

$$\tilde{f}(x_j|y_j) = \frac{f(x_j|y_j) I(x_j \in C_{nj}^*)}{\Pi(C_{nj}^*|y_j)},$$

where

$$\Pi(C_{nj}^*|y_j) = \int_{C_{nj}^*} f(x|y_j)\, dx$$

and

$$\tilde{p}_y(x_q, \ldots, x_1) = \frac{p_y(x_q, \ldots, x_1) I(x_q, \ldots, x_1 \in C_n)}{\Pi(C_n)},$$

where

$$\Pi(C_n) = \int_{C_n} p_y(x_q, \ldots, x_1)\, dx_q \cdots dx_1.$$

Also define

$$\tilde{p}_{nz}(x_j, y_j) = \frac{p_{nz}(x_j, y_j)I(y_{j+1} \in C_n)}{\Pi_n(C_n)},$$

$$\tilde{p}_{z}(x_j, y_j) = \frac{p_{z}(x_j, y_j)I(y_{j+1} \in C_n)}{\Pi(C_n)},$$

$$\beta_{ny} = \sup_y |p_{ny}(y) - p_y(y)|,$$

and

$$\beta_{nz} = \sup_{x,y} |p_{nz}(x, y) - p_z(x, y)|.$$

LEMMA 1: *For any $c > 1$, $\beta_{nz} = O[(\log n)^c h_n^\ell]$ a.s., and $\beta_{ny} = o(\beta_{nz})$ a.s.*

PROOF: This is a slightly modified version of Theorem 2.2 of Bosq (1996) and is proved the same way as that theorem. *Q.E.D.*

LEMMA 2: $\Pi_n(C_n) - \Pi(C_n) = O(\beta_{ny})$.

PROOF: C_n is bounded uniformly over n because X has bounded support. Therefore,

$$\Pi_n(C_n) - \Pi(C_n) = \int_{C_n} [p_{ny}(y) - p_y(y)]dy = O(\beta_{ny}). \qquad Q.E.D.$$

LEMMA 3: $\Pi(C_n) = 1 - o(\lambda_n)$ *a.s.*

PROOF: Let $V(2\lambda_n)$ denote volume of $\{y : p_y(y) < 2\lambda_n\}$. By Assumption 6(ii) and Lemma 1, $\beta_{ny} = o[\lambda_n(\log n)^{c-2}]$ a.s. for any $c > 1$. Therefore, $\beta_{ny} = o(\lambda_n)$ a.s., and

$$\begin{aligned} 1 - \Pi(C_n) &= \int I[p_{ny}(y) < \lambda_n]p_y(y)dy \\ &\leq 2\lambda_n \int I[p_y(y) < 2\lambda_n]dy = 2\lambda_n V(2\lambda_n) \quad \text{a.s.} \end{aligned}$$

$V(2\lambda_n) \to 0$ as $n \to \infty$, because p_y is bounded away from 0 on compact subsets of the interior of its support and $p_y(y) \to 0$ as y approaches a boundary of the support from the interior. *Q.E.D.*

LEMMA 4:

$$\sup_{y:\ p_{ny}(y) \geq \lambda_n} |p_y(y)[\Pi_n(C_{nj}^*|y) - \Pi(C_{nj}^*|y)]| = O(\beta_{nz}) \qquad a.s.$$

PROOF: Some algebra shows that $p_y(y)[\Pi_n(C_{nj}^*|y) - \Pi(C_{nj}^*|y)] = A_n + B_n$, where

$$A_n = [p_y(y)/p_{ny}(y)] \int_{C_{nj}^*} [p_{nz}(x_j, \ldots, x_{j-q}) - p_z(x_j, \ldots, x_{j-q})]dx_j,$$

$y = (x_{j-1}, \ldots, x_{j-q})$, and

$$B_n = -\int_{C_{nj}^*} p_z(x_j, \ldots, x_{j-q})p_{ny}(y)^{-1}[p_{ny}(y) - p_y(y)]\, dx_j \cdots dx_{j-q+1}.$$

An application of the delta method and boundedness of the support of X give $A_n = O(\beta_{nz})$ and $B_n = O(\beta_{ny})$. *Q.E.D.*

LEMMA 5:

$$\sup_{x,y} |\tilde{p}_{nz}(x,y) - p_z(x,y)| = O(\beta_{nz})$$

and

$$\sup_{x,y:\ p_{ny}(y)\geq\lambda_n} |\tilde{f}_n(x|y) - \tilde{f}(x|y)| = O(\beta_{nz}/\lambda_n).$$

PROOF: These results follow from Lemmas 2 and 4 and the delta method. *Q.E.D.*

For integers j and k with $j \leq k$, let $U \equiv \{X_j, X_{j+1}, \ldots, X_k\}$ be random variables that are generated by the truncated Markov process whose transition density conditional on the original data is $\tilde{f}$. Let g_y and g_{ny} denote the stationary densities of q consecutive realizations of the truncated DGP and the bootstrap DGP, respectively. Let $\{r_n\}$ be a sequence of strictly positive integers for which $r_n/\log n$ is bounded as $n \to \infty$. Define $k = j + r_n$ and

$$\mathscr{A}_n = \{x_j, \ldots, x_k : p_{ny}(y_{j+q}) \geq \lambda_n, \ldots, p_{ny}(y_{k+1}) \geq \lambda_n\},$$

where $y_j = (x_{j-1}, \ldots, x_{j-q})$. Let $\overline{\mathscr{A}_n}$ denote the complement of $\mathscr{A}_n$. Define

$$\psi(u) = \prod_{i=j}^{k} u_i^{\nu_i},$$

where $\nu_i\, (i = j, \ldots, k)$ is a nonnegative integer and $1 \leq \sum_{i=j}^{k} \nu_i \leq 4$. Let $\widetilde{\boldsymbol{E}}$ and $\boldsymbol{E}$, respectively, denote the expectation operator under sampling from the truncated Markov process (conditional on the data) and the nontruncated process induced by f. Define $\tilde{r}_n = \max(0, r_n - q + 1)$. Let $\Pi^*(y_j) = \Pi(C_{nj}^*|y_j)$. A result will be called "uniform over ν" if it holds for all ν_i $(i = j, \ldots, k)$ satisfying $1 \leq \sum_{i=j}^{k} \nu_i \leq 4$. Lemmas 6–13 establish the rate of convergence to zero of the difference between bootstrap and population expectations of $\psi(U)$. The rate of convergence result is used in Lemma 14 to establish the rate of convergence of $\|\hat{\kappa} - \kappa\|$.

LEMMA 6: *Conditional on the data, the support of* Y_j $(j = 1, 2, \ldots)$ *in the truncated Markov process is* C_n *a.s.*

PROOF: The support of Y_1 is C_n by construction, and the support of Y_j $(j \geq 2)$ can be no larger than C_n. For $j \geq 2$, suppose that the support of Y_{j-1} is C_n but the support of Y_j is a proper subset of C_n. Then there is a measurable set $B \subset C_n$ such that $\boldsymbol{P}(B) > 0$ and $\boldsymbol{P}(B|y) = 0$ for all $y \in C_n$. Let $V(B)$ denote the volume of B. Then by Lemma 1 and Assumption 6(ii),

$$\text{(A1)}\qquad \begin{aligned} \boldsymbol{P}(B) &= \int_B p_y(y)\,dy \\ &= \int_B p_{ny}(y)\,dy + \int_B [p_y(y) - p_{ny}(y)]\,dy \\ &\geq (\lambda_n - \beta_{ny})V(B) \geq c_B \lambda_n V(B) \quad \text{a.s.} \end{aligned}$$

for all sufficiently large n and some $c_B < \infty$. Let $\overline{C}_n$ denote the complement of C_n. Let $f_y(\cdot|y_{j-1})$ denote the probability density of Y_j conditional on $Y_{j-1} = y_{j-1}$. Then for some constant $M_B < \infty$, Lemma 3 yields

$$\begin{aligned} \boldsymbol{P}(B) &= \int_{\overline{C}_n} dy \int_B dx f_y(x|y) p_y(y) + \int_{C_n} dy \int_B dx f_y(x|y) p_y(y) \\ &= \int_{\overline{C}_n} dy \int_B dx f_y(x|y) p_y(y) \\ &\leq M_B V(B) \int_{\overline{C}_n} p_y(y)\,dy = V(B) o(\lambda_n). \end{aligned}$$

This is a contradiction of (A1). *Q.E.D.*

LEMMA 7: *Define*

$$\tau_{n1}=\int_{\mathcal{A}_n}\psi(u)\left[\prod_{i=j+q}^{k}f(x_i|y_i)\right]\left[\frac{1-\prod_{i=j+q}^{k}\Pi^*(y_i)}{\prod_{i=j+q}^{k}\Pi^*(y_i)}\right]p_y(y_{j+q})\,dx_j\cdots dx_k.$$

Then conditional on the data, $\tau_{n1}=O(r_n\lambda_n)$ *uniformly over* ν *a.s. as* $n\to\infty$.

PROOF: Let $\omega_n=\sup\{1-\Pi^*(y):y\in C_n\}$. Some algebra and Lemma 6 imply that

$$|\tau_{n1}|\le B(1-\omega_n)^{-\tilde r_n}\int_{\mathcal{A}_n}\left[\prod_{i=j+q}^{k}f(x_i|y)\right]\left[1-\prod_{i=j+q}^{k}\Pi^*(y_i)\right]p_y(y_{j+q})\,dx_j\ldots dx_k,$$

for some constant $B<\infty$. Therefore, for some possibly different $B<\infty$,

$$|\tau_{n1}|\le B(1-\omega_n)^{-\tilde r_n}\int_{\mathcal{A}_n}\left[1-\prod_{i=j+q}^{k}\Pi^*(y_i)\right]p(x_j,\ldots,x_k)\,dx_j\cdots dx_k.$$

Let $\widetilde{\Pi}^*(y_j)=1-\Pi^*(y_j)$. Let w be a multi-index with $\tilde r_n$ components, each of which is 0 or 1. Let $|w|=\sum_i w_i$. Then

$$\begin{aligned}\prod_{i=j+q}^{k}\Pi^*(y_i)&=\prod_{i=j+q}^{k}[1-\widetilde{\Pi}^*(y_i)]\\&=1-\sum_{w:|w|\ge 1}\prod_{i=j+q}^{k}\widetilde{\Pi}^*(y_i)^{w_i}(-1)^{|w|},\end{aligned}$$

and

$$\begin{aligned}1-\prod_{i=j+q}^{k}\Pi^*(y_i)&=-\sum_{i=j+q}^{k}\widetilde{\Pi}^*(y_i)+\sum_{w:|w|\ge 2}\prod_{i=j+q}^{k}\widetilde{\Pi}^*(y_i)^{w_i}(-1)^{|w|}\\&\equiv -C_{n1}+C_{n2}.\end{aligned}$$

But

$$\begin{aligned}\int_{\mathcal{A}_n}C_{n1}p(x_j,\ldots,x_k)\,dx_j\cdots dx_k&\le\int C_{n1}p(x_j,\ldots,x_k)\,dx_j\cdots dx_k\\&=\tilde r_n\boldsymbol{P}(\overline{C}_n)=O(r_n\lambda_n)\end{aligned}$$

a.s. uniformly over ν. Now each value of $|w|$ generates $\tilde r_n!/[|w|!(\tilde r_n-|w|!)]$ terms in C_{n2}. Each term of the integral of C_{n2} is bounded by $\lambda_n w_n^{|w|-1}$ a.s. Therefore,

$$\begin{aligned}\int_{\mathcal{A}_n}C_{n2}p(x_j,\ldots,x_k)\,dx_j\cdots dx_k&\le\lambda_n\sum_{|w|\ge 2}\frac{\tilde r_n!w_n^{|w|-1}}{|w|!(r_n-|w|)!}\\&=\lambda_n\omega_n^{-1}[(1+\omega_n)^{\tilde r_n}-1-\tilde r_n\omega_n]\\&\le 0.5\lambda_n\tilde r_n^2\omega_n+o(r_n^2\omega_n)\quad\text{a.s.}\end{aligned}$$

uniformly over ν. It follows that

$$|\tau_{n1}|\le B(1-\omega_n)^{\tilde r_n}[O(\tilde r_n\lambda_n)+\lambda_nO(r_n^2\omega_n)]=O(r_n\lambda_n)\quad\text{a.s.}$$

uniformly over ν. By Assumption 4(v), $\omega_n=o[(\log n)^{-1}]$, and the result follows. *Q.E.D.*

LEMMA 8: *As $n \to \infty$ and conditional on the data,*

(A2) $$\int_{C_n} |g_y(y) - p_y(y)|dy = O(\lambda_n \log n) \quad a.s.$$

and

(A3) $$\int_{C_n} |g_{ny}(y) - p_{ny}(y)|dy = O(\lambda_n \log n) \quad a.s.$$

PROOF: Only (A2) is proved here. The proof of (A3) is similar. Let $\mathcal{B}$ denote the measurable subsets of C_n. Let $\boldsymbol{G}_y$ and $\boldsymbol{P}_y$ denote the probability measures associated with g_y and p_y. Then

$$\int_{C_n} |g_y(y) - p_y(y)|dy = \sup_{\mathcal{A}\subset\mathcal{B}}[\boldsymbol{G}_y(A) - \boldsymbol{P}_y(A)] - \inf_{\mathcal{A}\subset\mathcal{B}}[\boldsymbol{G}_y(A) - \boldsymbol{P}_y(A)]$$
$$\leq 2\sup_{\mathcal{A}\subset\mathcal{B}}|\boldsymbol{G}_y(A) - \mathbf{P}_y(A)|.$$

Thus, it suffices to investigate the convergence of

$$\sup_{\mathcal{A}\subset\mathcal{B}}|\boldsymbol{G}_y(A) - \boldsymbol{P}_y(A)|.$$

This is done here for the case $q = 1$. The argument for the case of $q > 1$ is similar but more complex notationally. Let s be the integer part of $b \log n$ for some $b > 0$. The s-step transition probability from a point sampled randomly from $\boldsymbol{P}_y$ to a measurable set A is

$$\boldsymbol{P}^{(s)}(A) = \int_A dx_S \int dx_1 \cdots dx_{s-1}\left[\prod_{i=2}^{s} f(x_i|x_{i-1})\right]p_y(x_1)$$
$$= \int_A dx_S \int_{\mathcal{A}_n} dx_1 \cdots dx_{s-1}\left[\prod_{i=2}^{s} f(x_i|x_{i-1})\right]p_y(x_1)$$
$$+ \int_A dx_s \int_{\overline{\mathcal{A}}_n} dx_1 \cdots dx_{s-1}\left[\prod_{i=2}^{s} f(x_i|x_{i-1})\right]p_y(x_1).$$

The s-step transition probability of the truncated process starting from the same initial density but restricted to $\mathcal{A}_n$is

$$\boldsymbol{G}^{(s)}(A) = \Pi(C_n)^{-1}\int_A dx_s \int_{\mathcal{A}_n} dx_1 \cdots dx_{s-1}\left[\prod_{i=2}^{s} \tilde{f}(x_i|x_{i-1})\right]p_y(x_1).$$

Therefore, $\boldsymbol{G}^{(s)}(A) - \boldsymbol{P}^{(s)}(A) = S_{n1} - S_{n2}$, where

$$S_{n1} = \int_A dx_s \int_{\mathcal{A}_n} dx_1 \cdots dx_{s-1}\left[\prod(C_n)^{-1}\prod_{i=2}^{s}\tilde{f}(x_i|x_{i-1}) - \prod_{i=2}^{s} f(x_i|x_{i-1})\right]p_y(x_1)$$

and

$$S_{n2} = \int_A dx_s \int_{\overline{\mathcal{A}}_n} dx_1 \ldots dx_{s-1}\left[\prod_{i=2}^{s} f(x_i|x_{i-1})\right]p_y(x_1).$$

Arguments like those made in the proof of Lemma 7 show that $S_{n1} = O(s\lambda_n)$ a.s. uniformly over $A \subset \mathcal{B}$. In addition $S_{n2} \leq c\boldsymbol{P}(\overline{\mathcal{A}}_n)$ for some constant $c < \infty$. But

$$\overline{\mathcal{A}_n} \subset \bigcup_{i=1}^{s-1}\{x_i \notin C_n\}$$

so

$$P(\overline{\mathscr{A}_n}) \leq \sum_{i=1}^{s-1}[1-\Pi(C_n)] = O(s\lambda_n) \quad \text{a.s.}$$

It follows that

$$\text{(A4)} \qquad G^{(s)}(A) - P^{(s)}(A) = O(s\lambda_n) \quad \text{a.s.}$$

uniformly over $A \subset \mathscr{B}$. Similar arguments show that the s step transition probabilities from a point $\xi \in C_n$ to $A \subset \mathscr{B}$ satisfy

$$\text{(A5)} \qquad G^{(s)}(\xi, A) - P^{(s)}(\xi, A) = O(s\lambda_n) \quad \text{a.s.}$$

uniformly over $A \subset \mathscr{B}$ and $\xi \in C_n$. In addition,

$$\text{(A6)} \qquad \sup_{A\subset\mathscr{B}} |G_y(A) - P_y(A)| \leq \sup_{A\subset\mathscr{B}} |G^{(s)}(A) - P^{(s)}(A)| + \sup_{A\subset\mathscr{B}} |G^{(s)}(A) - G_y(A)| + \sup_{A\subset\mathscr{B}} |P^{(s)}(A) - P_y(A)|.$$

The first term on the right-hand side of (A6) is $O(s\lambda_n)$ a.s. By a result of Nagaev (1961), it follows from (A5) and Assumption 4(ii) that for some $\delta < 1$, the second term is a.s. $O(\delta^{s/k-1})$, where k is as in Assumption 4(ii) (Nagaev (1961)). The third term is $O(\rho^s)$ for some $\rho < 1$ because $\{X_j\}$ is GSM and, therefore, uniformly ergodic (Doukhan (1995, p. 21)). For b sufficiently large, the second two terms are $o(\lambda_n)$. *Q.E.D.*

LEMMA 9: *Define*

$$\tau_{n2} = \int_{\mathscr{A}_n} \psi(u)\left[\prod_{i=j+q}^{k} \tilde{f}(x_j|y_j)\right][g_y(y_{j+q}) - p_y(y_{j+q})]\, dx_j \cdots dx_k.$$

Then conditional on the data, $\tau_{n2} = O(\lambda_n \log n)$ *a.s. uniformly over* ν *as* $n \to \infty$.

PROOF:

$$\tau_{n2} \leq B\int_{C_n} |g_y(y) - p_y(y)| dy$$

for some constant $B < \infty$. The result now follows from Lemma 8. *Q.E.D.*

LEMMA 10: *Define*

$$\tau_{n3} = \int_{\overline{A}_n} \psi(u)\left[\prod_{i=j+q}^{k} f(x_i|y_i)\right] p_y(y_{j+q})\, dx_j \cdots dx_k.$$

Then conditional on the data, $\tau_{n3} = O(r_n\lambda_n)$ *a.s. uniformly over* ν *as* $n \to \infty$.

PROOF: $\tau_{n3} \leq BP(\overline{\mathscr{A}_n})$ for some constant $B < \infty$. The result now follows from arguments like those used to prove Lemma 8. *Q.E.D.*

LEMMA 11: *Conditional on the data,* $\widetilde{E}[\psi(U)] - E[\psi(U)] = O(\lambda_n \log n)$ *a.s. uniformly over* ν.

PROOF: Some algebra show that $\widetilde{E}[\psi(U)] - E[\psi(U)] = \tau_{n1} + \tau_{n2} + \tau_{n3}$. Now combine Lemmas 7, 9, and 10. *Q.E.D.*

LEMMA 12: *The bootstrap process* $\{\widehat{X}_j\}$ *is geometrically strongly mixing a.s.*

PROOF: It follows from Lemma 5 and Assumption 4(ii) that

$$\sup_{\xi,\zeta\in C_n;A\subset\mathscr{B}(C_n)} |\widehat{\boldsymbol{P}}^m(\xi, A)-\widehat{\boldsymbol{P}}^m(\zeta, A)| < 1-\varepsilon \quad \text{a.s.}$$

for all sufficiently large n and some $\varepsilon > 0$. The lemma follows from a result of Nagaev (1961). *Q.E.D.*

LEMMA 13: *Let $\widehat{\boldsymbol{E}}$ denote expectation relative to $\widehat{\boldsymbol{P}}$. Then conditional on the data, $\widehat{\boldsymbol{E}}[\psi(U)]-\boldsymbol{E}[\psi(U)] = O[\lambda_n(\log n)^4]$ a.s. uniformly over ν.*

PROOF: $\widehat{\boldsymbol{E}}[\psi(U)]-\boldsymbol{E}[\psi(U)] = \{\widehat{\boldsymbol{E}}[\psi(U)]-\widetilde{\boldsymbol{E}}[\psi(U)]\}+\{\widetilde{\boldsymbol{E}}[\psi(U)]-\boldsymbol{E}[\psi(U)]\}$, so

$$\widehat{\boldsymbol{E}}[\psi(U)]-\boldsymbol{E}[\psi(U)] = \widehat{\boldsymbol{E}}[\psi(U)]-\widetilde{E}[\psi(U)]+O(\lambda_n \log n) \quad \text{a.s.} \tag{A7}$$

uniformly over ν by Lemma 11. Define $\Delta_n = \widehat{E}[\psi(U)]-\widetilde{E}[\psi(U)]$. Then by (A3),

$$\Delta_n = \int_{\mathscr{A}_n} \psi(u)\left\{\left[\prod_{i=j+q}^{k} \tilde{f}_n(x_i|y_i)\right]g_{ny}(y_{j+q})-\left[\prod_{i=j+q}^{k} \tilde{f}(x_i|y_i)\right]g_y(y_{j+q})\right\} dx_j\cdots dx_k+O(\lambda_n \log n)$$

a.s. Let w be a multi-index with $\tilde{r}_n$ components, each of which is 0 or 1. Let $|w| = \sum_i w_i$. Then

$$\begin{aligned}\prod_{i=j+q}^{k} \tilde{f}_n(x_i|y_i) &= \sum_w \prod_{i=j+q}^{k} \{\tilde{f}_n(x_i|y_i)^{1-w}[\tilde{f}_n(x_i|y_i)-\tilde{f}(x_i|y_i)]^{w_i}\} \\ &= \prod_{i=j+q}^{k} \tilde{f}(x_i|y_i)+\sum_{l=j+q}^{k}\left[\prod_{\substack{i=j+q\\ i\neq l}}^{k} \tilde{f}(x_i|y_i)\right][\tilde{f}_n(x_l|y_l)-\tilde{f}(x_l|y_l)]+S_n,\end{aligned}$$

where

$$S_n = \sum_{|w|\geq 2}\prod_{i=j+q}^{k} \{\tilde{f}(x_i|y_i)^{1-w_i}[\tilde{f}_n(x_i|y_i)]-\tilde{f}(x_i|y_i)]^{w_i}\}.$$

Therefore,

$$\Delta_n = \tau_{n1}+\tau_{n2}+\tau_{n3}+\tau_{n4},$$

where

$$\tau_{n1} = \int_{\mathscr{A}_n} \psi(u)\left[\prod_{i=j+q}^{k} \tilde{f}(x_i|y_i)\right][g_{ny}(y_{j+q})-g_y(y_{j+q})]\, dx_j\cdots dx_k,$$

$$\tau_{n2} = \int_{\mathscr{A}_n} \psi(u)\sum_{l=j+q}^{k}\left[\prod_{\substack{i=j+q\\ i\neq l}}^{k} \tilde{f}(x_i|y_i)\right][\tilde{f}_n(x_l|y_l)-\tilde{f}(x_l|y_l)]g_y(y_{j+q})\, dx_j\cdots dx_k,$$

$$\tau_{n3} = \int_{\mathscr{A}_n} \psi(u)\sum_{l=j+q}^{k}\left[\prod_{\substack{i=j+q\\ i\neq l}}^{k} \tilde{f}_n(x_i|y_i)\right][\tilde{f}_n(x_l\mid y_l)-\tilde{f}(x_l|y_l)][g_{ny}(y_{j+q})-g_y(y_{j+q})]\, dx_j\cdots dx_k,$$

and

$$\tau_{n4} = \int_{\mathscr{A}_n} \psi(u)S_n g_{ny}(y_{j+q})dx_j,\ldots,dx_k.$$

It follows from Lemmas 5 and 8 that $\tau_{n1} = O(\lambda_n \log n)$ and $\tau_{n3} = O(\lambda_n \log n)$ a.s. uniformly over ν. Now consider τ_{n2}. A Taylor series approximation shows that

$$\tau_{n2} = \sum_{l=j+q+1}^{k} \left(\tau_{nl2}^{(1)} + \tau_{nl2}^{(2)} + \tau_{nl2}^{(3)}\right),$$

where

$$\tau_{nl2}^{(1)} = \int_{\mathscr{A}_n} \psi(u) \left[\prod_{\substack{i=j+q \\ i \neq l}}^{k} \tilde{f}(x_i|y_i)\right] \frac{p_{nz}(x_l, y_l) - p_z(x_l, y_l)}{p_y(y_l) \Pi(C_n^*|y_l)} I(x_l \in C_{nl}^*) g_y(y_{j+q})\, dx_j \cdots dx_k,$$

$$\tau_{nl2}^{(2)} = \int_{\mathscr{A}_n} \psi(u) \left[\prod_{\substack{i=j+q \\ i \neq l}}^{k} \tilde{f}(x_i|y_i)\right] \frac{f(x_l|y_l)[p_{ny}(y_l) - p_y(y_l)]}{p_y(y_l) \Pi(C_n^*|y_l)} I(x_l \in C_{nl}^*) g_y(y_{j+q})\, dx_j \cdots dx_k,$$

and

$$\tau_{nl2}^{(3)} = \int_{\mathscr{A}_n} \psi(u) \left[\prod_{\substack{i=j+q \\ i \neq l}}^{k} \tilde{f}(x_i|y_i)\right] \left[\frac{\tilde{f}(x_l \mid y_l)}{p_y(y_l) \Pi(\overline{C}_n^*|y_l)} O(\beta_{nz}) + O(\beta_{ny})\right] g_y(y_{j+q})\, dx_j \cdots dx_k.$$

For some constant $B < \infty$,

$$\begin{aligned}
\tau_{nl2}^{(1)} &\leq B\beta_{nz} \int \left[\prod_{\substack{i=j+q \\ i \neq l}}^{k} \tilde{f}(x_i|y_i)\right] \frac{1}{p_y(y_l)} g_y(y_{j+q})\, dx_j \cdots dx_k \\
&= B\beta_{nz} \int \left[\prod_{i=j+q}^{l-1} \tilde{f}(x_i|y_i)\right] \frac{1}{p_y(y_l)} g_y(y_{j+q})\, dx_j \cdots dx_{l-1} \\
&= B\beta_{nz} \int \left[\prod_{i=j+q}^{l-1} \tilde{f}(x_i|y_i)\right] \frac{1}{p_y(y_l)} p_z(y_{j+q})\, dx_j \cdots dx_{l-1} + O(\lambda_n \log n)
\end{aligned}$$

a.s. uniformly over ν, where the last line follows from Lemma 8. Some algebra shows that

$$\begin{aligned}
\left[\prod_{i=j+q}^{l-1} \tilde{f}(x_i|y_i)\right] \frac{p_y(y_{j+q})}{p_y(y_l)} &= f(x_{l-q-1}|y_l) \cdots f(x_j|y_{j+q+1}) \left[\prod_{i=j+q}^{l-1} \frac{I(y_{i+1} \in C_{n,i+1}^*)}{\Pi^*(y_i)}\right] \\
&\leq (1-\omega_n)^{l-j-q-2} f(x_{l-q-1}|y_1) \cdots f(x_j|y_{j+q+1}).
\end{aligned}$$

Therefore, $\tau_{nl2}^{(1)} \leq O(\lambda_n \log n)$ a.s. uniformly over ν. Similar arguments can be used to show that $\tau_{n\ell 2}^{(2)} = O(\lambda_n \log n)$ and $\tau_{nl2}^{(3)} = o(\lambda_n \log n)$ a.s. uniformly over ν. Therefore, $\tau_{n2} = O[\lambda_n (\log n)^2]$ a.s. uniformly over ν. Now consider τ_{n4}. Given w, let ℓ_w be the smallest value of i such that $w_i = 1$. Define

$$\delta_n = \sup_{x,\, y:\, p_{ny}(y) \geq \lambda_n} |\tilde{f}_n(x|y) - \tilde{f}(x|y)|.$$

Then

$$S_n \leq B_2 \sum_{|w| \geq 2} \delta_n^{|w|-1} \left[\prod_{i=1}^{\ell_w - 1} \tilde{f}(x_i|y_i)\right] \left[\tilde{f}_n(x_{\ell_w}|y_{\ell_w}) - \tilde{f}_n(x_{\ell_w}|y_{\ell_w})\right]$$

for some constant $B_2 < \infty$. Therefore

$$\tau_{n4} \leq B_2 \left(\sum_{|w| \geq 2} \delta_n^{|w|-1}\right) \int_{\mathscr{A}_n} \left[\prod_{i=1}^{\ell_w - 1} \tilde{f}(x_i|y_i)\right] |\tilde{f}_n(x_{\ell_w}|y_{\ell_w}) - \tilde{f}(x_{\ell_w}|y_{\ell_w})| g_{ny}(y_{j+q})\, dx_j \cdots dx_k.$$

Arguments like those made for τ_{n2} show that the integral is $O[\lambda_n(\log n)^2]$ a.s. Arguments like those made in the proof of Lemma 7 show that the sum is $O(r_n^2\delta_n)$. Therefore, $\tau_{n4} = o[\beta_{nz}(\log n)^4]$ a.s. uniformly over ν. The lemma follows by combining the rates of convergence of $\tau_{n1}, \tau_{n2}, \tau_{n3}$, and τ_{n4}. *Q.E.D.*

LEMMA 14: *Define* $\delta = \ell/[2\ell + d(q+1)]$. *For any* $\varepsilon > 0$, $\|\hat{\kappa} - \kappa\| = O(n^{-\delta+\varepsilon})$ *a.s.*

PROOF: Under Assumptions 3–4, T_n is a smooth function of sample moments. Therefore, the jth cumulant of T_n has the expansion

$$K_j = n^{-(j-2)/2}[k_{j1} + n^{-1}k_{j2} + O(n^{-2})]$$

with $k_{11} = 0$ and $k_{21} = 1$. The vector κ of Section 2.3 can be written $\kappa = (k_{12}, k_{22}, k_{31}, k_{41})$. Then the functions π_1 and π_2 in (2.6) are

$$\pi_1(z, \kappa) = -[k_{12} + (1/16)k_{31}(z^2 - 1)]$$

and

$$\pi_2(z, \kappa) = -z[(1/2)(k_{22} + k_{12}^2) + (1/24)(k_{41} + 4k_{12}k_{31})(z^2 - 3) + (1/72)k_{31}^2(z^4 - 10z^2 + 15)].$$

See, e.g., Hall (1992). The parameters k_{ij} in these expressions are independent of n. Define $\gamma_j = \boldsymbol{E}(X_1 - \mu)(X_{1+j} - \mu)$ and $\gamma_{nj} = n^{-1}\sum_{i=1}^{n-j}(X_i - m)(X_{i+j} - m)$. A Taylor series expansion of T_n about $m = \mu$ and $\gamma_{nj} = \gamma_j$ $(j = 0, \ldots, M)$ shows that κ is a continuous function of terms of the form $D(\mu, \gamma)\boldsymbol{E}n(m-\mu)^2$, $D(\mu, \gamma)\boldsymbol{E}n^2(m-\mu)^3$, $D(\mu, \gamma)\boldsymbol{E}n^2(m-\mu)^4$, $D(\mu, \gamma)\boldsymbol{E}n(m-\mu)(\gamma_{nj} - \gamma_j)$, $D(\mu, \gamma)$, $D(\mu, \gamma)\boldsymbol{E}n^2(m-\mu)^3(\gamma_{nj} - \gamma_j)$, $D(\mu, \gamma)\boldsymbol{E}n^{3/2}(m-\mu)(\gamma_{nj} - \gamma_j)^2$, and $D(\mu, \gamma)\boldsymbol{E}n^2(m-\mu)^2(\gamma_{nj} - \gamma_j)^2$, where D represents a differentiable function that may be different in different terms, and $\gamma = (\gamma_0, \ldots, \gamma_M)$. Similar expansions applied to $\widehat{T}_n$ show that $\hat{\kappa}$ is a continuous function of the same terms but with $\widehat{\boldsymbol{E}}, \hat{m}, \hat{\mu}$, and $\hat{\gamma}_j = \widehat{\boldsymbol{E}}(\widehat{X}_1 - \hat{\mu})(\widehat{X}_{1+j} - \hat{\mu})$ in place of $\boldsymbol{E}, m, \mu$, and γ_j. Thus, it is necessary to show that the differences between the bootstrap and population terms are $o(n^{-\delta+\varepsilon})$ a.s. This is done here for $D(\hat{\mu}, \hat{\gamma})\widehat{\boldsymbol{E}}n(\hat{m}-\hat{\mu})^2 - D(\mu, \gamma)\boldsymbol{E}n(m-\mu)^2$ and $D(\hat{\mu}, \hat{\gamma})\widehat{\boldsymbol{E}}n^2(\hat{m}-\hat{\mu})^3 - D(\mu, \gamma)\boldsymbol{E}n^2(m-\mu)^3$. The calculations for the remaining moments are similar but much lengthier. Lemma 13 implies that $D(\hat{\mu}, \hat{\gamma}) - D(\mu, \gamma) = O(n^{-\delta+\varepsilon})$ a.s., so it suffices to show that $\widehat{\boldsymbol{E}}n(\hat{m}-\hat{\mu})^2 - \boldsymbol{E}n(m-\mu)^2$ and $\widehat{\boldsymbol{E}}n^2(\hat{m}-\hat{\mu})^3 - \boldsymbol{E}n^2(m-\mu)^3$ are $O(n^{-\delta+\varepsilon})$.

First consider $\widehat{\boldsymbol{E}}n(\hat{m}-\hat{\mu})^2 - \boldsymbol{E}n(m-\mu)^2$. Let $\{J_n\}$ be an increasing sequence of positive integers such that $J_n/\log n \to b$ for some finite constant $b > 1/2$. Then it follows from Lemma 12 that

$$\widehat{\boldsymbol{E}}n(\hat{m}-\hat{\mu})^2 = \hat{\gamma}_0 + 2\sum_{j=2}^{J_\eta}(1 - j/n)\hat{\gamma}_j + o(n^{-1/2}) \quad \text{a.s.}$$

It further follows from Lemma 13 that $\widehat{\boldsymbol{E}}n(\hat{m}-\hat{\mu})^2 - \boldsymbol{E}n(m-\mu)^2 = o(n^{-\delta+\varepsilon})$ a.s. for any $\varepsilon > 0$. Now consider $\widehat{\boldsymbol{E}}n^{3/2}(\hat{m}-\hat{\mu})^3 - \boldsymbol{E}n^{3/2}(m-\mu)^3$. Let $V_i = X_i - \mu$ and $\widehat{V}_i = \widehat{X}_i - \hat{\mu}$. Then

$$n^2(\hat{m}-\hat{\mu})^3 = \frac{1}{n}\sum_{i,j,k=1}^{n}\widehat{V}_i\widehat{V}_j\widehat{V}_k$$
$$= \frac{1}{n}\sum_{i=1}^{n}\widehat{V}_i^3 + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\widehat{V}_i^2\widehat{V}_j + \widehat{V}_i\widehat{V}_j^2) + \frac{6}{n}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}\widehat{V}_i\widehat{V}_j\widehat{V}_k.$$

Since the bootstrap DGP is stationary, some algebra shows that

$$\widehat{\boldsymbol{E}}n^2(\hat{m}-\hat{\mu})^3 = \widehat{\boldsymbol{E}}\widehat{V}_1^3 + \sum_{i=1}^{n-1}(1 - i/n)\widehat{\boldsymbol{E}}(\widehat{V}_1^2\widehat{V}_{1+i} + \widehat{V}_1\widehat{V}_{1+i}^2) + 6\sum_{i=1}^{n-2}\sum_{j=1}^{n-i-1}[1 - (i+j)/n]\widehat{\boldsymbol{E}}\widehat{V}_1\widehat{V}_{1+i}\widehat{V}_{1+i+j}.$$

Define J_n as before. By Lemma 12 and Billingsley's inequality (Bosq (1996, inequality (1.11))),

$$\widehat{E}n^2(\hat{m}-\hat{\mu})^3 = \widehat{E}\widehat{V}_1^3 + \sum_{i=1}^{J_n}(1-i/n)\widehat{E}(\widehat{V}_1^2\widehat{V}_{1+i} + \widehat{V}_1\widehat{V}_{1+i}^2) + 6\sum_{i=1}^{J_n}\sum_{j=1}^{\min(n-i-1,J_n)}[1-(i+j)/n]\widehat{E}\widehat{V}_1\widehat{V}_{1+i}\widehat{V}_{1+i+j} + o(n^{-1/2}) \quad \text{a.s.}$$

A similar argument shows that

$$En^2(m-\mu)^3 = EV_1^3 + \sum_{i=1}^{J_n}(1-i/n)E(V_1^2V_{1+i} + V_1V_{1+i}^2) + 6\sum_{i=1}^{J_n}\sum_{j=1}^{\min(n-i-1,J_n)}[1-(i+j)/n]EV_1V_{1+i}V_{1+i+j} + o(n^{1/2}).$$

Now apply Lemma 13. *Q.E.D.*

A.2. Proofs of Theorems 3.1 and 3.2

PROOF OF THEOREM 3.1: Under Assumptions 3–4, T_n is a smooth function of sample moments. Therefore, Theorem 2.8 of Götze and Hipp (1983) and arguments like those used to prove Theorem 2 of Bhattacharya and Ghosh (1978) establish expansions (2.6) and (2.7). Repetition of Götze's and Hipp's proof shows that the expansions (2.8) and (2.9) are also valid. The theorem follows from Lemma 14 and the fact that g_1 and g_2 in (2.6)–(2.9) are polynomial functions of the components of κ and $\hat{\kappa}$. *Q.E.D.*

PROOF OF THEOREM 3.2: Equations (2.13) and (2.14) follow from Theorem 3.1. Cornish-Fisher inversions of (2.13) and (2.14) yield (2.15)–(2.18). The theorem follows by applying Lemma 14 and the delta method to (2.18). *Q.E.D.*

A.3. Proofs of Theorems 4.1–4.2

Assumptions 1 and 5–10 hold throughout this section.

LEMMA 15: *For any* $\varepsilon > 0$,

$$\lim_{n\to\infty} n^{1/2-\varepsilon}\|\theta_n - \theta_0\| = 0 \quad a.s.$$

PROOF: This follows from Lemmas 3 and 4 of Andrews (1999) and the Borel-Cantelli Lemma.

LEMMA 16: *Define* $S_n = n^{-1}\sum_{i=1+\tau}^n V(\mathscr{X}_i, \theta_0)$, $\widehat{S}_n = n^{-1}\sum_{i=1+\tau}^n V(\widehat{\mathscr{X}}_i, \theta_n)$, $S = E(S_n)$, *and* $\widehat{S} = \widehat{E}(\widehat{S}_n)$. *There is an infinitely differentiable function* Γ *such that* $\Gamma(S) = \Gamma(\widehat{S}) = 0$,

$$\lim_{n\to\infty}\sup_z n^{3/2}|P(t_{nr} \le z) - P[n^{1/2}\Gamma(S_n) \le z]| = 0,$$

and

$$\lim_{n\to\infty}\sup_z n^{3/2}|\widehat{P}(\hat{t}_{nr} \le z) - \widehat{P}[n^{1/2}\Gamma(\widehat{S}_n) \le z]| = 0 \quad a.s.$$

PROOF: This is a slightly modified version of Lemma 13 of Andrews (1999) and is proved using the methods of proof of that lemma. *Q.E.D.*

PROOF OF THEOREM 4.1: $\boldsymbol{P}[n^{1/2}\Gamma(S_n) \le z]$ has an Edgeworth expansion. By Lemma 16, $\boldsymbol{P}(t_{nr} \le z)$ has the same expansion to $O(n^{-3/2})$, and

$$\boldsymbol{P}(t_{nr} \le z) = \Phi(z) + \sum_{j=1}^{2} n^{-j/2}\pi_j(z, \kappa)\phi(z) + O(n^{-3/2})$$

uniformly over z, where κ is an infinitely differentiable function of the following moments: $\boldsymbol{E}n(S_n - S)^2$, $\boldsymbol{E}n^2(S_n - S)^3$, and $\boldsymbol{E}n^2(S_n - S)^4$. Similarly,

$$\widehat{P}(\hat{t}_{nr} \le z) = \Phi(z) + \sum_{j=1}^{2} n^{-j/2}\pi_j(z, \hat{\kappa})\phi(z) + O(n^{-3/2})$$

uniformly over z a.s., where $\hat{\kappa}$ is an infinitely differentiable function of $\widehat{\boldsymbol{E}}n(\widehat{S}_n - \widehat{S})^2, \widehat{\boldsymbol{E}}n^2(\widehat{S}_n - \widehat{S})^3$, and $\widehat{\boldsymbol{E}}n^2(\widehat{S}_n - \widehat{S})^4$. It is easy to see that the conclusion of Lemma 13 holds for moments of $V(\mathscr{X}_i, \theta_0)$ and $V(\widehat{\mathscr{X}}_i, \theta_0)$. Lemma 15 implies that $V(\widehat{\mathscr{X}}_i, \theta_n)$ can be replaced with $V(\widehat{\mathscr{X}}_i, \theta_0)$ in moments comprising $\hat{\kappa}$. Therefore, the conclusion of Lemma 14 holds for $\|\hat{\kappa} - \kappa\|$, and the theorem is proved. *Q.E.D.*

PROOF OF THEOREM 4.2: Let $\widetilde{T}_n$ be the version of T_n that is obtained by sampling $\{X_j^{(q)}\}$. Let $\widetilde{\boldsymbol{P}}$ be the probability measure corresponding to $\widetilde{T}_n$. Repetition of the steps leading to the conclusions of Theorems 3.1 and 3.2 shows that for every $\nu > 0$

$$\sup_z |\widehat{\boldsymbol{P}}(\widehat{T}_n \le z) - \widetilde{\boldsymbol{P}}(\widetilde{T}_n \le z)| = o(n^{-1+\nu}),$$

$$\sup_z |\widehat{\boldsymbol{P}}(|\widehat{T}_n| \le z) - \widetilde{\boldsymbol{P}}(|\widetilde{T}_n| \le z)| = o(n^{-3/2+\nu}) \quad \text{a.s.},$$

and

$$\widetilde{\boldsymbol{P}}(|\widetilde{T}_n| > \hat{z}_{n\alpha}) = \alpha + o(n^{-3/2+\nu}).$$

Therefore, it suffices to show that

$$\sup_z |\widetilde{\boldsymbol{P}}(\widetilde{T}_n \le z) - \boldsymbol{P}(T_n \le z)| = (n^{-3/2+\nu}).$$

$\widetilde{\boldsymbol{P}}(\widetilde{T}_n \le z)$ and $\boldsymbol{P}(T_n \le z)$ have Edgeworth expansions, so it is enough to show that $\|\tilde{\kappa} - \kappa\| = O(n^{1/2+\nu})$, where $\tilde{\kappa} = (\tilde{k}_{12}, \tilde{k}_{22}, \tilde{k}_{31}, \tilde{k}_{41})$ and $\tilde{k}_{ij}$ is defined as in Lemma 14 but relative to the distribution of $\widetilde{T}_n$ instead of the distribution of T_n. As in the proof of Lemma 14, $\|\tilde{\kappa} - \kappa\| = O(n^{-1/2+\nu})$ follows if $|\widetilde{\boldsymbol{E}}\psi(u) - \boldsymbol{E}\psi(u)| = O(n^{-1/2+\varepsilon})$ for any $\varepsilon > 0$, where $\widetilde{\boldsymbol{E}}$ denotes the expectation relative to the measure induced by $\{X_j^{(q)}\}$. Let $\omega_j = \{X_{j-1}, X_{j-2}, \ldots\}$. Define $\psi(u)$, r_n, the multi-index w, and $|w|$ as in the proof of Lemma 7. Then for a suitable constant $C_1 < \infty$,

$$\begin{aligned} |\widetilde{\boldsymbol{E}}\psi(u) - \boldsymbol{E}\psi(u)| &\le \int |\psi(u)| \left| \prod_{i=j}^{k} f^{(q)}(x_i|y_i) - \prod_{i=j+q}^{k} f(x_i|\omega_i) \right| dx_j \cdots dx_k d\boldsymbol{P}_w(\omega_j) \\ &\le C_1 \int \left| \prod_{i=j}^{k} f^{(q)}(x_i|y_i) - \prod_{i=j+q}^{k} f(x_i|\omega_i) \right| dx_j \cdots dx_k d\boldsymbol{P}_w(\omega_j) \\ &= C_1 S_n, \end{aligned}$$

where

$$S_n = \sum_{|w|\ge 1} \int \left[\prod_{i=j}^{k} f(x_i|\omega_i)^{1-w_i} |f^{(q)}(x_i|y_i) - f(x_i|\omega_i)|^{w_i} \right] dx_j \cdots dx_k d\boldsymbol{P}_\omega(\omega_j).$$

Each value of $|w|$ generates $r_n!/[|w|!(r_n - |w|)]!$ terms in the sum on the right-hand side of S_n. Each term is bounded by $(C_2 e^{-bq})^{|w|}$ for some constant $C_2 < \infty$. Therefore,

$$S_n \le \sum_{|w|\ge 1} \frac{r_n!}{|w|!(r_n - |w|)!} (C_2 e^{-bq})^{|w|} = (1 + C_2 e^{-bq})^{r_n} - 1.$$

$S_n = O(n^{-1/2+\varepsilon})$ for any $\varepsilon > 0$ follows from $r_n = O(\log n)$ and $q = c(\log n)^2$. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1999): "Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," Cowles Foundation Discussion Paper No. 1230, Cowles Foundation for Research in Economics, Yale University, New Haven, CT.

——— (2002): "Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119–162.

ANDREWS, D. W. K., AND J. C. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 60, 953–966.

BHATTACHARYA, R. N., AND J. K. GHOSH (1978): "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434–451.

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

——— (1990): "Bootstrap in Moving Average Models," *Annals of the Institute of Statistical Mathematics*, 42, 753–768.

BOSQ, D. (1996): *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* New York: Springer-Verlag.

BROWN, B. W., W. K. NEWEY, AND S. M. MAY (2000): "Efficient Bootstrapping for GMM," unpublished manuscript, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.

BÜHLMANN, P. (1997): "Sieve Bootstrap for Time Series," *Bernoulli*, 3, 123–148.

——— (1998): "Sieve Bootstrap for Smoothing in Nonstationary Time Series," *Annals of Statistics*, 26, 48–83.

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.

CHAN, K. S., AND H. TONG (1998): "A Note on Testing for Multi-Modality with Dependent Data," Working Paper, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA.

CHENG, B., AND H. TONG (1992): "Consistent Non-Parametric Order Determination and Chaos," *Journal of the Royal Statistical Society*, Series B, 54, 427–429.

CHOI, E., AND P. HALL (2000): "Bootstrap Confidence Regions Computed from Autoregressions of Arbitrary Order," *Journal of the Royal Statistical Society*, Series B, 62, 461–477.

DATTA, S., AND W. P. MCCORMICK (1995): "Some Continuous Edgeworth Expansions for Markov Chains with Applications to Bootstrap," *Journal of Multivariate Analysis*, 52, 83–106.

DAVIDSON, R., AND J. G. MACKINNON (1999): "The Size Distortion of Bootstrap Tests," *Econometric Theory*, 15, 361–376.

DOUKHAN, P. (1995): *Mixing: Properties and Examples*. New York: Springer-Verlag.

GÖTZE, F., AND C. HIPP (1983): "Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, 211–239.

GÖTZE, F., AND H. R. KÜNSCH (1996): "Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations," *Annals of Statistics*, 24, 1914–1933.

HALL, P. (1985): "Resampling a Coverage Process," *Stochastic Process Applications*, 19, 259–269.

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

HALL, P., AND J. L. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891–916.

HALL, P., J. L. HOROWITZ, AND B.-Y. JING (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

HANSEN, B. (1999): "Non-Parametric Dependent Data Bootstrap for Conditional Moment Models," Working Paper, Department of Economics, University of Wisconsin, Madison, WI.

HE, C., AND T. TERÄSVIRTA (1999): "Properties of Moments of a Family of GARCH Processes," *Journal of Econometrics*, 92, 173–192.

HOROWITZ, J. L. (1994): "Bootstrap-Based Critical Values for the Information-Matrix Test," *Journal of Econometrics*, 61, 395–411.

——— (1997): "Bootstrap Methods in Econometrics," in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, ed. by D. M. Kreps and K. F. Wallis. Cambridge: Cambridge University Press, pp. 188–222.

——— (2001): "The Bootstrap in Econometrics," in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier, Ch. 52, pp. 3159–3228.

INOUE, A., AND M. SHINTANI (2000): "Bootstrapping GMM Estimators for Time Series," unpublished manuscript, Department of Agricultural and Resource Economics, North Carolina State University.

KREISS, J.-P. (1992): "Bootstrap Procedures for AR(∞) Processes," in *Bootstrapping and Related Techniques*, Lecture Notes in Economics and Mathematical Systems, 376, ed. by K. H. Jöckel, G. Rothe, and W. Sender. Heidelberg: Springer-Verlag, pp. 107–113.

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

NAGAEV, S. V. (1961): "More Exact Statements of Limit Theorems for Homogeneous Markov Chains," *Theory of Probability and Its Applications*, 6, 62–81.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

PAPARODITIS, E. (1996): "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes," *Journal of Multivariate Analysis*, 57, 277–296.

PAPARODITIS, E., AND D. N. POLITIS (2001): "A Markovian Local Resampling Scheme for Nonparametric Estimators in Time Series Analysis," *Econometric Theory*, 17, 540–566.

——— (2002): "The Local Bootstrap for Markov Processes," *Journal of Statistical Planning and Inference*, 108, 301–328.

RAJARSHI, M. B. (1990): "Bootstrap in Markov-Sequences Based on Estimates of Transition Density," *Annals of the Institute of Statistical Mathematics*, 42, 253–268.

SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

ZVINGELIS, J. J. (2001): "On Bootstrap Coverage Probability with Dependent Data," in *Computer-Aided Econometrics*, ed. by D. Giles. New York: Marcel Dekker.