

BOOTSTRAP METHODS FOR TIME SERIES

by

Wolfgang Härdle
Institute for Statistics and Econometrics
Humboldt Universität zu Berlin
Berlin, GERMANY

Joel Horowitz
Department of Economics
Northwestern University
Evanston, IL

Jens-Peter Kreiss
Institute for Mathematical Stochastics
Technical University of Braunschweig
Braunschweig, GERMANY

August 2001

Abstract

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. The methods that are available for implementing the bootstrap and the accuracy of bootstrap estimates depend on whether the data are a random sample from a distribution or a time series. This paper is concerned with the application of the bootstrap to time-series data when one does not have a finite-dimensional parametric model that reduces the data generation process to independent random sampling. We review the methods that have been proposed for implementing the bootstrap in this situation and discuss the accuracy of these methods relative to that of first-order asymptotic approximations. We argue that methods for implementing the bootstrap with time-series data are not as well understood as methods for data that are sampled randomly from a distribution. Moreover, the performance of the bootstrap as measured by the rate of convergence of estimation errors tends to be poorer with time series than with random samples. This is an important problem for applied research because first-order asymptotic approximations are often inaccurate and misleading with time-series data and samples of the sizes encountered in applications. We conclude that there is a need for further research in the application of the bootstrap to time series, and we describe some of the important unsolved problems.

The research of Wolfgang Härdle and Jens-Peter Kreiss was supported in part by Deutsche Forschungsgemeinschaft Sonderforschungsbereich 373, "Quantifikation und Simulation Ökonomischer Prozesse." The research of Joel L. Horowitz was supported in part by NSF Grant SES-9910925 and by the Alexander von Humboldt Foundation.

BOOTSTRAP METHODS FOR TIME SERIES

1. Introduction

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 applications, the bootstrap provides approximations to distributions of statistics, coverage probabilities of confidence intervals, and rejection probabilities of tests that are at least as accurate as the approximations of first-order asymptotic distribution theory. Often, the bootstrap provides approximations that are more accurate than those of first-order asymptotic theory.

The methods that are available for implementing the bootstrap and the improvements in accuracy that it achieves relative to first-order asymptotic 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. Beran and Ducharme (1991), Hall (1992), Efron and Tibshirani (1993), and Davison and Hinkley (1997) provide detailed discussions of bootstrap methods and their properties for data that are sampled randomly from a distribution.

The situation is more complicated when the data are a time series because bootstrap sampling must be carried out in a way that suitably captures the dependence structure of the data generation process (DGP). This is not difficult if one has a finite-dimensional parametric model (e.g., a finite-order ARMA model) that reduces the DGP to independent random sampling. In this case and under suitable regularity conditions, the bootstrap has properties that are essentially the same as they are when the data are a random sample from a distribution. See, for example, Andrews (1999) and Bose (1988, 1990).

This paper is concerned with the situation in which one does not have a finite-dimensional parametric model that reduces the DGP to independent random sampling. We review the methods that have been proposed for implementing the bootstrap in this situation and discuss the ability of these methods to achieve asymptotic refinements. We argue that methods for implementing the bootstrap with time-series data are not as well understood as methods for data that are sampled randomly from a distribution. Moreover, the performance of the bootstrap as measured by the order of the asymptotic refinements that are available from known methods tends to be poorer with time series than with random samples. This is an important problem for

applied research because first-order asymptotic approximations are often inaccurate and misleading with time-series data and samples of the sizes encountered in applications. We conclude that there is a need for further research in the application of the bootstrap to time series, and we describe some of the important unsolved problems.

Section 2 of this paper describes the estimation and inference problems that will be discussed in the remainder of the paper. Section 2 also provides background information on the performance of the bootstrap when the data are a random sample from a distribution and on the theory underlying the bootstrap's ability to provide asymptotic refinements. Section 3 reviews the block bootstrap, which is the oldest and best known nonparametric method for implementing the bootstrap with time-series data. The block bootstrap imposes relatively few *a priori* restrictions on the DGP, but this flexibility comes at the price of estimation errors that converge to zero relatively slowly. Section 4 discusses methods that make stronger assumptions about the DGP but offer the possibility of faster converging estimation errors. Section 5 presents conclusions and suggestions for further research. The regularity conditions required by bootstrap methods for time-series tend to be highly technical, and they vary among investigators and methods for implementing the bootstrap. To enable us to concentrate on important ideas rather than technicalities, we do not give detailed regularity conditions in this paper. They are available in the references that are cited.

We assume throughout this paper that the DGP is stationary and weakly dependent. Bootstrap methods for DGP's that do not satisfy this condition, notably long-memory and unit-root processes, are important topics for research but are at a much more preliminary stage of development than are methods for stationary, weakly dependent processes.

2. Problem Definition and Background Information

This section has three parts. Section 2.1 describes the estimation and inference problems that will be treated in the remainder of the paper. Section 2.2 reviews the performance of the bootstrap when the data are a random sample from a distribution. This performance provides a useful benchmark for judging the bootstrap's performance when the data are a time series. Section 2.3 reviews the theory underlying the bootstrap's ability to provide asymptotic refinements.

2.1 Statement of the Problem

Let $\{X_i : i=1, \dots, n\}$ be observations from the sequence $\{X_i : -\infty < i < \infty\}$, where $X_i \in \mathbb{R}^d$ for each integer i and some integer d satisfying $1 \leq d < \infty$. Unless otherwise stated, we assume that $\{X_i\}$ is a realization of a discrete-time stochastic process (the DGP) that is strictly stationary and geometrically strongly mixing (GSM). We also assume that $\mu \equiv \mathbf{E}(X_1)$ exists. Define $m_n = n^{-1} \sum_{i=1}^n X_i$. Let $\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ be a function. In this paper, we are concerned with making inferences about $\theta = \theta(\mu)$ based on the estimator $\theta_n \equiv \theta(m_n)$.

As is discussed by Hall and Horowitz (1996) and Andrews (2002), a wide variety of estimators that are important in applications can be approximated with (higher-order) asymptotically negligible errors by functions of the form $\theta(m_n)$. Thus, the focus on estimators of this form is not highly restrictive. In particular, generalized-method-of-moments (GMM) estimators can be approximated this way under mild regularity conditions (Hall and Horowitz 1996, Andrews 2001). GMM is a method for estimating a possibly vector-valued parameter ψ that is identified by the moment condition $\mathbf{E}g(X, \psi) = 0$, where g is a function whose dimension equals or exceeds that of ψ . The class of GMM estimators includes linear and nonlinear least squares estimators and maximum likelihood estimators, among many others. Hansen (1982) provides details of the GMM estimation method and gives conditions under which GMM estimators are $n^{1/2}$ -consistent and asymptotically normal.

Assume now that $\mathbf{E}\theta(m_n)$ and $\sigma_\theta^2 \equiv \text{Var}[\theta(m_n)]$ exist. Define σ_∞^2 to be the variance of the asymptotic distribution of $n^{1/2}(\theta_n - \theta)$, and let s_n^2 be a weakly consistent estimator of σ_∞^2 . In the remainder of this paper, we discuss the use of the bootstrap to estimate the following quantities, all of which are important in applied research:

1. The bias and variance of θ_n , that is $\mathbf{E}\theta_n - \theta$ and σ_θ^2 .
2. The one-sided distribution functions $\mathbf{P}[n^{1/2}(\theta_n - \theta) \leq \tau]$, $\mathbf{P}[n^{1/2}(\theta_n - \theta)/\sigma_\infty \leq \tau]$, and $\mathbf{P}[n^{1/2}(\theta_n - \theta)/s_n \leq \tau]$ for any real τ .
3. The symmetrical distribution functions $\mathbf{P}[n^{1/2} |(\theta_n - \theta)| \leq \tau]$, $\mathbf{P}[n^{1/2} |(\theta_n - \theta)|/\sigma_\infty \leq \tau]$, and $\mathbf{P}[n^{1/2} |(\theta_n - \theta)|/s_n \leq \tau]$.

4. The coverage probabilities of one-sided and symmetrical confidence intervals for θ . These are $-\infty < \theta \leq \theta_n - \hat{z}_{1-\alpha} s_n$, $\theta_n - \hat{z}_\alpha s_n \leq \theta < \infty$, and $\theta_n - \hat{z}_{\alpha/2} s_n \leq \theta \leq \theta_n + \hat{z}_{\alpha/2} s_n$, where \hat{z}_α is a bootstrap estimator of the $1-\alpha$ quantile of the distribution of $n^{1/2}(\theta_n - \theta)/s_n$.

5. The probabilities that one-sided and symmetrical tests reject the correct null hypothesis $H_0: \theta = \theta_0$. For one-sided tests, the test statistic is $n^{1/2}(\theta_n - \theta_0)/s_n$ with bootstrap critical values \hat{z}_α and $\hat{z}_{1-\alpha}$ for upper- and lower- tail tests, respectively. For a symmetrical test, the test statistic is $n^{1/2} |(\theta_n - \theta)|/s_n$, and the bootstrap critical value is $\hat{z}_{\alpha/2}$.

The conclusions regarding coverage probabilities of confidence intervals and rejection probabilities of tests are identical, so only the rejection probabilities of tests are treated explicitly here.

2.2 Performance of the Bootstrap when the Data Are a Random Sample

The rates of convergence of the errors made by first-order asymptotic approximations and by bootstrap estimators with data that are a random sample provide useful benchmarks for judging the bootstrap's performance with time series. As will be discussed in Sections 3 and 4, the errors made by the bootstrap converge to zero more slowly when the data are a time series than when they are a random sample. In some cases, the rates of convergence for time series data are close to those for a random sample, but in others they are only slightly faster than the rates with first-order approximations.

If the data are a random sample, then under regularity conditions that are given by Hall (1992),

1. The bootstrap estimates σ_θ^2 consistently and reduces the bias of θ_n to $O(n^{-2})$.

That is $E(\theta_n - \theta - \hat{B}_n) = O(n^{-2})$, where \hat{B}_n is the bootstrap estimator of $E(\theta_n - \theta)$. By contrast, $E(\theta_n - \theta) = O(n^{-1})$. Horowitz (2001a) gives an algorithm for computing \hat{B}_n .

If, in addition, $n^{1/2}(\theta_n - \theta)/s_n \rightarrow^d N(0,1)$, then:

2. The errors in the bootstrap estimates of the one-sided distribution functions $P[n^{1/2}(\theta_n - \theta)/\sigma_\infty \leq \tau]$ and $P[n^{1/2}(\theta_n - \theta)/s_n \leq \tau]$ are $O_p(n^{-1})$. The errors made by first order asymptotic approximations are $O(n^{-1/2})$. By ‘‘error’’ we mean the difference between a bootstrap estimator and the population probability that it estimates.

3. The errors in the bootstrap estimates of the symmetrical distribution functions $P[n^{1/2} |(\theta_n - \theta)|/\sigma_\infty \leq \tau]$ and $P[n^{1/2} |(\theta_n - \theta)|/s_n \leq \tau]$ are $O_p(n^{-3/2})$, whereas the errors made by first-order approximations are $O(n^{-1})$.

4. When the bootstrap is used to obtain the critical value of a one-sided hypothesis test, the resulting difference between the true and nominal rejection probabilities under the null hypothesis (error in the rejection probability or ERP) 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 error in the coverage probability (ECP) of a one-sided confidence interval. In some cases, the bootstrap can reduce the ERP (ECP) of a one-sided test (confidence interval) to $O(n^{-3/2})$ (Hall 1992, p. 178; Davidson and MacKinnon 1999).

5. When the bootstrap is used to obtain the critical value of a symmetrical hypothesis test, the resulting ERP 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.

2.3 Why the Bootstrap Provides Asymptotic Refinements

This section outlines the theory underlying the bootstrap's ability to provide asymptotic refinements. To minimize the length of the discussion, we concentrate on the distribution function of the asymptotically $N(0,1)$ statistic $T_n \equiv n^{1/2}(\theta_n - \theta)/s_n$ and the ERP of a symmetrical hypothesis test based on this statistic. Similar arguments apply to one-sided tests and to confidence intervals. Hall (1992) gives regularity conditions for the results of this section when the data that are a random sample from a distribution. The references cited in Sections 3-4 give regularity conditions for time series.

Let \hat{P} denote the probability measure induced by bootstrap sampling, and let \hat{T}_n denote a bootstrap analog of T_n . If the data are a random sample from a population, then it suffices to let \hat{P} be the empirical distribution of the data. Bootstrap samples are then drawn by sampling the data $\{X_i : i = 1, \dots, n\}$ randomly with replacement. If $\{\hat{X}_i : i = 1, \dots, n\}$ is such a sample, then

$$\hat{T}_n \equiv n^{1/2}(\hat{\theta}_n - \theta_n)/\hat{s}_n,$$

where $\hat{\theta}_n = \theta(\hat{m}_n)$, $\hat{m}_n = n^{-1} \sum_{i=1}^n \hat{X}_i$, and \hat{s}_n^2 is obtained by replacing the $\{X_i\}$ with $\{\hat{X}_i\}$ in the formula for s_n^2 . For example, if Σ_n denotes the sample covariance matrix of X and

$s_n^2 = \nabla \theta(m_n)' \Sigma_n \nabla \theta(m_n)$, then $\hat{s}_n^2 = \nabla \theta(\hat{m}_n)' \hat{\Sigma}_n \nabla \theta(\hat{m}_n)$, where $\hat{\Sigma}_n$ is the sample covariance matrix of the bootstrap sample. Bootstrap versions of \hat{T}_n for time-series data are presented in Sections 3-4. The discussion in this section does not depend on or require knowledge of the details of the bootstrap versions of \hat{T}_n .

The arguments showing that the bootstrap yields asymptotic refinements are based on Edgeworth expansions of $\mathbf{P}(T_n \leq z)$ and $\hat{\mathbf{P}}(\hat{T}_n \leq z)$. Additional notation is needed to describe the expansions. Let Φ and ϕ , respectively, denote the standard normal distribution function and density. The j 'th cumulant of T_n ($j \leq 4$) 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 and I is the indicator function (Hall 1992, p. 46). Define $\kappa = (\kappa_1, \dots, \kappa_4)'$. Conditional on the data $\{X_i : i = 1, \dots, n\}$, the j 'th cumulant of \hat{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 $\{X_i\}$. They are nonstochastic relative to bootstrap sampling but are random variables relative to the stochastic process that generates $\{X_i\}$. Define $\hat{\kappa} = (\hat{\kappa}_1, \dots, \hat{\kappa}_4)'$.

$\mathbf{P}(T_n \leq z)$ has the Edgeworth expansion

$$(2.1) \quad \mathbf{P}(T_n \leq z) = \Phi(z) + \sum_{j=1}^2 n^{-j/2} g_j(z, \kappa) \phi(z) + O(n^{-3/2})$$

uniformly over z , where $g_j(z, \kappa)$ is a polynomial function of z for each κ , a polynomial 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, $\mathbf{P}(|T_n| \leq z)$ has the expansion

$$(2.2) \quad \mathbf{P}(|T_n| \leq z) = 2\Phi(z) - 1 + 2n^{-1} g_2(z, \kappa) \phi(z) + O(n^{-2})$$

uniformly over z . Conditional on the data, the bootstrap probabilities $\hat{\mathbf{P}}(\hat{T}_n \leq z)$ and $\hat{\mathbf{P}}(|\hat{T}_n| \leq z)$ have the expansions

$$(2.3) \quad \hat{\mathbf{P}}(\hat{T}_n \leq z) = \Phi(z) + \sum_{j=1}^2 n^{-j/2} g_j(z, \hat{\kappa}) \phi(z) + O(n^{-3/2})$$

and

$$(2.4) \quad \hat{\mathbf{P}}(|\hat{T}_n| \leq z) = 2\Phi(z) - 1 + 2n^{-1} g_2(z, \hat{\kappa}) \phi(z) + O(n^{-2})$$

uniformly over z almost surely. Let $\nabla_{\kappa} g_j$ denote the gradient of g_j with respect to its second argument. Then a Taylor series expansion yields

$$(2.5) \quad |\hat{\mathbf{P}}(\hat{T}_n \leq z) - \mathbf{P}(T_n \leq z)| = n^{-1/2} [\nabla_{\kappa} g_1(z, \kappa)(\hat{\kappa} - \kappa)] \phi(z) + O\left(n^{-1/2} \|\hat{\kappa} - \kappa\|^2\right) + O(n^{-1})$$

and

$$(2.6) \quad |\hat{\mathbf{P}}(|\hat{T}_n| \leq z) - \mathbf{P}(|T_n| \leq z)| = 2n^{-1} [\nabla_{\kappa} g_2(z, \kappa)(\hat{\kappa} - \kappa)] \phi(z) + O\left(n^{-1} \|\hat{\kappa} - \kappa\|^2\right) + O(n^{-2})$$

almost surely uniformly over z . Thus, the leading terms of the errors made by bootstrap estimators of one-sided and symmetrical distribution functions are $n^{-1/2} [\nabla_{\kappa} g_1(z, \kappa)(\hat{\kappa} - \kappa)] \phi(z)$ and $2n^{-1} [\nabla_{\kappa} g_2(z, \kappa)(\hat{\kappa} - \kappa)] \phi(z)$, respectively. If the data are a random sample from a distribution, then $\|\hat{\kappa} - \kappa\| = O_p(n^{-1/2})$, so the errors of the bootstrap estimators are $O_p(n^{-1})$ and $O_p(n^{-3/2})$ for one-sided and symmetrical distribution functions, respectively. The root-mean-square estimation errors (RMSE's) also converge at these rates. As will be discussed in Sections 3 and 4, the rate of convergence of $\|\hat{\kappa} - \kappa\|$ is slower than $n^{-1/2}$ when the data are a time series. Thus, the errors of bootstrap estimators of distribution functions are larger when the data are a time series than when they are a random sample.

Now consider the ERP of a symmetrical hypothesis test. Let $z_{\alpha/2}$ and $\hat{z}_{\alpha/2}$, respectively, denote the $1-\alpha/2$ quantiles of the distributions of T_n and \hat{T}_n . Then $\mathbf{P}(|T_n| \leq z_{\alpha/2}) = \hat{\mathbf{P}}(|\hat{T}_n| \leq \hat{z}_{\alpha/2}) = 1-\alpha$. The bootstrap-based symmetrical test at the nominal α level accepts H_0 if $|T_n| \leq \hat{z}_{\alpha/2}$. Thus, the ERP of the test is $\mathbf{P}(|T_n| \leq \hat{z}_{\alpha/2}) - (1-\alpha)$.

To derive the ERP, use (2.1) and (2.3) to obtain

$$(2.7) \quad 2\Phi(z_{\alpha}) - 1 + 2n^{-1} g_2(z_{\alpha}, \kappa) \phi(z_{\alpha}) = 1 - \alpha + O(n^{-2})$$

and

$$(2.8) \quad 2\Phi(\hat{z}_{\alpha}) - 1 + 2n^{-1} g_2(\hat{z}_{\alpha}, \hat{\kappa}) \phi(z_{\alpha}) = 1 - \alpha + O(n^{-2})$$

almost surely. Let v_{α} denote the $1-\alpha/2$ quantile of the $N(0,1)$ distribution. Then Cornish-Fisher inversions of (2.7) and (2.8) give

$$z_{n\alpha} = v_{\alpha} - n^{-1} g_2(v_{\alpha}, \kappa) + O(n^{-2})$$

and

$$\hat{z}_{n\alpha} = v_{\alpha} - n^{-1} g_2(v_{\alpha}, \hat{\kappa}) + O(n^{-2})$$

almost surely (Hall 1992). Therefore,

$$(2.9) \quad \mathbf{P}(|T_n| \leq \hat{z}_{n\alpha}) = \mathbf{P}\{|T_n| \leq z_{n\alpha} + n^{-1}[\mathbf{g}_2(v_\alpha, \hat{\kappa}) - \mathbf{g}_2(v_\alpha, \kappa)] + O(n^{-2})\}.$$

Now suppose that $\|\hat{\kappa} - \kappa\| = O(n^{-a})$ almost surely for some $a > 0$. Since the components of κ can be estimated with errors that are no smaller than $O_p(n^{-1/2})$, it follows from (2.9) that

$$\mathbf{P}(|T_n| \leq \hat{z}_{\alpha/2}) = 1 - \alpha + O(n^{-1-a}).$$

Thus, in general, the rate of convergence of the ERP (and of the ECP for a symmetrical confidence interval) is determined by the rate of convergence of $\|\hat{\kappa} - \kappa\|$.

If the data are a random sample from a distribution, then it is possible to carry out an Edgeworth expansion of the right-hand side of (2.9). This yields

$$(2.10) \quad \mathbf{P}(|T_n| \leq \hat{z}_{\alpha/2}) = 1 - \alpha + O(n^{-2})$$

See Hall (1992, pp. 108-114). Thus, the ERP of a symmetrical test (and the ECP of a symmetrical confidence interval) based on the bootstrap critical value $\hat{z}_{\alpha/2}$ is $O(n^{-2})$ when the data are a random sample from a population. As will be discussed in Sections 3 and 4, this rate of the ERP is not available with current bootstrap methods for time series. Rather, the ERP for time-series data is $O(n^{-1-a})$ for some a satisfying $0 < a < 1/2$.

3. The Block Bootstrap

The block bootstrap is the best-known method for implementing the bootstrap with time-series data. It consists of dividing the data into blocks of observations and sampling the blocks randomly with replacement. The blocks may be non-overlapping (Hall 1985, Carlstein 1986) or overlapping (Hall 1985, Künsch 1989, Politis and Romano 1993). To describe these blocking methods more precisely, let the data consist of observations $\{X_i : i = 1, \dots, n\}$. With non-overlapping blocks of length ℓ , block 1 is observations $\{X_j : j = 1, \dots, \ell\}$, block 2 is observations $\{X_{\ell+j} : j = 1, \dots, \ell\}$, and so forth. With overlapping blocks of length ℓ , block 1 is observations $\{X_j : j = 1, \dots, \ell\}$, block 2 is observations $\{X_{j+1} : j = 1, \dots, \ell\}$, and so forth. The bootstrap sample is obtained by sampling blocks randomly with replacement and laying them end-to-end in the order sampled. It is also possible to use overlapping blocks with lengths that are sampled randomly from the geometric distribution (Politis and Romano 1993). The block bootstrap with random block lengths is also called the *stationary bootstrap* because the resulting bootstrap data series is stationary, whereas it is not with overlapping or non-overlapping blocks of non-stochastic lengths.

Regardless of the blocking method that is used, the block length (or average block length in the stationary bootstrap) must increase with increasing sample size n to make bootstrap estimators of moments and distribution functions consistent (Carlstein 1986, Künsch 1989, Hall *et al.* 1995). Similarly, the block length must increase with increasing sample size to enable the block bootstrap to achieve asymptotically correct coverage probabilities for confidence intervals and rejection probabilities for hypothesis tests. When the objective is to estimate a moment or distribution function, the asymptotically optimal block length may be defined as the one that minimizes the asymptotic mean-square error of the block bootstrap estimator. When the objective is to form a confidence interval or test a hypothesis, the asymptotically optimal block length may be defined as the one that minimizes the ECP of the confidence interval or ERP or the test. The asymptotically optimal block length and the corresponding rates of convergence of block bootstrap estimation errors, ECP's and ERP's depend on what is being estimated (e.g., bias, a one-sided distribution function, a symmetrical distribution function, etc.). The optimal block lengths and the rates of convergence of block bootstrap estimation errors with non-stochastic block lengths are discussed in detail in Section 3.2. The accuracy of the stationary bootstrap is discussed in Section 3.3. The performance of some modified forms of the block bootstrap are discussed in Sections 3.4-3.5. Before presenting results on the performance, it is necessary to deal with certain problems that arise in centering and Studentizing statistics based on the block bootstrap. These issues are discussed in Sections 3.1.

3.1 Centering and Studentizing with the Block Bootstrap

Two problems are treated in this section. The first is the construction of a block bootstrap version of the centered statistic $\Delta_n \equiv \theta(m_n) - \theta(\mu)$ that does not have excessive bias. This problem is discussed in Section 3.1.1. The second problem is Studentization of the resulting block bootstrap version of $n^{1/2}\Delta_n$. This is topic of Section 3.2.2. We consider only non-stochastic block lengths in this section. The stationary bootstrap is discussed in Section 3.3.

3.1.1 Centering

The problem of centering and its solution can be seen most simply by assuming that X_i is a scalar and θ is the identity function. Thus, $\Delta_n = m_n - \mu$ and $E\Delta_n = 0$. An obvious block bootstrap version of Δ_n is $\hat{\Delta}_n = \hat{m}_n - m_n$, where $\hat{m}_n = n^{-1} \sum_{i=1}^n \hat{X}_i$, and $\{\hat{X}_i\}$ is the block bootstrap sample using either non-overlapping or overlapping blocks. Let \hat{E} denote the

expectation operator with respect to the probability measure induced by block bootstrap sampling. If the blocks are non-overlapping, then $\hat{\mathbf{E}}\hat{m}_n = m_n$, so $\hat{\mathbf{E}}\hat{\Delta}_n = 0$. With overlapping blocks, however, it can be shown that

$$\hat{\mathbf{E}}\hat{m}_n = m_n + [\ell(n - \ell - 1)]^{-1}[\ell(\ell - 1)m_n - \tau_1 - \tau_2],$$

where ℓ is the block length $\tau_1 = \sum_{j=1}^{\ell-1}(\ell - j)X_j$, $\tau_2 = \sum_{j=n-\ell-2}^n[j - (n - \ell + 1)]X_j$, and it is assumed for simplicity that n is an integer multiple of ℓ (Hall *et al.* 1995). Thus, $\hat{\mathbf{E}}\hat{m}_n \neq m_n$ with overlapping blocks. The resulting bias decreases the rate of convergence of the estimation errors of the block bootstrap with overlapping blocks. This problem can be solved by centering the overlapping block bootstrap estimator at $\hat{\mathbf{E}}\hat{m}_n$ instead of at m_n . The resulting bootstrap version of Δ_n is $\hat{\Delta}_n = \hat{m}_n - \hat{\mathbf{E}}\hat{m}_n$. More generally, the block bootstrap version of $\theta(m_n) - \theta(\mu)$ is $\theta(\hat{m}_n) - \theta(\hat{\mathbf{E}}\hat{m}_n)$. This centering can also be used with non-overlapping blocks because $\hat{\mathbf{E}}\hat{m}_n = m_n$ with non-overlapping blocks.

3.1.2 Studentization

This section addresses the problem of Studentizing $n^{1/2}\hat{\Delta}_n = n^{1/2}[\theta(\hat{m}_n) - \theta(\hat{\mathbf{E}}\hat{m}_n)]$. The source of the problem is that blocking distorts the dependence structure of the DGP. As a result, the most obvious methods for Studentizing the bootstrap version $n^{1/2}\Delta_n$ create excessively large estimation errors. Various forms of this problem have been discussed by Lahiri (1992), Davison and Hall (1993), and Hall and Horowitz (1996). The discussion in this section is based on Hall and Horowitz (1996).

To illustrate the essential issues with a minimum of complexity, assume that the blocks are non-overlapping, θ is the identity function, and $\{X_i\}$ is a sequence of uncorrelated (though not necessarily independent) scalar random variables. Let V denote the variance operator relative to the process that generates $\{X_i\}$. Then $n^{1/2}\Delta_n = n^{1/2}(m_n - \mu)$, $n^{1/2}\hat{\Delta}_n = n^{1/2}(\hat{m}_n - m_n)$ and $V(n^{1/2}\Delta_n) = \mathbf{E}(X_1 - \mu)^2$. The natural choice for s_n^2 is the sample variance, $s_n^2 = n^{-1}\sum_{i=1}^n(X_i - m_n)^2$, in which case $s_n^2 - \text{Var}(n^{1/2}\Delta_n) = O_p(n^{-1/2})$.

Now consider Studentization of $n^{1/2}\hat{\Delta}_n$. Let ℓ and B , respectively, denote the block length and number of blocks, and assume that $B\ell = n$. Let \hat{V} denote the variance operator relative

to the block bootstrap DGP. An obvious bootstrap analog of s_n^2 is $\hat{s}_n^2 = n^{-1} \sum_{i=1}^n (\hat{X}_i - \hat{m}_n)^2$, which leads to the Studentized statistic $\tilde{T}_n \equiv n^{1/2} \hat{\Delta}_n / \hat{s}_n$. However, $\hat{V}(n^{1/2} \hat{\Delta}_n) = \tilde{s}_n^2$, where

$$\tilde{s}_n^2 = n^{-1} \sum_{b=0}^B \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (X_{b\ell+i} - m_n)(X_{b\ell+j} - m_n)$$

(Hall and Horowitz 1996). Moreover, $\hat{s}_n^2 - \tilde{s}_n^2 = O[(\ell/n)^{1/2}]$ almost surely. The consequences of this relatively large error in the estimator of the variance of $n^{1/2} \hat{\Delta}_n$ can be seen by carrying out Edgeworth expansions of the one-sided distribution functions $\mathbf{P}(n^{1/2} \Delta_n / s_n \leq z)$ and $\hat{\mathbf{P}}(n^{1/2} \hat{\Delta}_n / \hat{s}_n \leq z)$. These have the forms

$$\mathbf{P}(n^{1/2} \Delta_n / s_n \leq z) = \Phi(z) + n^{-1/2} g_1(z, \kappa) \phi(z) + O(n^{-1})$$

and

$$\hat{\mathbf{P}}(n^{1/2} \hat{\Delta}_n / \hat{s}_n \leq z) = \Phi(z) + n^{-1/2} g_1(z, \hat{\kappa}) \phi(z) + z \phi(z) (\tau_n - 1) + O(\ell^{1/2} / n + 1/n)$$

almost surely, where $\tau_n = s_n / \tilde{s}_n$ and g_1 , κ , and $\hat{\kappa}$ are as defined in Section 2.3 (Hall and Horowitz 1996). Therefore, $\hat{\mathbf{P}}(n^{1/2} \hat{\Delta}_n / \hat{s}_n \leq z) - \mathbf{P}(n^{1/2} \Delta_n / s_n \leq z) = O[(\ell/n)^{1/2}]$. By contrast, the error made by first-order asymptotic approximations is $O(n^{-1/2})$. Thus, the block bootstrap does not provide asymptotic refinements and, in fact, is less accurate than first-order approximations when $n^{1/2} \hat{\Delta}_n$ is Studentized with \hat{s}_n .

This problem can be mitigated by Studentizing $n^{1/2} \hat{\Delta}_n$ with \tilde{s}_n (Lahiri 1996a) or the estimator $n^{-1} \sum_{b=0}^{B-1} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (X_{b\ell+i} - \hat{m}_n)(X_{b\ell+j} - \hat{m}_n)$ (Götze and Künsch 1996). The error in the block bootstrap estimator of a one-sided distribution function is then $o(n^{-1/2})$ almost surely (Lahiri 1996a, Götze and Künsch 1996). However, the distributions of the symmetrical probabilities $\mathbf{P}(|n^{1/2} \Delta_n| / s_n \leq z)$ and $\hat{\mathbf{P}}(|n^{1/2} \hat{\Delta}_n| / \tilde{s}_n \leq z)$ differ by $O(n^{-1})$, so the block bootstrap does not provide asymptotic refinements for symmetrical distributions, confidence intervals and tests.

Refinements for both one-sided and symmetrical distribution functions, confidence intervals, and tests can be obtained by replacing \tilde{T}_n with the “corrected” statistic $\hat{T}_n = \tau_n \tilde{T}_n$ (Hall and Horowitz 1996, Andrews 2002). In the remainder of this paper, \hat{T}_n will be called a “corrected”

bootstrap test statistic and τ_n will be called a correction factor. The estimation errors resulting from the use of corrected statistics are discussed in Section 3.3.

The use of a correction factor can be generalized to the case in which θ is not the identity function and X_i is a vector. Suppose that $\text{cov}(X_i, X_j) = 0$ whenever $|i - j| > M$ for some $M < \infty$. (This assumption is weaker than M -dependence, which requires X_i and X_j to be independent when $|i - j| > M$.) Then $s_n^2 = \nabla \theta(m_n)' \Sigma_n \nabla \theta(m_n)$, where

$$(3.1) \quad \Sigma_n = n^{-1} \sum_{i=1}^n \left[(X_i - m_n)(X_i - m_n)' + \sum_{j=1}^M H(X_i, X_{i+j}, m_n) \right],$$

and $H(X_i, X_{i+j}, m_n) = (X_i - m_n)(X_{i+j} - m_n)' + (X_i - m_n)'(X_{i+j} - m_n)$. The bootstrap version of s_n^2 is $\hat{s}_n^2 = \nabla \theta(\hat{m}_n)' \hat{\Sigma}_n \nabla \theta(\hat{m}_n)$, where

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

Define

$$\tilde{\Sigma}_n = n^{-1} \sum_{b=0}^B \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (X_{b\ell+i} - m_n)(X_{b\ell+j} - m_n)',$$

$\tilde{s}_n^2 = \nabla \theta(m_n)' \tilde{\Sigma}_n \nabla \theta(m_n)$, and $\tau_n = s_n / \tilde{s}_n$. Then with non-overlapping blocks, the corrected block bootstrap version of $n^{1/2}(\theta_n - \theta) / s_n$ is $\tau_n n^{1/2}[\theta(\hat{m}_n) - \theta(\hat{m}_n)] / \hat{s}_n$ (Hall and Horowitz 1996). Andrews (2002) gives the overlapping-blocks version of the statistic.

The foregoing discussion assumes that $\text{cov}(X_i, X_j) = 0$ when $|i - j| > M$ for some $M < \infty$. When this assumption is not made, Σ_n must be replaced by a kernel-type estimator of the covariance matrix of $m_n - \mu$. See, for example, (e.g., Newey and West 1987, 1994; Andrews 1991, Andrews and Monahan 1992, Götze and Künsch 1996). In contrast to the covariance-matrix estimator (3.1), kernel covariance matrix estimators are not functions of sample moments. This complicates the analysis of rates of convergence of estimation errors. As was discussed in Section 2.3, this analysis is based on Edgeworth expansions of the distributions of the relevant statistics. The most general results on the validity of such expansions assumes that the statistic of interest is a function of sample moments (Götze and Hipp 1983, 1994; Lahiri 1996b). Consequently, as will be discussed in Section 3.2, the properties of the block bootstrap are less well understood when Studentization is with a kernel covariance matrix estimator than when Studentization is with a function of sample moments.

3.2 The Accuracy of Block Bootstrap Estimates

This section summarizes results on the magnitudes of the estimation errors made by the block bootstrap. The earliest result appears to be due to Carlstein (1986), who gave conditions under which the block bootstrap with non-overlapping blocks provides a consistent estimator of the variance of $\{X_1\}$ based on observations $\{X_i : i = 1, \dots, n\}$ from a strictly stationary time series. For the special case of an AR(1) DGP, Carlstein also calculated the block length that minimizes the asymptotic mean-square error (AMSE) of the variance estimator (the asymptotically optimal block length). This length increases at the rate $\ell \propto n^{1/3}$. The corresponding AMSE of the variance estimator is $O(n^{-2/3})$.

Künsch (1989) investigated the use of the block bootstrap to estimate a distribution function. He gave conditions under which the overlapping-blocks bootstrap consistently estimates the CDF of a sample average. Lahiri (1991, 1992) was the first to investigate the ability of the overlapping-blocks bootstrap to provide asymptotic refinements for estimation of the CDF of a normalized function of a sample mean. He also investigated refinements for the CDF of a Studentized function of a sample mean for the special case in which the DGP is M -dependent. Lahiri (1991, 1992) gave conditions under which the error in the bootstrap estimator of the one-sided distribution function of a normalized or Studentized function of a sample mean is $o(n^{-1/2})$ almost surely. In contrast, the errors made by first-order asymptotic approximations are $O(n^{-1/2})$, so the overlapping-blocks bootstrap provides an asymptotic refinement.

Lahiri's results were refined and extended by Hall, *et al.* (1995) and Zvingelis (2001), who give exact rates of convergence of block bootstrap estimators of certain moments, one-sided and symmetrical distribution functions, and the ERP's of tests. We first summarize the results of Hall, *et al.* (1995). The following notation will be used. Define $\Delta_n = \theta(m_n) - \theta(\mu)$, $\beta = \mathbf{E}\Delta_n$, and $\sigma_\infty^2 = V(n^{1/2}\Delta_n)$, where the moments are assumed to exist. Define the block bootstrap analogs $\hat{\Delta}_n = \theta(\hat{m}_n) - \theta(\hat{\mathbf{E}}\hat{m}_n)$, $\hat{\beta} = \hat{\mathbf{E}}\hat{\Delta}_n$, and $\hat{\sigma}_\infty^2 = \hat{V}(n^{1/2}\hat{\Delta}_n)$. The block bootstrap may use either overlapping or non-overlapping blocks. Define one-sided and symmetrical distribution functions of the normalized statistic $n^{1/2}\Delta_n/\sigma_\infty$ by $F_1(z) = \mathbf{P}(n^{1/2}\Delta_n/\sigma_\infty \leq z)$ and $F_2(z) = \mathbf{P}(n^{1/2}|\Delta_n|/\sigma_\infty \leq z)$. Define bootstrap analogs of F_1 and F_2 by $\hat{F}_1(z) = \hat{\mathbf{P}}(n^{1/2}\hat{\Delta}_n/\hat{\sigma}_\infty \leq z)$ and $\hat{F}_2(z) = \hat{\mathbf{P}}(n^{1/2}|\hat{\Delta}_n|/\hat{\sigma}_\infty \leq z)$. Finally, let $(\psi, \hat{\psi})$ denote either $(\beta, \hat{\beta})$ or $(\sigma_\infty^2/n, \hat{\sigma}_\infty^2/n)$, and let ϕ denote the standard normal density function.

Hall, *et al.* (1995) show that there are constants C_j ($j = 1, \dots, 6$) such that

$$(3.3) \quad E(\psi - \hat{\psi})^2 \sim n^{-2}(C_1\ell^{-2} + C_2n^{-1}\ell),$$

$$(3.4) \quad E[F_1(z) - \hat{F}_1(z)]^2 \sim n^{-1}(C_3\ell^{-2} + C_4n^{-1}\ell^2)\phi(z)^2,$$

and

$$(3.5) \quad E[F_2(z) - \hat{F}_2(z)]^2 \sim n^{-2}(C_5\ell^{-2} + C_6n^{-1}\ell^3)\phi(z)^2,$$

where the symbol “ \sim ” indicates that the quantity on the right-hand side is the leading term of an asymptotic expansion. The constants C_j do not depend on n or ℓ . The terms involving C_1 , C_3 , and C_5 correspond to the bias of the bootstrap estimator, and the terms involving C_2 , C_4 , and C_6 correspond to the variance. As will be discussed in Section 3.2.1, the variance terms are smaller if the blocks are overlapping than if they are non-overlapping. It follows from (3.3)-(3.5) that the asymptotically optimal block length (in the sense of minimizing the AMSE) is $\ell \propto n^{1/3}$ for bias or variance estimation, $\ell \propto n^{1/4}$ for estimating a one-sided distribution function, and $\ell \propto n^{1/5}$ for estimating a symmetrical distribution function. The corresponding minimum asymptotic RMSE's are proportional to $n^{-4/3}$ for bias or variance estimation, $n^{-3/4}$ for estimating a one-sided distribution function of a normalized statistic, and $n^{-6/5}$ for estimating a symmetrical distribution function of a normalized statistic. These results hold for overlapping and non-overlapping blocks. By contrast, the errors made by first-order asymptotic approximations are $O(n^{-1/2})$ for estimating a bias, variance or one-sided distribution function; and $O(n^{-1})$ for estimating a symmetrical distribution function. With data that are a random sample from a distribution, the bootstrap's asymptotic RMSE's are $O(n^{-3/2})$ for bias or variance estimation, $O(n^{-1})$ for estimating a one-sided distribution function and $O(n^{-3/2})$ for estimating a symmetrical distribution function. Thus, estimation errors of the block bootstrap with the asymptotically optimal block length converge to zero more rapidly than do the errors made by first-order approximations but less rapidly than do those of the bootstrap for data that are a random sample. The RMSE of the block bootstrap estimator of a symmetrical distribution function converges only slightly more rapidly than the RMSE of first-order asymptotic theory.

Zvingelis (2001) has extended these results to the ECP's of confidence intervals and the ERP's of hypothesis tests based on Studentized statistics. Zvingelis uses non-overlapping blocks, and he assumes that $\text{cov}(X_i, X_j) = 0$ whenever $|i - j|$ exceeds some finite value. The

Studentized statistic that he considers is $T_n = n^{1/2}[\theta(m_n) - \theta(\mu)]/s_n$, where $s_n^2 = \nabla\theta(m_n)' \Sigma_n \nabla\theta(m_n)$ and Σ_n is defined in (3.1). The bootstrap analog is $\hat{T}_n = \tau_n n^{1/2}[\theta(\hat{m}_n) - \theta(m_n)]/\hat{s}_n$, where $\hat{s}_n^2 = \nabla\theta(\hat{m}_n)' \hat{\Sigma}_n \nabla\theta(\hat{m}_n)$, $\hat{\Sigma}_n$ is defined in (3.2), and τ_n is the correction factor defined in Section 3.1.2. Zvingelis defines an asymptotically optimal block length as one that maximizes the rate of convergence of the ECP of a confidence interval or the ERP of a test. To state his results, let \hat{z}_α be the $1 - \alpha$ quantile of the distribution of \hat{T}_n relative to the probability measure induced by block bootstrap sampling. Then

1. The asymptotically optimal bandwidth for a one-sided confidence interval or test is $\ell \propto n^{1/4}$. The corresponding ECP and ERP are $O(n^{-3/4})$. Thus, for example, $\mathbf{P}(T_n \leq \hat{z}_\alpha) - (1 - \alpha) = O(n^{-3/4})$.

2. The asymptotically optimal bandwidth for a symmetrical confidence interval or test is $\ell \propto n^{1/4}$. The corresponding ECP and ERP are $O(n^{-5/4})$. Thus, for example, $\mathbf{P}(|T_n| \leq \hat{z}_{\alpha/2}) - (1 - \alpha) = O(n^{-5/4})$.

The errors made by first-order asymptotic approximations are $O(n^{-1/2})$ for one-sided confidence intervals and tests, and $O(n^{-1})$ for symmetrical confidence intervals and tests. Thus, the ECP and ERP of block bootstrap confidence intervals and tests converge more rapidly than do the ECP and ERP based on first-order approximations. However, the rates of convergence of the block bootstrap ECP and ERP are slower than the rates obtained with the bootstrap for data that are a random sample from a population. These rates are $O(n^{-1})$ (sometimes $O(n^{-3/2})$) for one-sided confidence intervals and tests, and $O(n^{-2})$ for symmetrical confidence intervals and tests.

Götze and Künsch (1996), Lahiri (1996a), and Inoue and Shintani (2001) have investigated the application of the block bootstrap to statistics that are Studentized with a kernel covariance matrix estimator and overlapping blocks. In Götze and Künsch (1996) and Lahiri (1996a), the Studentized statistic is $T_n = n^{1/2} \Delta_n / s_n$, where $s_n^2 = \nabla\theta(m_n)' \bar{\Sigma}_n \nabla\theta(m_n)$,

$$\bar{\Sigma}_n = n^{-1} \sum_{i=0}^{\ell} \sum_{j=1}^{n-\ell} [2 - I(i=0)] \omega(i/\ell) (X_i - m_n)(X_{i+j} - m_n)',$$

I is the indicator function, $\omega: [0, 1) \rightarrow [0, 1]$ is a kernel function satisfying $\omega(0) = 1$ and $\omega(u) = 0$ if $|u| \geq 1$, and ℓ is both the block length and the width of the lag window. The bootstrap version of T_n is $n^{1/2}[\theta(\hat{m}_n) - \theta(\hat{\mathbf{E}}\hat{m}_n)]/\hat{s}_n$, where $\hat{s}_n^2 = \nabla\theta(\hat{m}_n)' \hat{\Sigma}_n \nabla\theta(\hat{m}_n)$,

$$\hat{\Sigma}_n = \sum_{[i/\ell]=[j/\ell]} m_{ij}^{-1} (\hat{X}_i - \hat{m}_n)(\hat{X}_j - \hat{m}_n)',$$

$[u]$ denotes the integer part of u , $m_{ij} = n$ in Götze and Künsch (1996), and $m_{ij} = n(1 - |i - j|/\ell)$ in Lahiri (1996). In Lahiri (1996), T_n is a Studentized slope coefficient of a linear mean-regression model. Lahiri gives conditions under which

$$D_n \equiv \sup_z |\hat{\mathbf{P}}(\hat{T}_n \leq z) - \mathbf{P}(T_n \leq z)| = o_p(n^{-1/2}).$$

In contrast, the error made by the asymptotic normal approximation is $O(n^{-1/2})$, so the block bootstrap provides an asymptotic refinement. Götze and Künsch (1996) refine this result. They show that the rate of convergence of D_n is maximized by using a rectangular kernel function ω and setting $\ell \propto n^{1/4}$. This yields

$$\sup_z |\hat{\mathbf{P}}(\hat{T}_n \leq z) - \mathbf{P}(T_n \leq z)| = O_p(n^{-3/4+\varepsilon})$$

for any $\varepsilon > 0$. The rectangular kernel has the disadvantage of not guaranteeing that $\hat{\Sigma}_n$ is positive definite. This problem can be overcome at the price of a slightly slower rate of convergence of D_n by setting $\omega(u) = (1 - u^2)I(|u| \leq 1)$. With this quadratic kernel,

$$\sup_z |\hat{\mathbf{P}}(\hat{T}_n \leq z) - \mathbf{P}(T_n \leq z)| = O_p(n^{-2/3+\varepsilon})$$

for any $\varepsilon > 0$. Götze and Künsch (1996) show that an asymptotic refinement cannot be achieved with the triangular kernel $\omega(u) = (1 - |u|)I(|u| \leq 1)$. Finally, Götze and Künsch (1996) show that

$$\mathbf{P}(T_n \leq \hat{z}_\alpha) = 1 - \alpha + o(n^{-1/2}),$$

where \hat{z}_α is the $1 - \alpha$ quantile of the distribution of \hat{T}_n under block bootstrap sampling. Thus, the nonoverlapping block bootstrap provides an asymptotic refinement for the ECP of a one-sided confidence interval and the ERP of a one-sided hypothesis test.

Inoue and Shintani (2001) have extended the results of Götze and Künsch (1996) and Lahiri (1996). Inoue and Shintani (2001) apply the overlapping-blocks bootstrap to a Studentized estimator of a slope coefficient in a (possibly overidentified) linear model that is estimated by instrumental variables. They assume that the block length, ℓ , and width of the lag window are equal, and they require the rate of increase of ℓ to be faster than $n^{1/6}$ but slower than $n^{1/4}$. To summarize their results, let ω be the kernel used in the covariance matrix estimator, and define q to be the largest integer such that

$$\lim_{u \rightarrow 0} [1 - \omega(u)] / |u|^q < \infty.$$

Let T_n and \hat{T}_n , respectively, denote the Studentized slope estimator and its block-bootstrap analog. Let \hat{z}_α denote the $1 - \alpha$ quantile of the distribution of \hat{T}_n under bootstrap sampling. Inoue and Shintani (2001) give conditions under which

$$\mathbf{P}(T_n \leq \hat{z}_\alpha) = 1 - \alpha + O(\ell/n) + O(\ell^{-q})$$

and

$$\mathbf{P}(|T_n| \leq \hat{z}_{\alpha/2}) = 1 - \alpha + o(\ell/n) + O(\ell^{-q}).$$

Because of Inoue and Shintani (2001) require the rate of increase of ℓ to exceed $n^{1/6}$, their tightest bound on the ERP of a symmetrical test is $o(a_n n^{-5/6})$ where $a_n \rightarrow \infty$ at an arbitrarily slow rate as $n \rightarrow \infty$. It is not known whether the bootstrap can achieve an ERP of $o(n^{-1})$ when a kernel covariance matrix estimator is used for Studentization.

3.2.1 Relative Accuracy of the Bootstrap with Overlapping and Nonoverlapping Blocks

Hall, *et al.* (1995) and Lahiri (1999) have compared the estimation errors made by the overlapping- and non-overlapping-blocks bootstraps. They find that when the asymptotically optimal block length is used for estimating a bias or variance, then the AMSE with non-overlapping blocks is $1.5^{2/3}$ times the AMSE with overlapping blocks. Thus the AMSE is approximately 31% larger with non-overlapping blocks. The rates of convergence of the AMSE's are equal, however. Hall, *et al.* (1995) also compare the AMSE's for estimation of a one-sided distribution function of a normalized sample average (that is, for estimating $\mathbf{P}[n^{1/2}(m_n - \mu)/\sigma_\infty \leq z]$). The AMSE is $1.5^{1/2}$ times or 22% larger with non-overlapping blocks than with overlapping ones. The bootstrap is less accurate with non-overlapping blocks because the variance of the bootstrap estimator is larger with non-overlapping blocks than with overlapping ones. The bias of the bootstrap estimator is the same for non-overlapping and overlapping blocks.

3.3 The Stationary Bootstrap

In the stationary bootstrap, the index i of the first observation in a block, X_i , is sampled from the discrete uniform distribution on $\{1, \dots, n\}$. The block length ℓ is sampled from the geometric distribution. That is, $\mathbf{P}(\ell = m) = (1 - p)^{m-1} p$ for $m = 1, 2, \dots$ and some $p \in (0, 1)$. If

$i + \ell > n$, then the block is $\{X_i, \dots, X_n, X_1, \dots, X_{\ell-n+i-1}\}$. The stationary bootstrap was proposed by Politis and Romano (1993), who showed that it generates a stationary bootstrap data series and gave conditions under which it consistently estimates the CDF of a smooth function of a sample mean. Lahiri (1999) calculated the AMSE of the stationary bootstrap estimator of a bias or variance when p is optimally chosen. He found that the AMSE of the stationary bootstrap always exceeds the AMSE of the bootstrap with non-stochastic block lengths, regardless of whether the blocks are overlapping or non-overlapping. Indeed, the asymptotic relative efficiency of the stationary bootstrap compared to the block bootstrap with non-random block lengths can be arbitrarily close to zero. More precisely, let $AMSE_{SB}$ and $AMSE_{NR}$, respectively, denote the asymptotic AMSE's of the stationary bootstrap and the block bootstrap with overlapping or non-overlapping blocks with non-random lengths. Then $AMSE_{NR}/AMSE_{SB} < 1$ always and can be arbitrarily close to zero. Lahiri (1999) reports that similar results apply to estimation of distribution functions. Thus, at least in terms of AMSE, the stationary bootstrap is unattractive relative to the block bootstrap with fixed-length blocks.

3.4 Subsampling

The block bootstrap's distortions of the dependence structure of a time series can be avoided by using a subsampling method proposed by Politis and Romano (1994) and Politis, *et al.* (1999). To describe this method, let $t_n \equiv t_n(X_1, \dots, X_n)$ be an estimator of the population parameter θ , and set $\Delta_n = \rho(n)(t_n - \theta)$, where the normalizing factor $\rho(n)$ is chosen so that $\mathbf{P}(\Delta_n \leq z)$ converges to a nondegenerate limit at continuity points of the latter. For $i = 1, \dots, n - \ell + 1$, let $\{X_j : j = i, \dots, i + \ell - 1\}$ be a subset of $\ell < n$ consecutive observations taken from the sample $\{X_i : i = 1, \dots, n\}$. Define $N_{n\ell}$ to be the total number of subsets that can be formed. Let $t_{\ell,k}$ denote the estimator t_ℓ evaluated at the k 'th subset. The subsampling method estimates $\mathbf{P}(\Delta_n \leq z)$ by

$$G_{n\ell}(z) \equiv \frac{1}{N_{n\ell}} \sum_{k=1}^{N_{n\ell}} I[\rho(\ell)(t_{\ell,k} - t_n) \leq z].$$

The intuition behind this method is that each subsample is a realization of length ℓ of the true DGP. Therefore, $\mathbf{P}(\Delta_\ell \leq z)$ is the exact sampling distribution of $\rho(\ell)(t_\ell - \theta)$, and

$$(3.6) \quad \mathbf{P}(\Delta_\ell \leq z) = \mathbf{E}\{I[\rho(\ell)(t_\ell - \theta) \leq z]\}.$$

$G_{n\ell}(z)$ is the estimator of the right-hand side of (3.6) that is obtained by replacing the population expectation by the average over subsamples and θ by t_n . If n is large but ℓ/n is small, then random fluctuations in t_n are small relative to those in t_ℓ . Accordingly, the sampling distributions of $\rho(\ell)(t_\ell - t_n)$ and $\rho(n)(t_n - \theta)$ are close. Similarly, if $N_{n\ell}$ is large, the average over subsamples is a good approximation to the population average. These ideas were formalized by Politis and Romano (1994), who give conditions under which the subsampling method consistently estimates $\mathbf{P}(\Delta_n \leq z)$ and the coverage probability of a confidence interval for θ .

Hall and Jing (1996) investigated the accuracy of the subsampling method for estimating one-sided and symmetrical distribution functions of a Studentized, asymptotically normal statistic. Hall and Jing (1996) find that when ℓ is chosen optimally, the rates of convergence of the RMSE's are $n^{-1/4}$ and $n^{-1/3}$, respectively, for one-sided and symmetrical distribution functions. These rates are slower than those of first-order asymptotic approximations and the block bootstrap. Hall and Jing (1996) also describe an extrapolation technique that accelerates the rate of convergence of the RMSE. This method will now be summarized for the case of estimating a one-sided distribution function. Let t_n and t_ℓ be the statistics based on samples of sizes n and ℓ , respectively. Assume that $\mathbf{P}(t_n \leq z) \rightarrow N(0,1)$ as $n \rightarrow \infty$ with a similar result for t_ℓ . Under regularity conditions, $\mathbf{P}(t_n \leq z)$ and $\mathbf{P}(t_\ell \leq z)$ have Edgeworth expansions

$$(3.7) \quad \mathbf{P}(t_n \leq z) = \Phi(z) + n^{-1/2}g(z)\phi(z) + O(n^{-1})$$

and

$$(3.8) \quad \mathbf{P}(t_\ell \leq z) = \Phi(z) + \ell^{-1/2}g(z)\phi(z) + O(\ell^{-1}),$$

where g is a polynomial function of z that does not depend on n or ℓ . Solving (3.8) for g and substituting the result into (3.7) yields

$$\mathbf{P}(t_n \leq z) = [1 - (\ell/n)^{1/2}]\Phi(z) + (\ell/n)^{1/2}\mathbf{P}(t_\ell \leq z) + O[(n\ell)^{-1/2}].$$

Let $\hat{\mathbf{P}}(t_\ell \leq z)$ denote the estimator of $\mathbf{P}(t_\ell \leq z)$ that is obtained by subsampling with blocks of length ℓ . Hall and Jing (1996) proposed estimating $\mathbf{P}(t_n \leq z)$ by

$$\hat{\mathbf{P}}(t_n \leq z) = [1 - (\ell/n)^{1/2}]\Phi(z) + (\ell/n)^{1/2}\hat{\mathbf{P}}(t_\ell \leq z).$$

They showed that the fastest possible rate of convergence in probability of $\hat{\mathbf{P}}(t_n \leq z) - \mathbf{P}(t_n \leq z)$ is $O_p(n^{-2/3})$. This occurs when $\ell \propto n^{1/3}$. They also used extrapolation to estimate the symmetrical probability $\mathbf{P}(|t_n| \leq z)$ and found that the fastest possible rate of convergence in

probability of $\hat{P}(|t_n| \leq z) - P(|t_n| \leq z)$ is $O_p(n^{-8/7})$. These rates are faster than those of first-order asymptotic approximations. However, the rate for estimating a one-sided probability is slower than that provided by the block bootstrap. The rate for estimating a symmetrical probability is also slower than that provided by the block bootstrap, at least when $\{X_i\}$ is uncorrelated after finitely many lags. Thus, in terms of rates of convergence of estimation errors, subsampling with or without extrapolation does not improve on the block bootstrap.

3.5 Modifications of the Block Bootstrap

This section describes attempts to improve the performance of the block bootstrap by reducing the influence of the “discontinuities” in the bootstrap data series that occur at block boundaries. Carlstein, *et al.* (1998) proposed sampling blocks according to a data-based Markov chain so as to increase the likelihood that consecutive blocks match at their ends. They gave conditions under which this *matched-block bootstrap* (MBB) reduces the bias of a bootstrap estimator of a variance. However, the MBB increases the rate of convergence of the bias only if the DGP is a Markov process. The MBB does not reduce the variance of the estimator. Carlstein, *et al.* (1998) did not investigate the performance of the MBB for estimating a distribution function, the coverage probability of a confidence interval, or the rejection probability of a hypothesis test.

Paparoditis and Politis (2001) proposed downweighting points \hat{X}_i in the block bootstrap data series that are near block endpoints. They give conditions under which this *tapered bootstrap* procedure increases the rate of convergence of the MSE of the bootstrap estimator of $Var(n^{1/2}m_n)$ to $O(n^{-4/5})$. This may be contrasted with the rate $O(n^{-2/3})$ that is provided by the block bootstrap without tapering. Paparoditis and Politis (2001) did not investigate the ability of the tapered block bootstrap to provide asymptotic refinements for distribution estimation, ECP’s of confidence intervals, or ERP’s of hypothesis tests. Thus, it remains unknown whether the MBB or tapered bootstrap can increase the rates of convergence of the errors made by the block bootstrap for estimating distribution functions, coverage probabilities, and rejection probabilities.

4. Methods That Impose Stronger Restrictions on the DGP

This section describes bootstrap methods that make stronger *a priori* assumptions about the structure of the DGP than does the block bootstrap. In return for stronger assumptions, some of these methods achieve higher orders of asymptotic refinement (that is, faster rates of convergence of estimation errors) than does the block bootstrap.

4.1 The Sieve Bootstrap for Linear Processes

A substantial improvement over the performance of the block bootstrap is possible if the DGP is known to be a linear process. That is, the DGP has the form

$$(4.1) \quad X_i - \mu = \sum_{j=1}^{\infty} \alpha_j (X_{i-j} - \mu) + U_i,$$

where $\mu = \mathbf{E}(X_i)$ for all i , $\{U_i\}$ is a sequence of independently and identically distributed (*iid*) random variables, and $\{X_i\}$ may be a scalar or a vector process. Assume that $\sum_{j=1}^{\infty} \alpha_j^2 < \infty$ and that all of the roots of the power series $1 - \sum_{j=1}^{\infty} \alpha_j z^j$ are outside of the unit circle. Bühlmann (1997, 1998), Kreiss (1988, 1992), and Paparoditis (1996) proposed approximating (4.1) by an AR(p) model in which $p = p(n)$ increases with increasing sample size. Let $\{a_{nj} : j=1, \dots, p\}$ denote least squares or Yule-Walker estimates of the coefficients of the approximating process, and let $\{U_{nj}\}$ denote the centered residuals. The sieve bootstrap consists of generating bootstrap samples according to the process

$$(4.2) \quad \hat{X}_i - m = \sum_{j=1}^p a_{nj} (\hat{X}_{i-j} - m) + \hat{U}_j,$$

where $m = n^{-1} \sum_{i=1}^n X_i$ and the \hat{U}_j are sampled randomly with replacement from the U_{nj} . Bühlmann (1997), Kreiss (1992, 2000), and Paparoditis (1996) have given conditions under which this procedure consistently estimates the distributions of sample averages, sample autocovariances and autocorrelations, and the regression coefficients a_{nj} among other statistics.

Choi and Hall (2000) investigated the ability of the sieve bootstrap to provide asymptotic refinements to the coverage probability of a one-sided confidence interval for the mean of a linear statistic when the X_i 's are scalar random variables. A linear statistic has the form

$$\theta_n = (n - \ell + 1)^{-1} \sum_{i=1}^{n-\ell+1} G(X_i, \dots, X_{i+\ell-1}),$$

where $\ell \geq 1$ is a fixed integer and G is a known function. Define $\theta = \mathbf{E}[G(X_1, \dots, X_\ell)]$. Choi and Hall (2000) considered the problem of finding a one-sided confidence interval for θ , although they note that their conclusions also apply to a one-sided confidence interval for $H(\theta)$, where H is a continuously differentiable function with $H'(\theta) \neq 0$.

To obtain a confidence interval for θ , define $\sigma_\theta^2 = \text{Var}[(n - \ell + 1)^{1/2}(\theta_n - \theta)]$. Then

$$\sigma_\theta^2 = \gamma_G(0) + 2 \sum_{j=1}^{n-\ell} [1 - j(n-\ell+1)^{-1}] \gamma_G(j),$$

where $\gamma_G(j) = \text{cov}[G(X_1, \dots, X_\ell), G(X_{j+1}, \dots, X_{j+\ell})]$. The bootstrap analogs of θ_n and θ are

$$\hat{\theta}_n = (n-\ell+1)^{-1} \sum_{i=1}^{n-\ell+1} G(\hat{X}_i, \dots, \hat{X}_{i+\ell-1})$$

and $\hat{\theta} = \hat{E}[G(\hat{X}_1, \dots, \hat{X}_{1+\ell})]$, respectively, where \hat{E} is the expectation operator relative to the probability distribution induced by sieve bootstrap sampling. The variance of $(n-\ell+1)^{1/2}(\hat{\theta}_n - \hat{\theta})$ conditional on the data is

$$\hat{\sigma}_\theta^2 = \hat{\gamma}_G(0) + 2 \sum_{j=1}^{n-\ell} [1 - j(n-\ell+1)^{-1}] \hat{\gamma}_G(j),$$

where $\hat{\gamma}_G(j) = \text{cov}[G(\hat{X}_1, \dots, \hat{X}_\ell), G(\hat{X}_{j+1}, \dots, \hat{X}_{j+\ell})]$ and the covariance is relative to the bootstrap probability distribution. Define $\hat{T}_n = (n-\ell+1)^{1/2}(\hat{\theta}_n - \hat{\theta})/\hat{\sigma}_\theta$, where $\hat{\sigma}_\theta$ is a bootstrap estimator of $\hat{\sigma}_\theta$, possibly obtained through a double bootstrap procedure. Let $\hat{z}_{n\alpha}$ satisfy

$$\hat{P}(\hat{T}_n \leq \hat{z}_{n\alpha}) = 1 - \alpha,$$

where \hat{P} denotes the probability measure induced by sieve bootstrap sampling. Then a lower $100(1-\alpha)$ percent confidence interval for θ is $\theta_n - (n-\ell+1)^{-1/2} \hat{z}_{n\alpha} \hat{\sigma}_\theta \leq \theta$, and an upper $100(1-\alpha)$ percent confidence interval is $\theta_n - (n-\ell+1)^{-1/2} \hat{z}_{n,1-\alpha} \hat{\sigma}_\theta \geq \theta$. Choi and Hall (2000) gave conditions under which the ECP's of these intervals are $O(n^{-1+\varepsilon})$ for any $\varepsilon > 0$. This is only slightly larger than the ECP of $O(n^{-1})$ that is available when the data are a random sample. Thus, in comparison to the block bootstrap, the sieve bootstrap provides a substantial improvement in the rate of convergence of the ECP when the DGP is a linear process.

4.2 The Bootstrap for Markov Processes

Asymptotic refinements of order higher than those provided by the block bootstrap are also possible if the DGP is a (possibly higher-order) Markov process or can be approximated by such a process. The class of Markov and approximate Markov processes contains many nonlinear autoregressive, ARCH, and GARCH processes, among others, that are important in applications.

When the DGP is a Markov process, the bootstrap can be implemented by estimating the Markov transition density nonparametrically. Bootstrap samples are generated by sampling the

stochastic process implied by the estimated transition density. Call this procedure the Markov bootstrap (MB). The MB was proposed by Rajarshi (1990), who gave conditions under which it consistently estimates the asymptotic distribution of a statistic. Datta and McCormick (1995) gave conditions under which the error in the MB 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 MB is consistent or provides asymptotic refinements. Chan and Tong (1998) proposed using the MB in a test for multimodality in the distribution of dependent data. Paparoditis and Politis (2001) proposed estimating the Markov transition probability by resampling the data in a suitable way.

Horowitz (2001b) investigated the ability of the MB to provide asymptotic refinements for confidence intervals and tests based on Studentized statistics. To describe the results, let the DGP be an order q Markov process that is stationary and geometrically strongly mixing. As before, let the data be $\{X_i : i = 1, \dots, n\}$. For any $j > q$, define $Y_j = (X_{j-1}, \dots, X_{j-q})$. Also define $\mu = \mathbf{E}(X_1)$ and $m = n^{-1} \sum_{i=1}^n X_i$. Horowitz (2001b) considered a confidence interval for $H(\mu)$, where H is a sufficiently smooth function. The confidence interval is based on the statistic

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

where s_n^2 is a consistent estimator of the variance of the asymptotic distribution of $n^{1/2}[H(m) - H(\mu)]$. Thus, for example, a symmetrical $1 - \alpha$ confidence interval is

$$(4.3) \quad H(m) - \hat{z}_{n\alpha} s_n \leq H(\mu) \leq H(m) + \hat{z}_{n\alpha} s_n,$$

where $\hat{z}_{n\alpha}$ is the MB estimator of the $1 - \alpha$ quantile of the distribution of $|T_n|$. Horowitz (2001b) used a kernel estimator of the Markov transition density to implement the MB. The estimator is $f_n(x|y) = p_{nz}(x, y)/p_{ny}(y)$, where

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

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

and K_f and K_p are kernel functions. The MB procedure samples the process induced by a version of $f_n(x|y)$ that is trimmed to keep the bootstrap process away from regions where the probability density of Y is close to zero. Define $C_n = \{y : p_{ny}(y) \geq \lambda_n\}$, where $\lambda_n \rightarrow 0$ at a

suitable rate as $n \rightarrow \infty$. Then the MB process used by Horowitz (2001b) consists of the following steps:

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

MB 2. Having obtained $\hat{Y}_j \equiv (\hat{X}_{j-1}, \dots, \hat{X}_{j-q})$ for any $j \geq q+2$, draw \hat{X}_j from the distribution whose density is $f_n(\cdot | \hat{Y}_j)$. Retain \hat{X}_j and set $\hat{Y}_{j+1} = (\hat{X}_j, \dots, \hat{X}_{j-q+1})$ if $(\hat{X}_j, \dots, \hat{X}_{j-q+1}) \in C_n$. Otherwise, discard the current \hat{X}_j and draw a new one. Continue this process until an \hat{X}_j is obtained for which $(\hat{X}_j, \dots, \hat{X}_{j-q+1}) \in C_n$.

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

MB 4. Set $\hat{z}_{n\alpha}$ equal to the $1 - \alpha$ quantile of the empirical distribution of $|\hat{T}_n|$ that is obtained by repeating steps 1-3 many times.

For any $\varepsilon > 0$, Horowitz (2001b) gives conditions under which the ECP of the resulting symmetrical confidence interval (4.3) is $O(n^{-3/2+\varepsilon})$. The ECP for a one-sided confidence interval is $O(n^{-1+\varepsilon})$. The corresponding ECP's for confidence intervals based on the block bootstrap with the asymptotically optimal block length are $O(n^{-4/3})$ and $O(n^{-3/4})$, respectively. Thus, the ECP's of the MB converge to zero more rapidly than do those of the block bootstrap.

These results also apply to approximate Markov processes. An approximate Markov process is a process whose density conditional on the past can be approximated by its density conditional on lags of order up to q with an error of $O(e^{-bq})$ for some $b > 0$. The MB for an approximate Markov process is implemented by generating bootstrap samples from the estimated order q Markov process, where $q \rightarrow \infty$ at a logarithmic rate as $n \rightarrow \infty$.

4.3 The Nonparametric Autoregressive Bootstrap

In this section it is assumed that the DGP is the autoregressive process

$$(4.4) \quad X_i = g(X_{i-1}, \dots, X_{i-p}) + \sigma(X_{i-1}, \dots, X_{i-q})U_i; \quad i = 0, 1, 2, \dots,$$

where g and σ are unknown functions, $\{U_i\}$ is a sequence of *iid* random variables with zero mean and unit variance, and $p, q \geq 1$ are integers. Equation (4.4) is a nonparametric version of an ARCH model. Allowing the possibility of a non-constant conditional variance function (conditional heteroskedasticity) is important in applications in finance (Engle 1982, Bollerslev *et al.* 1992, Gouriéroux 1997).

The DGP (4.4) is a Markov process to which the bootstrap procedure of Section 4.2 may be applied. In this section, however, we describe a procedure due to Franke, Kreiss, and Mammen (2000) (hereinafter FKM) that takes advantage of the specific structure of (4.4). This procedure generates bootstrap samples by replacing m , σ , and U_i in (4.4) with nonparametric estimates. It is assumed that g , σ , and the distribution of U are such that $\{X_i\}$ is strictly stationary and strongly geometrically mixing. See Diebolt and Guegan (1990), Meyn and Tweedie (1993), and FKM for regularity conditions that insure these properties.

Let the data consist of observations of $\{X_i : i = 1, \dots, n\}$. FKM use Nadaraya-Watson kernel estimators of g and σ . Other estimators such as local polynomials could also be used. In the case that $p = q = 1$, the estimators are

$$(4.5) \quad g_{nh}(x) = \frac{1}{(n-1)p_{nh}(x)} \sum_{i=1}^{n-1} X_{i+1} K\left(\frac{x - X_i}{h_n}\right)$$

and

$$(4.6) \quad \sigma_{nh}^2(x) = \frac{1}{(n-1)p_{nh}(x)} \sum_{i=1}^{n-1} [X_{i+1} - g_{nh}(X_i)]^2 K\left(\frac{x - X_i}{h_n}\right),$$

where K , the kernel, is a probability density function,

$$(4.7) \quad p_{nh}(x) = \frac{1}{(n-1)} \sum_{i=1}^{n-1} K\left(\frac{x - X_i}{h_n}\right),$$

and $\{h_n\}$ is a sequence of positive constants (bandwidths) that converges to zero as $n \rightarrow \infty$.

Product kernels may be used when $p > 1$ or $q > 1$.

Let $r = \max(p, q)$. As in the Markov bootstrap of Section 4.2, the procedure used to generate bootstrap samples based on (4.4) must avoid regions in which the density of $(X_{i-1}, \dots, X_{i-r})$ is close to zero. In addition, the estimates of g and σ that are used to generate

bootstrap samples must be based on a bandwidth \hat{h}_n that converges to zero more slowly than h_n so that $\hat{h}_n / h_n \rightarrow \infty$ as $n \rightarrow \infty$. To this end, let $g_{n\hat{h}}$ and $\sigma_{n\hat{h}}$, respectively, denote estimates of g and σ based on the bandwidth \hat{h}_n . Let $\tilde{g}_{n\hat{h}}$ and $\tilde{\sigma}_{n\hat{h}}$ denote $g_{n\hat{h}}$ and $\sigma_{n\hat{h}}$ restricted to a compact set C on which the density of $(X_{i-1}, \dots, X_{i-r})$ exceeds zero. Let $\{v_i\}$ denote the residuals $\{X_i - g_{n\hat{h}}(X_{i-1}, \dots, X_{i-q})\}$ corresponding to points $(X_{i-1}, \dots, X_{i-r})$ in C . Then bootstrap samples are generated by the recursion

$$\hat{X}_i = \tilde{g}_{n\hat{h}}(\hat{X}_{i-1}, \dots, \hat{X}_{i-p}) + \tilde{\sigma}_{n\hat{h}}(\hat{X}_{i-1}, \dots, \hat{X}_{i-q})\hat{U}_i,$$

where $\{\hat{U}_i\}$ is obtained by sampling the standardized v_i 's randomly with replacement. The process is initialized by setting $(\hat{X}_r, \dots, \hat{X}_1) = (X_r, \dots, X_1)$. Bootstrap estimators of g and σ are obtained by applying (4.5)-(4.7) with the bandwidth h_n to the bootstrap sample. Denote the resulting bootstrap estimators of g and σ by \hat{g}_{nh} and $\hat{\sigma}_{nh}$.

Let $\hat{\mathbf{P}}$ denote the probability measure induced by bootstrap sampling. Let y_p and y_q , respectively, denote the vectors (x_p, \dots, x_1) and (x_q, \dots, x_1) . FKM give conditions under which

$$\sup_z \left| \hat{\mathbf{P}}\{(nh_n^p)^{1/2}[\hat{g}_{nh}(y_p) - \tilde{g}_{n\hat{h}}(y_p)] \leq z\} - \mathbf{P}\{(nh_n^p)^{1/2}[g_{nh}(y_p) - g(y_p)] \leq z\} \right| \rightarrow^p 0$$

and

$$\sup_z \left| \hat{\mathbf{P}}\{(nh_n^q)^{1/2}[\hat{s}_{nh}(y_q) - \tilde{s}_{n\hat{h}}(y_q)] \leq z\} - \mathbf{P}\{(nh_n^q)^{1/2}[s_{nh}(y_q) - \sigma(y_q)] \leq z\} \right| \rightarrow^p 0.$$

Hafner (1996) and Härdle *et al.* (2001) use this nonparametric autoregressive bootstrap procedure to construct pointwise confidence intervals for g and σ . Franke, Kreiss, Mammen, and Neumann (2000) give conditions under which the nonparametric autoregressive bootstrap can be used to obtain uniform confidence bands for g and to carry out inference about the parameters of a misspecified (finite-dimensional) parametric model. Neumann and Kreiss (1998) use a wild-bootstrap version of the nonparametric autoregressive bootstrap to test parametric models of the conditional mean and variance functions. In addition, they give conditions under which the ECP of a uniform confidence band for g converges to zero at the rate $(\log n)^{3/2} / (nh_n)^{1/4}$. In contrast, first order asymptotic approximations yield the rate $(\log n)^{-1}$ (Hall 1991). Kreiss, Neumann, and Yao (1998) also use the nonparametric autoregressive bootstrap for testing.

4.4 The Periodogram Bootstrap

In this section, it is assumed that the data are generated by a stationary, univariate process with mean zero and the possibly infinite-order moving average representation

$$(4.8) \quad X_i = \sum_{k=-\infty}^{\infty} b_k \xi_{i-k} .$$

The innovations $\{\xi_i\}$ are *iid* with $E(\xi_i) = 0$, $E(\xi_i^2) = 1$, and $E(\xi_i^5) < \infty$. It is also assumed that the coefficients $\{b_k\}$ satisfy $\sum_{k=-\infty}^{\infty} |kb_k| < \infty$.

We begin by considering the problem of estimating the distribution of a kernel estimator the spectral density at frequency ω , $f(\omega)$. To construct the estimator, let

$$I_n(\omega) \equiv n^{-1} \left| \sum_{k=1}^n X_k e^{tk\omega} \right|^2, \quad -\pi \leq \omega \leq \pi$$

with $t = \sqrt{-1}$ denote the periodogram of the sample $\{X_i : i = 1, \dots, n\}$. Let N denote the largest integer that is less than or equal to $n/2$, and define the frequencies $\omega_k = 2\pi k/n$, $-N \leq k \leq N$. Then the kernel estimator of $f(\omega)$ is given by

$$f_n(\omega; h_n) = \frac{1}{nh_n} \sum_{k=-N}^N K\left(\frac{\omega - \omega_k}{h_n}\right) I_n(\omega_k),$$

where the kernel function K is symmetrical and non-negative, and $\{h_n\}$ is a sequence of bandwidths that converges to zero as $n \rightarrow \infty$. We seek an estimator of the probability distribution of $T_n(\omega, h_n) \equiv (nh_n)^{1/2} [f_n(\omega, h_n) - f(\omega)] / f(\omega)$.

Franke and Härdle (1992) (hereinafter FH) proposed using a frequency domain version of the bootstrap to estimate the distribution of $T_n(\omega, h_n)$. Nordgaard (1992) and Theiler *et al.* (1994) considered frequency domain bootstrap methods for Gaussian processes. Here, we do not assume that $\{X_i\}$ is Gaussian. The method of FH is based on the relation

$$I_n(\omega_k) = f(\omega_k) \varepsilon_k, \quad k = 1, \dots, N.$$

The residuals ε_k are approximately *iid* for large n . FH use a scaled version of $\{I_n(\omega_k) : k = 1, \dots, N\}$ to construct an estimate of $\{\varepsilon_k : k = 1, \dots, N\}$. They obtain bootstrap samples of the ε_k 's by sampling the estimates randomly with replacement. The bootstrap samples are used to construct bootstrap estimates of $f(\omega)$, and the distribution of $T_n(\omega, h_n)$ is estimated by

the distribution of a bootstrap analog that is obtained from the bootstrap estimates of $f(\omega)$. FH give conditions under which this procedure consistently estimates the distribution of $T_n(\omega, h_n)$.

To describe FH procedure more specifically, let $h_n \propto n^{-1/5}$. Also, define bandwidths $h_{n0} \propto n^{-1/4}$ and h_{n1} , where $h_{n1} \rightarrow 0$ and $h_n/h_{n1} \rightarrow 0$ as $n \rightarrow \infty$. Define

$$\hat{\varepsilon}_k = I_n(\omega_k) / f_n(\omega_k; h_{n0}), \quad k = 1, \dots, N,$$

$\bar{\varepsilon} = N^{-1} \sum_{j=1}^N \hat{\varepsilon}_j$, and $\tilde{\varepsilon}_k = \hat{\varepsilon}_k / \bar{\varepsilon}$. The FH procedure is:

FH1. Draw bootstrap residuals $\varepsilon_1^*, \dots, \varepsilon_N^*$ by sampling $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_N$ randomly with replacement.

FH2. Compute the bootstrap periodogram values $I_n^*(\omega_k) \equiv f_n(\omega_k, h_{n1}) \varepsilon_k^*$.

FH3. Compute the bootstrap spectral density estimate

$$f_n^*(\omega; h_n, h_{n1}) = \frac{1}{nh_n} \sum_{k=-N}^N K\left(\frac{\omega - \omega_k}{h_n}\right) I_n^*(\omega_k).$$

FH give conditions under which

$$d_M \left[(nh_n)^{1/2} \frac{f_n(\omega; h_n) - f(\omega)}{f(\omega)}, (nh_n)^{1/2} \frac{f_n^*(\omega; h_n, h_{n1}) - f_n(\omega; h_{n1})}{f_n(\omega; h_{n1})} \right] \rightarrow^P 0$$

as $n \rightarrow \infty$, where d_M is the Mallows metric. This result implies that the bootstrap distribution of

$$T_n^*(\omega; h_n, h_{n0}, h_{n1}) \equiv (nh_n)^{1/2} \frac{f_n^*(\omega; h_n, h_{n1}) - f_n(\omega; h_{n1})}{f_n(\omega; h_{n1})}$$

consistently estimates the distribution of $T_n(\omega; h_n)$ (Bickel and Freedman 1981). That is

$$\sup_z |\hat{P}[T_n^*(\omega; h_n, h_{n0}, h_{n1}) \leq z] - P[T_n(\omega; h_n) \leq z]| \rightarrow^P 0$$

as $n \rightarrow \infty$.

Dahlhaus and Janas (1996) (hereinafter DJ) applied a modified version of the FH procedure to a class of *ratio statistics* that includes sample autocorrelation coefficients and a normalized estimate of the spectral density function. DJ give conditions under which the difference between the CDF of a normalized ratio statistic and the CDF of a bootstrap analog is uniformly $o(n^{-1/2})$ almost surely. By contrast, the error made by the asymptotic normal approximation to the CDF of a normalized ratio statistic is $O(n^{-1/2})$, so the frequency domain bootstrap provides an asymptotic refinement under the conditions of DJ (1996). Some of these conditions are quite restrictive. In particular, X_i must have a known mean of zero, and the

innovations must satisfy $E(\xi_i^3) = 0$. DJ point out that the asymptotic refinement is not achieved if these conditions do not hold. In addition, DJ assume that the variance of the asymptotic distribution of the ratio statistic is known.

Kreiss and Paparoditis (2000) describe a procedure called the *autoregressive aided periodogram bootstrap* (AAPB). In this procedure, bootstrap samples $\{\hat{X}_i\}$ are generated using the sieve procedure of Section 4.1. These samples are used to calculate a bootstrap version of the periodogram and bootstrap versions of statistics that are functionals of the periodogram. Kreiss and Paparoditis (2000) give conditions under which the AAPB procedure consistently estimates the distribution of ratio statistics and statistics based on the integrated periodogram. These conditions are sufficiently general to permit application of the AAPB to moving average processes that are not invertible and, therefore, do not have an autoregressive representation.

5. Conclusions

The block bootstrap is the best known method for implementing the bootstrap with time-series data when one does not have a finite-dimensional parametric model that reduces the DGP to independent random sampling. The block bootstrap makes relatively weak assumptions about the structure of the DGP, but its ERP's and ECP's converge to zero only slightly faster than those of first-order asymptotic approximations. Faster rates of convergence can be achieved by imposing additional structure on the DGP. It is an open question whether it is possible to develop methods that are more accurate than the block bootstrap but impose less structure on the DGP than do the Markov bootstrap and the sieve bootstrap for linear processes. This is a question of considerable practical importance that merits further research. First-order asymptotic approximations are often inaccurate with the sample sizes encountered in applications, and the bootstrap is one of the few practical methods that has the potential to achieve significantly better accuracy.

The existing theoretical explanation of the bootstrap's ability to provide asymptotic refinements is based on Edgeworth expansions. This theory is an imperfect tool for understanding the performance of the bootstrap, and the imperfections limit the theory's usefulness for developing improved bootstrap methods for time-series. If the DGP is GSM, then the parameters κ that enter the Edgeworth expansions (2.1) and (2.2) can be estimated analytically with errors that are only slightly larger than $O(n^{-1/2})$. By substituting these estimates into an analytic Edgeworth expansion, one can obtain theoretical ECP's and ERP's that are comparable to or smaller than those of the block, sieve and Markov bootstraps. In Monte

Carlo experiments, however, analytic Edgeworth expansions are often much less accurate than the bootstrap. Thus, Edgeworth expansions are imperfect guides to the relative accuracy of alternative methods for achieving asymptotic refinements. This is a serious problem for research on bootstrap methods for time series because comparing the relative accuracies of alternative bootstrap approaches is an essential element of this research. Accordingly the development of a more complete theory of the performance of the bootstrap is a potentially important, though undoubtedly very difficult, area for future research.

REFERENCES

- Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817-858.
- Andrews, D.W.K. (2002). Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators, *Econometrica*, forthcoming.
- Andrews, D.W.K. and J.C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 59, 817-858.
- Beran, R. and G.R. Ducharme (1991), *Asymptotic Theory for Bootstrap Methods in Statistics*, Les Publications CRM, Centre de recherches mathématiques, Université de Montréal, Montréal, Canada.
- Bickel, P. and D. Freedman (1981). Some Asymptotic Theory for the Bootstrap, *Annals of Statistics*, 9, 1196-1217.
- Bollerslev, T. and R. Chou, and K. Kroner (1992). ARCH Modelling in Finance: A Review of Theory and Empirical Evidence, *Journal of Econometrics*, 52, 5-60.
- Bose, A. (1988). Edgeworth Correction by Bootstrap in Autoregressions, *Annals of Statistics*, 16, 1709-1722.
- Bose, A. (1990). Bootstrap in Moving Average Models, *Annals of the Institute of Statistical Mathematics*, 42, 753-768.
- Bühlmann, P. (1997). Sieve Bootstrap for Time Series, *Bernoulli*, 3, 123-148.
- Bühlmann, P. (1998). Sieve Bootstrap for Smoothing 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.
- Carlstein, E. K.-A. Do, P. Hall, T. Hesterberg, and H.R. Künsch (1998). Matched-block bootstrap for dependent data, *Bernoulli*, 4, 305-328.
- 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.
- 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.
- Dahlhaus, R. and D. Janas (1996). A Frequency Domain Bootstrap for Ratio Statistics in Time Series, *Annals of Statistics*, 24, 1934-1963.
- 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.
- Davison, A.C. and P. Hall (1993). On Studentizing and Blocking Methods for Implementing the Bootstrap with Dependent Data, *Australian Journal of Statistics*, 35, 215-224.
- Davison, A.C. and D.V. Hinkley (1997). *Bootstrap Methods and Their Application* (Cambridge University Press, Cambridge, U.K).
- Diebolt, J. and Guegan, D. (1990). Probabilistic properties of the general nonlinear Markovian process of order one and applications to time series modelling, Rapport Technique de L.S.T.A. 125, Université Paris XIII.
- Efron, B. and R.J. Tibshirani (1993). *An Introduction to the Bootstrap*. (Chapman & Hall, New York).
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica*, 50, 987-1008.
- Franke, J. and W. Härdle (1992). On Bootstrapping Kernel Spectral Estimates, *Annals of Statistics*, 20, 121-145.
- Franke, J., J.-P. Kreiss, and E. Mammen (2000). Bootstrap of Kernel Smoothing in Nonlinear Time Series, *Bernoulli*, forthcoming.
- Franke, J., J.-P. Kreiss, E. Mammen, and M.H. Neumann (2000). Properties of the Nonparametric Autoregressive Bootstrap, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany..
- 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 C. Hipp (1994). Asymptotic distribution of statistics in time series, *Annals of Statistics*, 22, 2062-2088.
- Götze F. and H.R. Künsch (1996). Blockwise bootstrap for dependent observations: higher order approximations for studentized statistics, *Annals of Statistics*, 24, 1914-1933.
- Gouriéroux, C. (1997). *ARCH Models and Financial Applications* (Springer-Verlag, New York).
- Härdle, W., and T. Kleinow, and A. Korostelev, and C. Logeay, and E. Platen (2001). Semiparametric Diffusion Estimation and Application to a Stock Market Index, Preprint No. 24, SFB 373, Humboldt-Universität Berlin.
- Hafner, C. (1996). *Nonlinear Time Series Analysis with Applications to foreign Exchange Rate Volatility* (Physica-Verlag Heidelberg).
- Hall, P. (1985). Resampling a coverage process, *Stochastic Process Applications* 19, 259-269.

- Hall, P. (1991). On the distribution of suprema, *Probability Theory and Related Fields*, 89, 447-455.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion* (Springer-Verlag, New York).
- 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.
- Hall, P. and B.-Y. Jing (1996). On sample reuse methods for dependent data, *Journal of the Royal Statistical Society, Series B*, 58, 727-737.
- Hansen, B. (1999). “Non-Parametric Dependent Data Bootstrap for Conditional Moment Models” working paper, Department of Economics, University of Wisconsin, Madison, WI.
- Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.
- Horowitz, J.L. (2001a). The bootstrap, in J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*, Vol. 5 (North-Holland Publishing Co., Amsterdam).
- Horowitz, J.L. (2001b). Bootstrap Methods for Markov Processes, working paper, Department of Economics, Northwestern University, Evanston, IL.
- Kreiss, J.-P. (1988). Asymptotic Statistical Inference for a Class of Stochastic Processes, Habilitationsschrift, University of Hamburg, Germany.
- Kreiss, J.-P. (1992). Bootstrap Procedures for $AR(\infty)$ Processes, in K.H. Jöckel, G. Rothe, and W. Sender (eds.), *Bootstrapping and Related Techniques*, Lecture Notes in Economics and Mathematical Systems, 376 (Springer-Verlag, Heidelberg), pp. 107-113.
- Kreiss, J.-P. (2000). Residual and Wild Bootstrap for Infinite Order Autoregressions, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
- Kreiss, J.-P., and M.H. Neumann, and Q. Yao (1998). Bootstrap Tests for Simple Structure in Nonparametric Time Series Regression, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
- Kreiss, J.-P. and E. Paparoditis (2000). Autoregressive Aided Periodogram Bootstrap for Time Series, working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.
- Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, 17, 1217-1241.
- Inoue A. and M. Shintani (2001). Bootstrapping GMM estimators for time series, working paper, Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC.

- Lahiri, S.N. (1991). Second order optimality of stationary bootstrap, *Statistics and Probability Letters*, 11, 335-341.
- Lahiri, S.N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data, in R. LePage and L. Billard (eds.), *Exploring the Limits of Bootstrap* (Wiley, New York).
- Lahiri, S.N. (1996a). On Edgeworth expansion and moving block bootstrap for Studentized M -estimators in multiple linear regression models, *Journal of Multivariate Analysis*, 56, 42-59.
- Lahiri, S.N. (1996b). Asymptotic expansions for sums of random vectors under polynomial mixing rates, *Sankhya*, Series A, 58, Pt. 2, 206-224.
- Lahiri, S.N. (1999). Theoretical comparisons of block bootstrap methods, *Annals of Statistics*, 27, 386-404.
- Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability* (Springer-Verlag, New York).
- Neumann, M.H., and J.-P. Kreiss (1998). Regression-type Inference in Nonparametric Autoregression, *Annals of Statistics*, 26, 1570-1613.
- Newey, W.K. and K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.
- Newey, W.K. and K.D. West (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, 61, 631-653.
- Nordgaard, A. (1992). Resampling a Stochastic Process Using a Bootstrap Approach, in: *Bootstrapping and Related Techniques* in K.H. Jöckel, G. Rothe, W. Sandler (eds.), *Lecture Notes in Economics and Mathematical Systems* 376 (Springer-Verlag, Berlin).
- 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 (2000). The Local Bootstrap for Markov Processes, *Journal of Statistical Planning and Inference*, forthcoming.
- Paparoditis, E. and D.N. Politis (2001). Tapered block bootstrap, *Journal of Statistical Planning and Inference*, forthcoming.
- Politis, D.N. and J.P. Romano (1993). The stationary bootstrap, *Journal of the American Statistical Association*, 89, 1303-1313.
- Politis, D.N. and J.P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions, *Annals of Statistics*, 22, 2031-2050.
- Politis, D.N., and J.P. Romano, and M. Wolf (1999). *Subsampling* (Springer-Verlag, New York).

- Rajarshi, M.B. (1990): Bootstrap in Markov-Sequences Based on Estimates of Transition Density,” *Annals of the Institute of Statistical Mathematics*, 42, 253-268.
- Theiler, J., and L.S. Paul, and D.M. Rubin (1994). Detecting Nonlinearity in Data with Long Coherence Time, in: *Time Series Prediction*, A. Weigend and N. Gershenfeld (eds.) (Addison-Welsey, Reading, MA).
- Zvingelis, J. (2001). On Bootstrap Coverage Probability with Dependent Data, in *Computer-Aided Econometrics*, D. Giles (ed.) (Marcel Dekker).