Statistical Science
2003, Vol. 18, No. 2, 211–218

The Bootstrap in Econometrics

Joel L. Horowitz

Abstract. This paper presents examples of problems in estimation and hypothesis testing that demonstrate the use and performance of the bootstrap in *econometric* settings. The examples are illustrated with two empirical applications. The paper concludes with a discussion of topics on which further research is needed.

Key words and phrases: Hypothesis test, asymptotic distribution, asymptotic refinement.

1. INTRODUCTION

Many important statistics in econometrics have complicated asymptotic distributions that depend on nuisance parameters and, therefore, cannot be tabulated. Examples include the conditional Kolmogorov test statistic of Andrews (1997) and Manski's (1975, 1985) maximum score estimator for a binary-response model. The bootstrap and related resampling techniques provide practical methods for estimating the asymptotic distributions of such statistics. In other cases, the statistics of interest have a standard distribution (e.g., bivariate normal) but with a complicated covariance matrix that is difficult to work with analytically (e.g., Horowitz and Manski, 2000). Again, the bootstrap provides a practical method for carrying out inference. Finally, it is not unusual for first-order asymptotic approximations to be very inaccurate with the sample sizes that are found in applications. When this happens, the difference between the true and nominal coverage probabilities (error in the coverage probability, or ECP) of confidence intervals can be very large. Similarly, the differences between the true and nominal probabilities that a test rejects a correct null hypothesis (error in the rejection probability, or ERP) can be very large. Consequently, inference based on first-order asymptotic approximations can be highly misleading. White's (1982) information matrix test is a well-known example of this, but there are many others. The bootstrap often greatly reduces the ECPs of confidence intervals and the ERPs of tests, thereby making reliable inference possible.

Joel L. Horowitz is Professor of Economics, Department of Economics, Northwestern University, Evanston, Illinois 60201.

Section 2 of this paper provides examples of problems in estimation and hypothesis testing that illustrate the use and performance of the bootstrap in econometric settings. Section 3 presents two empirical applications that are based on the examples of Section 2. Section 4 discusses topics on which further research is needed.

2. APPLICATIONS OF THE BOOTSTRAP AND OTHER RESAMPLING TECHNIQUES IN ECONOMETRICS

Section 2.1 provides examples of statistics with complicated asymptotic distributions that depend on nuisance parameters and, therefore, cannot be tabulated. Section 2.2 provides examples in which the bootstrap reduces the ERP of a test.

2.1 Statistics with Nonpivotal Asymptotic Distributions

EXAMPLE 1 [The conditional Kolmogorov test of Andrews (1997)]. Let $Y \in \mathbb{R}^V$ and $X \in \mathbb{R}^K$ be random variables. Let $H(y|x) = \boldsymbol{P}(Y \le y|X = x)$ and let $\Theta \subset \mathbb{R}^L$ be a parameter set. Andrews (1997) develops a test of the hypothesis $H_0: H(y|x) = F(y|x,\theta)$ almost surely for some function F and some $\theta \in \Theta$. The alternative hypothesis is that there is no $\theta \in \Theta$ such that $H(y|x) = F(y|x,\theta)$ for almost every x. Let $\{Y_i, X_i : i = 1, \dots, n\}$ be a random sample of (Y, X), θ_n be an estimator of θ that is $n^{-1/2}$consistent under H_0 and $\boldsymbol{I}$ be the indicator function. Then the conditional Kolmogorov test statistic is

$$CK_n = \max_{j \le n} \left| n^{-1/2} \sum_{i=1}^{n} \left[I(Y_i \le Y_j) - F(Y_j|X_i,\theta_n) \right] \boldsymbol{I}(X_i \le X_j) \right|.$$

Andrews gives conditions under which

$$CK_n \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z)|,$$

where $Z = \text{supp}(Y, X)$ and $\nu(\cdot)$ is a mean-zero Gaussian process whose covariance function depends on θ and the distribution of X. Critical values of CK_n cannot be tabulated. A bootstrap critical value can be obtained by drawing bootstrap samples from the estimated parametric model $F(y|x, \theta_n)$. Specifically, for each $i = 1, \ldots, n$, let $\hat{Y}_i$ be drawn randomly from the distribution whose cumulative distribution function (CDF) is $F(y|X_i, \theta_n)$. Let CK_{nb} denote the version of the test statistic that is obtained by using the bootstrap sample $\{\hat{Y}_i, X_i : i = 1, \ldots, n\}$. The α-level bootstrap critical value is then the $1 - \alpha$ quantile of the empirical distribution of CK_{nb} that is obtained from repeated bootstrap sampling. Andrews (1997) gives conditions under which the bootstrap procedure consistently estimates the asymptotic distribution of CK_n, and he provides Monte Carlo evidence indicating that the ERP of the test with bootstrap critical values is small with samples of practical size.

EXAMPLE 2 [Manski's (1975, 1985) maximum score estimator]. Manski considers the binary-response model

$$Y = \boldsymbol{I}(X'\beta - U \geq 0), \tag{1}$$

where $X \in \mathbb{R}^K$ is an observed random variable, X' is the transpose of the column vector X, β is a K-vector of constant parameters to be estimated and U is an unobserved random variable that satisfies $\text{median}(U|X = x) = 0$ almost surely. The distribution of U is assumed to satisfy mild regularity conditions but is otherwise unknown. This version of the binary-response model contains probit and logit models as special cases (i.e., if U has the standard normal or logistic distributions) but is much more flexible than a probit or logit model. For example, (1) permits U to have heteroscedasticity of unknown form. Let $\{Y_i, X_i : i = 1, \ldots, n\}$ be a random sample of (Y, X). The maximum score estimator of β is

$$b_n = \arg\max_{b \in B} \sum_{i=1}^{n} (2Y_i - 1)\boldsymbol{I}(X_i'b \geq 0), \tag{2}$$

where B is a compact parameter set. Since β is identified only up to scale, it may be assumed without loss of generality that $\|\beta\| = 1$ and that B is the surface of the unit sphere in $\mathbb{R}^K$. Manski (1985) gives conditions under which $b_n \to \beta$ almost surely as $n \to \infty$. The asymptotic distribution of the centered, scaled maximum score estimator was derived by Cavanagh (1987) and Kim and Pollard (1990). These investigators showed that $n^{1/3}(b_n - \beta)$ is distributed as the maximum of a Gaussian process with quadratic drift. The asymptotic distribution depends on unknown population parameters and, therefore, cannot be tabulated.

Manski and Thompson (1986) proposed using the bootstrap to carry out inference with the maximum score estimator and gave Monte Carlo evidence indicating that the bootstrap estimates the mean-square error of b_n accurately when $\boldsymbol{P}(Y = 1|X = x)$ is a continuous function of x. However, there is as yet no proof that the bootstrap consistently estimates the asymptotic distribution of the maximum score estimator. Moreover, Delgado, Rodríguez-Poo and Wolf (2001) provide Monte Carlo evidence indicating that the bootstrap does not provide correct critical values for testing hypotheses about β, which suggests that the bootstrap is inconsistent.

An alternative approach is to base inference on the subsampling procedure of Politis and Romano (1994). This procedure consistently estimates the asymptotic distribution of a statistic under very weak assumptions. Delgado, Rodríguez-Poo and Wolf (2001) describe the application of subsampling to the maximum score estimator and propose a data-based method for choosing the sizes of the subsamples. They also present Monte Carlo evidence in which the ERPs of tests of hypotheses about β are quite small with samples of 100–200 observations.

A third possibility is to replace the indicator function in (2) with a smooth function. Specifically, let K be a smooth function, possibly but not necessarily a distribution function, that satisfies $K(-\infty) = 0$ and $K(\infty) = 1$. Let $\{h_n : n = 1, 2, \ldots\}$ be a sequence of strictly positive constants (bandwidths) that satisfies $h_n \to 0$ as $n \to \infty$. The smoothed maximum score estimator of β, b_{ns}, satisfies

$$b_{ns} = \arg\max_{b \in B} \sum_{i=1}^{n} (2Y_i - 1)K(X_i'b/h_n). \tag{3}$$

Horowitz (1992) shows that under assumptions that are stronger than those of Manski (1975, 1985) but still quite weak, $n^r(b_{ns} - \beta)$ is asymptotically normal, where $2/5 \leq r < 1/2$ and the exact value of r depends on the smoothness of the distribution of $X'\beta$ and of $\boldsymbol{P}(Y = 1|X = x)$. Monte Carlo evidence suggests that the asymptotic normal approximation can be inaccurate with samples of practical size. However, Horowitz (2002) shows that the bootstrap, which is implemented by sampling the data randomly with replacement, provides asymptotic refinements for tests of hypotheses about β and produces low ERPs for these tests. Thus, the bootstrap provides a practical way to carry out inference with the smoothed maximum score estimator.

EXAMPLE 3 [The Box–Pierce (1970) test for serial correlation of a time series]. The Box–Pierce statistic is often used to test the hypothesis that a time series is independently and identically distributed (i.i.d.) against the alternative that at least one of its first K autocorrelation coefficients is nonzero. The test statistic is asymptotically chi-square distributed with K degrees of freedom when the null hypothesis is true, but it has a complicated asymptotic distribution that depends on nuisance parameters when, as often happens, for example, with data in finance, the time series is uncorrelated but serially dependent. Horowitz, Lobato, Nankervis and Savin (2003) proposed using a double blocks-of-blocks bootstrap procedure to obtain critical values for the Box–Pierce statistic when the time series of interest is not assumed to be i.i.d. under the null hypothesis. This procedure makes weaker assumptions about the data generation process than do several other tests of the null hypothesis of no serial correlation in the presence of serial dependence (e.g., Diebold, 1986; Romano and Thombs, 1996; Guo and Phillips, 1998; Lobato, Nankervis and Savin, 2001). Horowitz, Lobato, Nankervis and Savin (2003) present Monte Carlo evidence indicating that the Box–Pierce test with critical values based on their bootstrap procedure has low ERPs and power that is comparable to the power of a modified (though difficult to compute) version of the Box–Pierce statistic (Lobato, Nankervis and Savin, 2002) that is asymptotically chi-square distributed when the time series is serially uncorrelated but possibly serially dependent.

2.2 Asymptotic Refinements

It is well known that the bootstrap provides asymptotic refinements for tests and confidence intervals that are based on asymptotically pivotal statistics (see, e.g., Hall, 1992, among many others). This section presents two important econometric examples in which the bootstrap provides very large reductions in the ERP of a test or the ECP of a confidence interval.

EXAMPLE 4 [White's (1982) information matrix (IM) test]. This is a specification test for parametric models estimated by maximum likelihood. It tests the hypothesis that the Hessian and outer-product forms of the information matrix are equal. Rejection implies that the model is misspecified. The test statistic is asymptotically chi-square distributed, but the asymptotic distribution is a poor approximation of the finite-sample distribution. Monte Carlo experiments carried out by many investigators have shown that with asymptotic critical values and sample sizes in the range found in applications, the true and nominal probabilities of rejecting a correct model can differ by a factor of 10 or more (Taylor, 1987; Kennan and Neumann, 1988; Orme, 1990; Horowitz, 1994).

The IM test statistic satisfies the assumptions of Hall's (1992) smooth function model, so the bootstrap provides asymptotic refinements. Horowitz (1994) reports the results of Monte Carlo experiments that investigate the ERPs of the IM test with bootstrap critical values. Some of these results are summarized in Table 1. The table gives results for two asymptotically equivalent forms of the test: the Chesher (1983) and Lancaster (1984) form and White's (1982) original form. The Chesher–Lancaster form is easier to compute than the White form, which requires estimation of expected values of third derivatives of the log density, but the asymptotic chi-square approximation is especially poor for the Chesher–Lancaster form. The experiments reported here consisted of applying the two forms of the IM test to Tobit and binary probit models. [In a Tobit model, $Y = \max(X'\beta + U, 0)$, where $U \sim N(0, \sigma^2)$ independently of X, and β and σ^2 are unknown parameters.] The details are described in Horowitz (1994). It can be seen from Table 1 that the ERPs are very large when critical values based on the asymptotic chi-square distribution are used. When bootstrap critical values are used, however, the ERPs are very small. In the experiments reported in Table 1 as well as the additional experiments reported in Horowitz (1994), the bootstrap essentially eliminates the differences between the true and nominal rejection probabilities of the two forms of the IM test.

EXAMPLE 5 (Estimation of covariance structures). In estimation of covariance structures, the objective is to estimate the covariance matrix of a $k \times 1$ vector X subject to restrictions that reduce the number of unique, unknown elements to $r < k(k+1)/2$. Estimates of the r unknown elements can be obtained

TABLE 1
Empirical rejection probabilities of nominal 0.05-level information matrix tests of probit and Tobit models

		Rejection probability using			
	Distribution	Asymptotic critical values		Bootstrap critical values	
N	**of X**	**White**	**Chesher–Lancaster**	**White**	**Chesher-Lancaster**
		Binary probit models			
50	$N(0,1)$	0.385	0.904	0.064	0.056
	$U(-2,2)$	0.498	0.920	0.066	0.036
100	$N(0,1)$	0.589	0.848	0.053	0.059
	$U(-2,2)$	0.632	0.875	0.058	0.056
		Tobit models			
50	$N(0,1)$	0.112	0.575	0.083	0.047
	$U(-2,2)$	0.128	0.737	0.051	0.059
100	$N(0,1)$	0.065	0.470	0.038	0.039
	$U(-2,2)$	0.090	0.501	0.046	0.052

by minimizing the weighted distance between sample moments and the estimated population moments. Weighting all sample moments equally produces the equally weighted minimum distance (EWMD) estimator, whereas choosing the weights to maximize asymptotic estimation efficiency produces the optimal minimum distance (OMD) estimator.

The OMD estimator dominates the EWMD estimator in terms of asymptotic efficiency, but it has poor finite-sample performance in applications (Abowd and Card, 1989). Altonji and Segal (1994, 1996) carried out an extensive Monte Carlo investigation of the finite-sample performance of the OMD estimator. They found that the estimator is badly biased with samples of the sizes often found in applications and that its finite-sample root-mean-square estimation error (RMSE) often greatly exceeds the RMSE of the asymptotically inefficient EWMD estimator. In addition, the true coverage probabilities of asymptotic confidence intervals based on the OMD estimator tend to be much lower than the nominal coverage probabilities.

Horowitz (1998) reports the results of a Monte Carlo investigation of the ability of the bootstrap to reduce the ERPs of nominal 95% symmetrical confidence intervals based on the OMD estimator. In each experiment, X has 10 components, and the sample size is $n = 500$. The jth component of X, X_j $(j = 1, \ldots, 10)$ is generated by $X_j = (Z_j + \rho Z_{j+1})/(1 + \rho^2)^{1/2}$, where $Z_1, \ldots, Z_{11}$ are i.i.d. random variables with means of 0 and variances of 1, and $\rho = 0.5$. The Z's are sampled from five different distributions depending on the experiment. It is assumed that ρ is known and that the components of X are known to be identically distributed and to follow MA(1) processes. The estimation problem is to infer the scalar parameter θ that is identified by the moment conditions $\mathrm{Var}(X_j) = \theta$ $(j = 1, \ldots, 10)$ and $\mathrm{Cov}(X_j, X_{j-1}) = \rho\theta/(1 + \rho^2)$ $(j = 2, \ldots, 10)$.

The results of the experiments are summarized in Table 2. The coverage probabilities of confidence intervals based on asymptotic critical values are far below the nominal value of 0.95 except in the experiment with uniform Z's. However, the use of bootstrap critical values greatly reduces the ERPs. In the experiments with normal, Student t, uniform or exponential Z's, the bootstrap essentially eliminates the errors in the coverage probabilities of the confidence intervals.

TABLE 2
Empirical coverage probabilities of nominal 95% symmetrical confidence intervals based on the OMD estimator

Distribution of Z	**Asymptotic critical value**	**Bootstrap critical value**
Uniform	0.93	0.96
Normal	0.85	0.95
Student t with 10 d.f.	0.79	0.95
Exponential	0.54	0.96
Lognormal	0.03	0.91

3. EMPIRICAL APPLICATIONS

This section presents two empirical applications of the bootstrap. One consists of smoothed maximum score estimation of a model of the choice between automobile and transit for travel to work. The other consists of OMD estimation of the covariance structure of year-to-year changes in the logarithms of annual earnings and hours worked.

3.1 Smoothed Maximum Score Estimation of a Work Trip Mode Choice Model

Horowitz (1993) used the smoothed maximum score method to estimate the parameters of a model of the choice between automobile and transit for work trips in the Washington, DC, area. The model is given by (1), and the estimator of the parameter vector, β, is given by (3). The explanatory variables are defined in Table 3. Scale normalization is achieved by setting the coefficient of DCOST equal to 1. The data consist of 842 observations sampled randomly from the Washington, DC, area transportation study. Each record contains information about a single trip to work, including the chosen mode (automobile or transit) and the values of the explanatory variables. Column 2 of Table 3 shows the smoothed maximum score estimates of the model's parameters. Column 3 shows the half-widths of nominal 90% symmetrical confidence intervals based on the asymptotic normal approximation (the half-width equals 1.67 times the standard error of the estimate). Column 4 shows half-widths obtained from the bootstrap. The bootstrap confidence intervals are 2.5–3 times wider than the intervals based on the asymptotic normal approximation. The bootstrap confidence interval for the coefficient of DOVTT contains 0, but the confidence interval based on the asymptotic normal approximation does not. Therefore, the hypothesis that the coefficient of DOVTT is 0 is not rejected at the 0.1 level based on the bootstrap but is rejected based on the asymptotic normal approximation. Horowitz (2002) shows that the bootstrap provides asymptotic refinements for hypothesis tests and confidence intervals based on the smoothed maximum score estimator. Horowitz (2002) also presents Monte Carlo evidence indicating that the bootstrap reduces the ECPs of confidence intervals and the ERPs of hypothesis tests.

TABLE 3
Smoothed maximum score estimates of a work trip mode choice model

		Half-width of nominal 90% confidence interval based on	
Variable*	**Estimated coefficient**	**Asymptotic normal approximation**	**Bootstrap**
INTRCPT	−1.5761	0.2812	0.7664
AUTOS	2.2418	0.2989	0.7488
DOVTT	0.0269	0.0124	0.0310
DIVTT	0.0143	0.0033	0.0087
DCOST	1.0**		

*INTRCPT, intercept term equal to 1; AUTOS, number of cars owned by traveler's household; DOVTT, transit out-of-vehicle travel time minus automobile out-of-vehicle travel time (minutes); DIVTT, transit in-vehicle travel time minus automobile in-vehicle travel time; DCOST, transit fare minus automobile travel cost ($).
**Coefficient equal to 1 by scale normalization.

3.2 Estimation of Covariance Structures

This section reports results of OMD estimation of the covariance structure of year-to-year changes in the logarithms of annual earnings and working hours of male heads of households in the Panel Study of Income Dynamics (PSID). The results are taken from Horowitz (1998), who provides further details on the estimation method.

The data are those used by Altonji and Segal (1994, 1996). They are based on observations of annual earnings and hours worked of 1536 individuals over the 11-year period 1969–1979. The data provide 10 observations per individual on year-to-year changes in the logarithms of earnings and hours. Accordingly, the covariance matrix of the observations contains 210 separate moments for the various years and lags. However, Abowd and Card (1989) found that the covariances of observations separated by more than two years are negligible. Accordingly, the estimates reported here are based on a vector of 98 variances and covariances with time lags of up to two years. The estimated model is stationary, so there are only $r = 11$ separate covariance parameters. Table 4 shows the estimated covariances and the half-widths of nominal 95% symmetrical confidence intervals based on the asymptotic normal approximation and the bootstrap. The bootstrap confidence intervals are 2–3 times wider than the intervals based on the asymptotic normal approximation. As was discussed in Section 2, there are both theoretical arguments and Monte Carlo evidence indicating that the bootstrap intervals have smaller ECPs than do the intervals based on the asymptotic normal approximation.

TABLE 4
OMD estimates of covariances of logarithms of earnings and hours

Covariance parameter*	Estimate	Half-width of nominal 95% confidence interval based on: Asymptotic normal approximation	Bootstrap
$E(t)$, $E(t)$	0.173	0.011	0.025
$E(t)$, $E(t-1)$	−0.056	0.005	0.012
$E(t)$, $E(t-2)$	−0.010	0.003	0.008
$H(t)$, $H(t)$	0.118	0.008	0.018
$H(t)$, $H(t-1)$	−0.039	0.004	0.009
$H(t)$, $H(t-2)$	−0.012	0.002	0.007
$E(t)$, $H(t)$	0.077	0.007	0.016
$E(t)$, $H(t-1)$	−0.021	0.003	0.005
$E(t)$, $H(t-2)$	−0.001	0.003	0.008
$H(t)$, $E(t-1)$	−0.023	0.004	0.010
$H(t)$, $E(t-2)$	−0.015	0.003	0.007

* E and H denote logarithms of earnings and hours, respectively. $E(t)$, $E(t)$ denotes variance of logarithm of earnings; $E(t)$, $H(t)$ denotes covariance of logarithms of earnings and hours; $E(t)$, $E(t-1)$ denotes covariance of logarithms of earnings and earnings lagged by one year, etc.

4. UNSOLVED PROBLEMS: THE BOOTSTRAP FOR TIME-SERIES DATA

Applied econometric research often involves inference based on time-series data. Accordingly, there is much interest in using the bootstrap to reduce the ERPs of tests and the ECPs of confidence intervals obtained from time-series data. This raises the question of how bootstrap samples should be generated. In many applications, it is undesirable to assume that the data generation process (DGP) belongs to a known, finite-dimensional, parametric family (e.g., a finite-order ARMA model), so a nonparametric method for bootstrap sampling is needed. If the DGP is strictly stationary, then the block bootstrap is the best-known such method. However, blocking distorts the dependence structure of the DGP, thereby causing the errors of bootstrap estimates to converge more slowly than is the case with i.i.d. data. For example, when the block length is chosen optimally, the errors in block-bootstrap estimates of one-sided and symmetrical distribution functions are $O(n^{-3/4})$ and $O(n^{-6/5})$, respectively (Hall, Horowitz and Jing, 1995), compared to $O(n^{-1})$ and $O(n^{-3/2})$ for the bootstrap with i.i.d. data. The errors made by the asymptotic normal approximation are $O(n^{-1/2})$ and $O(n^{-1})$ for one-sided and symmetrical distribution functions, respectively. Thus, the improvement in accuracy available from the block-bootstrap is considerably less than that from the bootstrap for i.i.d. data. In the case of estimating a symmetrical distribution function, the rate of convergence of the block-bootstrap estimation error is only slightly faster than the rate of convergence of the error made by the asymptotic normal approximation. Similarly, the rates of convergence of block bootstrap ERPs and ECPs are relatively slow (Zvingelis, 2001; Andrews, 2002a). Monte Carlo results have confirmed this disappointing performance of the block bootstrap (Hall and Horowitz, 1996).

The 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) proposed a sieve bootstrap for linear processes. 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) showed 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. However, the assumption of linearity is too strong for many applications.

Another possibility is to assume that the DGP is a (possibly higher order) Markov process or a process that is well approximated in a suitable sense by a Markov process. The Markov transition density can be estimated nonparametrically and bootstrap samples generated from the DGP that is implied by the estimated density. This approach has been investigated by Rajarshi (1990), Datta and McCormick (1995), Paparoditis and Politis (2001, 2002) and Horowitz (2003). Horowitz (2003) gives conditions under which the errors in Markov bootstrap estimates of one-sided and symmetrical distribution functions and the ERPs of one-sided and symmetrical hypothesis tests are $O(n^{-1+\varepsilon})$ and $O(n^{-3/2+\varepsilon})$, respectively, for any $\varepsilon > 0$. Thus, the Markov bootstrap improves on the block bootstrap when the DGP is a Markov process that satisfies appropriate regularity conditions. However, the Markov bootstrap suffers from a form of the curse of dimensionality of nonparametric estimation owing to the need to carry out nonparametric estimation of the transition density. Consequently, the Markov bootstrap is likely to be most suitable in practice for DGPs that are either low-order Markov processes or can be approximated well by low-order processes.

It is largely an open question whether there are practically useful bootstrap methods for time-series data whose assumptions about the structure of the DGP are weaker than those of the Markov bootstrap but that achieve rates of convergence of ERPs and ECPs that are the same as or faster than those of the Markov bootstrap. One possible approach is a modified block-bootstrap procedure that is described by Andrews (2002b). More generally, the incompleteness of the current theory of the bootstrap's ability to achieve asymptotic refinements is an obstacle to research about the application of the bootstrap to time series. The current theory is based on Edgeworth expansions. Essentially, the bootstrap amounts to carrying out a low-order Edgeworth expansion of the statistic in question. However, if the DGP is strictly stationary and geometrically strongly mixing, the relevant moments can be estimated with rates of convergence in probability that are only slightly slower than $O(n^{-1/2})$. This result can be used to show that the errors made by an empirical Edgeworth expansion converge more rapidly than do the errors of either the block or the Markov bootstrap. However, the results of Monte Carlo simulations show that the numerical accuracy of an empirical Edgeworth expansion is much worse than that of the bootstrap. Thus, the current Edgeworth-based theory provides an inadequate guide to the numerical performance of methods for improving upon first-order asymptotic approximations. This problem is especially serious with time-series data because a method whose errors converge relatively rapidly (an empirical Edgeworth expansion) has worse numerical performance than methods whose errors converge relatively slowly (the block and Markov bootstraps). Accordingly, it would be useful (though undoubtedly very difficult) to develop a theory of the bootstrap that provides a more reliable guide to its finite-sample performance.

Bootstrap methods for DGPs with unit roots or cointegration is another area where there are important opportunities for further research. Methods for obtaining consistent bootstrap estimates when the DGP has or may have a unit root are now available (Basawa et al., 1991; Datta, 1996; Ferretti and Romo, 1996; Inoue and Kilian, 2002; Paparoditis and Politis, 2003; Park, 2002). However, research is only just beginning on the ability of the bootstrap to provide asymptotic refinements when the DGP may have a unit root (Park, 2002). Similarly, there has been little research on applying the bootstrap to cointegrated data.

ACKNOWLEDGMENT

Supported in part by NSF Grant SES-99-10925.

REFERENCES

ABOWD, J. M. and CARD, D. (1989). On the covariance structure of earnings and hours changes. *Econometrica* **57** 411–445.

ALTONJI, J. G. and SEGAL, L. M. (1994). Small sample bias in GMM estimation of covariance structures. NBER Technical Working Paper 156, National Bureau of Economic Research, Cambridge, MA.

ALTONJI, J. G. and SEGAL, L. M. (1996). Small sample bias in GMM estimation of covariance structures. *J. Bus. Econom. Statist.* **14** 353–366.

ANDREWS, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica* **65** 1097–1128.

ANDREWS, D. W. K. (2002a). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* **70** 119–162.

ANDREWS, D. W. K. (2002b). The block–block bootstrap: Improved asymptotic refinements. Working paper, Cowles Foundation for Research in Economics, Yale Univ.

BASAWA, I. V., MALLIK, A. K., MCCORMICK, W. P., REEVES, J. H. and TAYLOR, R. L. (1991). Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Comm. Statist. Theory Methods* **20** 1015–1026.

BOX, G. E. P. and PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *J. Amer. Statist. Assoc.* **65** 1509–1526.

BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148.

BÜHLMANN, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *Ann. Statist.* **26** 48–83.

CAVANAGH, C. L. (1987). Limiting behavior of estimators defined by optimization. Unpublished manuscript.

CHESHER, A. (1983). The information matrix test. *Econom. Lett.* **13** 45–48.

CHOI, E. and HALL, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 461–477.

DATTA, S. (1996). On asymptotic properties of bootstrap for AR(1) processes. *J. Statist. Plann. Inference* **53** 361–374.

DATTA, S. and MCCORMICK, W. P. (1995). Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. *J. Multivariate Anal.* **52** 83–106.

DELGADO, M. A., RODRÍGUEZ-POO, J. M. and WOLF, M. (2001). Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Econom. Lett.* **73** 241–250.

DIEBOLD, F. X. (1986). Testing for serial correlation in the presence of ARCH. In *Proc. Business and Economics Statistics Section* 323–328. Amer. Statist. Assoc., Alexandria, VA.

FERRETTI, N. and ROMO, J. (1996). Unit root bootstrap tests for AR(1) models. *Biometrika* **83** 849–860.

GUO, B. B. and PHILLIPS, P. C. B. (1998). Testing for autocorrelation and unit roots in the presence of conditional heteroskedasticity of unknown form. Unpublished manuscript.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

HALL, P. and HOROWITZ, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* **64** 891–916.

HALL, P., HOROWITZ, J. L. and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574.

HOROWITZ, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* **60** 505–531.

HOROWITZ, J. L. (1993). Semiparametric estimation of a work-trip mode choice model. *J. Econometrics* **58** 49–70.

HOROWITZ, J. L. (1994). Bootstrap-based critical values for the information matrix test. *J. Econometrics* **61** 395–411.

HOROWITZ, J. L. (1998). Bootstrap methods for covariance structures. *J. Human Resources* **33** 39–61.

HOROWITZ, J. L. (2002). Bootstrap critical values for tests based on the smoothed maximum score estimator. *J. Econometrics* **111** 141–167.

HOROWITZ, J. L. (2003). Bootstrap methods for Markov processes. *Econometrica* **71** 1049–1082.

HOROWITZ, J. L., LOBATO, I., NANKERVIS, J. C. and SAVIN, N. E. (2003). Bootstrapping the Box–Pierce test: A robust test of uncorrelatedness. *J. Econometrics*. To appear.

HOROWITZ, J. L. and MANSKI, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data (with discussion). *J. Amer. Statist. Assoc.* **95** 77–88.

INOUE, A. and KILIAN, L. (2002). Bootstrapping autoregressive processes with possible unit roots. *Econometrica* **70** 377–391.

KENNAN, J. and NEUMANN, G. R. (1988). Why does the information matrix test reject so often? Working paper, Dept. Economics, Univ. Iowa, Iowa City.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

KREISS, J.-P. (1992). Bootstrap procedures for AR(∞) processes. In *Bootstrapping and Related Techniques* (K. H. Jöckel, G. Rothe and W. Sender, eds.) 107–113. Springer, Berlin.

LANCASTER, T. (1984). The covariance matrix of the information matrix test. *Econometrica* **52** 1051–1053.

LOBATO, I., NANKERVIS, J. C. and SAVIN, N. E. (2001). Testing for autocorrelation using a modified Box–Pierce Q test. *Internat. Econom. Rev.* **42** 187–205.

LOBATO, I., NANKERVIS, J. C. and SAVIN, N. E. (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory* **18** 730–743.

MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228.

MANSKI, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *J. Econometrics* **27** 313–333.

MANSKI, C. F. and THOMPSON, T. S. (1986). Operational characteristics of maximum score estimation. *J. Econometrics* **32** 85–108.

ORME, C. (1990). The small-sample performance of the information matrix test. *J. Econometrics* **46** 309–331.

PAPARODITIS, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *J. Multivariate Anal.* **57** 277–296.

PAPARODITIS, E. and POLITIS, D. N. (2001). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory* **17** 540–566.

PAPARODITIS, E. and POLITIS, D. N. (2002). The local bootstrap for Markov processes. *J. Statist. Plann. Inference* **108** 301–328.

PAPARODITIS, E. and POLITIS, D. N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica* **71** 813–855.

PARK, J. Y. (2002). Bootstrap unit root tests. *Econometrica*. To appear.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.

RAJARSHI, M. B. (1990). Bootstrap in Markov-sequences based on estimates of transition density. *Ann. Inst. Statist. Math.* **42** 253–268.

ROMANO, J. P. and THOMBS, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91** 590–600.

TAYLOR, L. W. (1987). The size bias of White's information matrix test. *Econom. Lett.* **24** 63–67.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–26.

ZVINGELIS, J. (2001). On bootstrap coverage probability with dependent data. In *Computer-Aided Econometrics* (D. Giles, ed.). Dekker, New York.