THE BOOTSTRAP

by

Joel L. Horowitz
Department of Economics
University of Iowa
Iowa City, IA 52242

March 1999

Prepared for the *Handbook of Econometrics*, vol. 5. I thank Donald Andrews, James Heckman, Hidehiko Ichimura, John Kennan, Clive Loader, George Neumann, Harry Paarsch, Bernard Salanié, and Gene Savin for comments on earlier drafts of this chapter. Parts of the chapter are taken from Horowitz (1997) and are used here with permission from the Cambridge University Press. The preparation of this chapter was supported in part by NSF Grant No. SBR-9617925.

THE BOOTSTRAP

1. INTRODUCTION

The bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data. It amounts to treating the data as if they were the population for the purpose of evaluating the distribution of interest. Under mild regularity conditions, the bootstrap yields an approximation to the distribution of an estimator or test statistic that is at least as accurate as the approximation obtained from first-order asymptotic theory. Thus, the bootstrap provides a way to substitute computation for mathematical analysis if calculating the asymptotic distribution of an estimator or statistic is difficult. The statistic developed by Härdle *et al*. (1991) for testing positive-definiteness of income-effect matrices, the conditional Kolmogorov test of Andrews (1997a), Stute's (1997) specification test for parametric regression models, and certain functions of time-series data (Blanchard and Quah 1989, Runkle 1987, West 1990) are examples in which evaluating the asymptotic distribution is difficult and bootstrapping has been used as an alternative.

In fact, the bootstrap is often more accurate in finite samples than first-order asymptotic approximations but does not entail the algebraic complexity of higher-order expansions. Thus, it can provide a practical method for improving upon first-order approximations. Such improvements are called *asymptotic refinements*. One use of the bootstrap's ability to provide asymptotic refinements is bias reduction. It is not unusual for an asymptotically unbiased estimator to have a large finite-sample bias. This bias may cause the estimator's finite-sample mean square error to greatly exceed the mean-square error implied by its asymptotic distribution. The bootstrap can be used to reduce the estimator's finite-sample bias and, thereby, its finite-sample mean-square error.

The bootstrap's ability to provide asymptotic refinements is also important in hypothesis testing. First-order asymptotic theory often gives poor approximations to the distributions of test statistics with the sample sizes available in applications. As a result, the nominal probability that a test based on an asymptotic critical value rejects a true null hypothesis can be very different from the true rejection probability (RP).[1] The information matrix test of White (1982) is a well-known example of a test in which large finite-sample errors in the RP can occur when asymptotic critical values are used (Horowitz 1994, Kennan and Neumann 1988, Orme 1990, Taylor 1987). Other illustrations are given later in this chapter. The bootstrap often provides a tractable way to reduce or eliminate finite-sample errors in the RP's of statistical tests.

The problem of obtaining critical values for test statistics is closely related to that of obtaining confidence intervals. Accordingly, the bootstrap can also be used to obtain confidence intervals with reduced errors in coverage probabilities. That is, the difference between the true and nominal coverage probabilities is often lower when the bootstrap is used than when first-order asymptotic approximations are used to obtain a confidence interval.

The bootstrap has been the object of much research in statistics since its introduction by Efron (1979). The results of this research are synthesized in the books by Beran and Ducharme (1991), Davison and Hinkley (1997), Efron and Tibshirani (1993), Hall (1992a), Mammen (1992), and Shao and Tu (1995). Hall (1994), Horowitz (1997), Maddala and Jeong (1996) and Vinod (1993) provide reviews with an econometric orientation. This chapter covers a broader range of topics than do these reviews. Topics that are treated here but only briefly or not at all in the reviews include bootstrap consistency, subsampling, bias reduction, time-series models with unit roots, semiparametric and nonparametric models, and certain types of non-smooth models. Some of these topics are not treated in existing books on the bootstrap.

The purpose of this chapter is to explain and illustrate the usefulness and limitations of the bootstrap in contexts of interest in econometrics. Particular emphasis is given to the bootstrap's ability to improve upon first-order asymptotic approximations. The presentation is informal and expository. Its aim is to provide an intuitive understanding of how the bootstrap works and a feeling for its practical value in econometrics. The discussion in this chapter does not provide a mathematically detailed or rigorous treatment of the theory of the bootstrap. Such treatments are available in the books by Beran and Ducharme (1991) and Hall (1992a) as well as in journal articles that are cited later in this chapter.

It should be borne in mind throughout this chapter that although the bootstrap often provides smaller biases, smaller errors in the RP's of tests, and smaller errors in the coverage probabilities of confidence intervals than does first-order asymptotic theory, bootstrap bias estimates, RP's, and confidence intervals are, nonetheless, approximations and not exact. Although the accuracy of bootstrap approximations is often very high, this is not always the case. Even when theory indicates that it provides asymptotic refinements, the bootstrap's numerical performance may be poor. In some cases, the numerical accuracy of bootstrap approximations may be even worse than the accuracy of first-order asymptotic approximations. This is particularly likely to happen with estimators whose asymptotic covariance matrices are "nearly singular," as in instrumental-variables estimation with poorly correlated instruments and regressors. Thus, the bootstrap should not be used blindly or uncritically.

However, in the many cases where the bootstrap works well, it essentially removes getting the RP or coverage probability right as a factor in selecting a test statistic or method for constructing a confidence interval. In addition, the bootstrap can provide dramatic reductions in the finite-sample biases and mean-square errors of certain estimators.

The remainder of this chapter is divided into five sections. Section 2 explains the bootstrap sampling procedure and gives conditions under which the bootstrap distribution of a statistic is a consistent estimator of the statistic's asymptotic distribution. Section 3 explains when and why the bootstrap provides asymptotic refinements. This section concentrates on data that are simple random samples from a distribution and statistics that are either smooth functions of sample moments or can be approximated with asymptotically negligible error by such functions (the smooth function model). Section 4 extends the results of Section 3 to dependent data and statistics that do not satisfy the assumptions of the smooth function model. Section 5 presents Monte Carlo evidence on the numerical performance of the bootstrap in a variety of settings that are relevant to econometrics, and Section 6 presents concluding comments.

2. THE BOOTSTRAP SAMPLING PROCEDURE AND ITS CONSISTENCY

The bootstrap is a method for estimating the distribution of a statistic or a feature of the distribution, such as a moment or a quantile. This section explains how the bootstrap is implemented in simple settings and gives conditions under which it provides a consistent estimator of a statistic's asymptotic distribution. This section also gives examples in which the consistency conditions are not satisfied and the bootstrap is inconsistent.

The estimation problem to be solved may be stated as follows. Let the data be a random sample of size n from a probability distribution whose cumulative distribution function (CDF) is F_0. Denote the data by $\{X_i: i = 1, ..., n\}$. Let F_0 belong to a finite- or infinite-dimensional family of distribution functions, $\Im$. Let F denote a general member of $\Im$. If $\Im$ is a finite-dimensional family indexed by the parameter q whose population value is q_0, write $F_0(x, q_0)$ for $P(X \leq x)$ and $F(x, q)$ for a general member of the parametric family. Let $T_n = T_n(X_1, \ldots, X_n)$ be a statistic (that is, a function of the data). Let $G_n(t, F_0) \equiv P(T_n \leq t)$ denote the exact, finite-sample CDF of T_n Let $G_n(\cdot, F)$ denote the exact CDF of T_n when the data are sampled from the distribution whose CDF is F. Usually, $G_n(t, F)$ is a different function of t for different distributions F. An exception occurs if $G_n(\cdot, F)$ does not depend on F, in which case T_n is said to be *pivotal*. For example, the t statistic for testing a hypothesis about the mean of a normal population is independent of unknown population parameters and, therefore, is pivotal. The same is true of the t statistic for testing a hypothesis about a slope coefficient in a normal linear regression model. Pivotal statistics are not available in most econometric applications, however, especially without making strong distributional assumptions (e.g., the assumption that the random component of a linear regression model is normally distributed). Therefore, $G_n(\cdot, F)$ usually depends on F, and $G_n(\cdot, F_0)$ cannot be calculated if, as is usually the case in applications, F_0 is unknown. The bootstrap is a method for estimating $G_n(\cdot, F_0)$ or features of $G_n(\cdot, F_0)$ such as its quantiles when F_0 is unknown.

Asymptotic distribution theory is another method for estimating $G_n(\cdot, F_0)$. The asymptotic distributions of many econometric statistics are standard normal or chi-square, possibly after centering and normalization, regardless of the distribution from which the data were sampled. Such statistics are called *asymptotically pivotal*, meaning that their asymptotic distributions do not depend on unknown population parameters. Let $G_\infty(\cdot, F_0)$ denote the asymptotic distribution of T_n. Let $G_\infty(\cdot, F)$ denote the asymptotic CDF of T_n when the data are sampled from the distribution whose CDF is F. If T_n is asymptotically pivotal, then $G_\infty(\cdot, F) \equiv G_\infty(\cdot)$ does not depend on F. Therefore, if n is sufficiently large, $G_n(\cdot, F_0)$ can be estimated by $G_\infty(\cdot)$ without knowing F_0. This method for estimating $G_n(\cdot, F_0)$ is often easy to implement and is widely used. However, as was discussed in Section 1, $G_\infty(\cdot)$ can be a very poor approximation to $G_n(\cdot, F_0)$ with samples of the sizes encountered in applications.

Econometric parameter estimators usually are not asymptotically pivotal (that is, their asymptotic distributions usually depend on one or more unknown population parameters), but many are asymptotically normally distributed. If an estimator is asymptotically normally distributed, then its asymptotic distribution depends on at most two unknown parameters, the mean and the variance, that can often be estimated without great difficulty. The normal distribution with the estimated mean and variance can then be used to approximate the unknown $G_n(\cdot, F_0)$ if n is sufficiently large.

The bootstrap provides an alternative approximation to the finite-sample distribution of a statistic $T_n(X_1, \ldots, X_n)$. Whereas first-order asymptotic approximations replace the unknown distribution function G_n with the known function G_∞, the bootstrap replaces the unknown distribution function F with a known estimator. Let F_n denote the estimator of F_0. Two possible choices of F_n are:

(1) The empirical distribution function (EDF) of the data:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x),$$

where I is the indicator function. It follows from the Glivenko-Cantelli theorem that $F_n(x) \to F_0(x)$ as $n \to \infty$ uniformly over x almost surely.

(2) A parametric estimator of F_0. Suppose that $F_0(\cdot) = F(\cdot, q_0)$ for some finite-dimensional q_0 that is estimated consistently by q_n. If $F(\cdot, q)$ is a continuous function of q in a neighborhood of q_0, then $F(x, q_n) \to F(x, q_0)$ as $n \to \infty$ at each x. The convergence is in probability or almost sure according to whether $q_n \to q_0$ in probability or almost surely.

Other possible F_n's are discussed in Section 3.7.

Regardless of the choice of F_n, the bootstrap estimator of $G_n(\cdot, F_0)$ is $G_n(\cdot, F_n)$. Usually, $G_n(\cdot, F_n)$ cannot be evaluated analytically. It can, however, be estimated with arbitrary accuracy by carrying out a Monte Carlo simulation in which random samples are drawn from F_n. Thus, the bootstrap is usually implemented by Monte Carlo simulation. The Monte Carlo procedure for estimating $G_n(t, F_0)$ is as follows

Monte Carlo Procedure for Bootstrap Estimation of $G_n(t, F_0)$

Step 1: Generate a bootstrap sample of size n, $\{X_i^*: i = 1, \ldots, n\}$, by sampling the distribution corresponding to F_n randomly. If F_n is the EDF of the estimation data set, then the bootstrap sample can be obtained by sampling the estimation data randomly with replacement.

Step 2: Compute $T_n^* \equiv T_n(X_1^*, \ldots, X_n^*)$.

Step 3: Use the results of many repetitions of steps 1 and 2 to compute the empirical probability of the event $T_n^* \le t$ (that is, the proportion of repetitions in which this event occurs).

Procedures for using the bootstrap to compute other statistical objects are described in Sections 3.1 and 3.3. Brown (1999) and Hall (1992a, Appendix II) discuss simulation methods that take advantage of techniques for reducing sampling variation in Monte Carlo simulation. The essential characteristic of the bootstrap, however, is the use of F_n to approximate F_0 in $G_n(\cdot, F_0)$, not the method that is used to evaluate $G_n(\cdot, F_n)$.

Since F_n and F_0 are different functions, $G_n(\cdot, F_n)$ and $G_n(\cdot, F_0)$ are also different functions unless T_n is pivotal. Therefore, the bootstrap estimator $G_n(\cdot, F_n)$ is only an approximation to the exact finite-sample CDF of T_n, $G_n(\cdot, F_0)$. Section 3 discusses the accuracy of this approximation. The remainder of this section is concerned with conditions under which $G_n(\cdot, F_n)$ satisfies the minimal criterion for adequacy as an estimator of $G_n(\cdot, F_0)$, namely consistency. Roughly speaking, $G_n(\cdot, F_n)$ is consistent if it converges in probability to the asymptotic CDF of T_n, $G_\infty(\cdot, F_0)$, as $n \to \infty$. Section 2.1 defines consistency precisely and gives conditions under which it holds. Section 2.2 describes some resampling procedures that can be used to estimate $G_n(\cdot, F_0)$ when the bootstrap is not consistent.

2.1 *Consistency of the Bootstrap*

Suppose that F_n is a consistent estimator of F_0. This means that at each x in the support of X, $F_n(x) \to F_0(x)$ in probability or almost surely as $n \to \infty$. If F_0 is a continuous function, then it follows from Polya's theorem that $F_n \to F_0$ in probability or almost surely uniformly over x. Thus, F_n and F_0 are uniformly close to one another if n is large. If, in addition, $G_n(t, F)$ considered as a functional of F is continuous in an appropriate sense, it can be expected that $G_n(t, F_n)$ is close to $G_n(t, F_0)$ when n is large. On the other hand, if n is large, then $G_n(\cdot, F_0)$ is uniformly close to the asymptotic distribution $G_\infty(\cdot, F_0)$ if $G_\infty(\cdot, F_0)$ is continuous. This suggests that the bootstrap estimator $G_n(\cdot, F_n)$ and the asymptotic distribution $G_\infty(\cdot, F_0)$ should be uniformly close if n is large and suitable continuity conditions hold. The definition of consistency of the bootstrap formalizes this idea in a way that takes account of the randomness of the function $G_n(\cdot, F_n)$. Let $\mathfrak{I}$ denote the space of permitted distribution functions.

Definition 2.1: *Let P_n denote the joint probability distribution of the sample* $\{X_i$: $i = 1, \ldots, n\}$. *The bootstrap estimator $G_n(\cdot, F_n)$ is consistent if for each* $e > 0$ *and* $F_0 \in \Im$

$$\lim_{n \to \infty} P_n \left[\sup_t \left| G_n(t, F_n) - G_\infty(t, F_0) \right| > e \right] = 0 .$$

A theorem by Beran and Ducharme (1991) gives conditions under which the bootstrap estimator is consistent. This theorem is fundamental to understanding the bootstrap. Let r denote a metric on the space $\Im$ of permitted distribution functions.

Theorem 2.1 (Beran and Ducharme 1991): $G_n(\cdot, F_n)$ *is consistent if for any* $e > 0$ *and* $F_0 \in \Im$: (i) $\lim_{n \to \infty} P_n[r(F_n, F_0) > e] = 0$; (ii) $G_\infty(t, F)$ *is a continuous function of* t *for each* $F \in \Im$; *and* (iii) *for any t and any sequence* $\{H_n\} \in \Im$ *such that* $\lim_{n \to \infty} r(H_n, F_0) = 0$, $G_n(t, H_n) \to G_\infty(t, F_0)$.

The following is an example in which the conditions of Theorem 2.1 are satisfied:

Example 2.1 (The distribution of the sample average): Let $\Im$ be the set of distribution functions F corresponding to populations with finite variances. Let $\bar{X}$ be the average of the random sample $\{X_i$: $i = 1, \ldots, n\}$. Define $T_n = n^{1/2}(\bar{X} - m)$, where $m = E(X)$. Let $G_n(t, F_0) = P_n\left[n^{1/2}(\bar{X} - m) \leq t\right]$. Consider using the bootstrap to estimate $G_n(t, F_0)$. Let F_n be the EDF of the data. Then the bootstrap analog of T_n is $T_n^* = n^{1/2}(\bar{X}^* - \bar{X})$, where $\bar{X}^*$ is the average of a random sample of size n drawn from F_n (the bootstrap sample). The bootstrap sample can be obtained by sampling the data $\{X_i\}$ randomly with replacement. T_n^* is centered at $\bar{X}$ because $\bar{X}$ is the mean of the distribution from which the bootstrap sample is drawn. The bootstrap estimator of $G_n(t, F_0)$ is $G_n(t, F_n) = P_n^*\left[n^{1/2}(\bar{X}^* - \bar{X}) \leq t\right]$, where P_n^* is the probability distribution induced by the bootstrap sampling process. $G_n(t, F_n)$ satisfies the conditions of Theorem 2.1 and, therefore, is consistent. Let r be the Mallows metric.[2] The Glivenko-Cantelli theorem and the strong law of large numbers imply that condition (i) of Theorem 2.1 is satisfied.

The Lindeberg-Levy central limit theorem implies that T_n is asymptotically normally distributed. The cumulative normal distribution function is continuous, so condition (ii) holds. By using arguments similar to those used to prove the Lindeberg-Levy theorem, it can be shown that condition (iii) holds. ∎

A theorem by Mammen (1992) gives necessary and sufficient conditions for the bootstrap to consistently estimate the distribution of a linear functional of F_0 when F_n is the EDF of the data. This theorem is important because the conditions are often easy to check, and many estimators and test statistics of interest in econometrics are asymptotically equivalent to linear functionals of some F_0. Hall (1990) and Gill (1989) give related theorems.

Theorem 2.2 (Mammen 1992): *Let* $\{X_i: \ i = 1, \ldots, n\}$ *be a random sample from a population. For a sequence of functions* g_n *and sequences of numbers* t_n *and* s_n, *define* $\bar{g}_n = n^{-1}\sum_{i=1}^{n} g_n(X_i)$ *and* $T_n = (\bar{g}_n - t_n)/s_n$. *For the bootstrap sample* $\{X_i^*: \ i = 1, \ldots, n\}$, *define* $\bar{g}_n^* = n^{-1}\sum_{i=1}^{n} g_n(X_i^*)$ *and* $T_n^* = (\bar{g}_n^* - \bar{g}_n)/s_n$. *Let* $G_n(t) = P(T_n \leq t)$ *and* $G_n^*(t) = P^*(T_n^* \leq t)$, *where* P^* *is the probability distribution induced by bootstrap sampling. Then* $G_n^*(\cdot)$ *consistently estimates* G_n *if and only if* $T_n \to^d N(0,1)$. ∎

If $E[g_n(X)]$ and $Var[g_n(X)]$ exist for each n, then the asymptotic normality condition of Theorem 2.2 holds with $t_n = E(\bar{g}_n)$ and $s_n^2 = Var(\bar{g}_n)$ or $s_n^2 = n^{-2}\sum_{i=1}^{n}[g_n(X_i) - \bar{g}_n]^2$. Thus, consistency of the bootstrap estimator of the distribution of the centered, normalized sample average in Example 2.1 follows trivially from Theorem 2.2.

The bootstrap need not be consistent if the conditions of Theorem 2.1 are not satisfied and is inconsistent if the asymptotic normality condition of Theorem 2.2 is not satisfied. In particular, the bootstrap tends to be inconsistent if F_0 is a point of discontinuity of the asymptotic distribution function $G_\infty(t, \cdot)$ or a point of superefficiency. Section 2.2 describes resampling methods that can sometimes be used to overcome these difficulties.

The following examples illustrate conditions under which the bootstrap is inconsistent. The conditions that cause inconsistency in the examples are unusual in econometric practice. The bootstrap is consistent in most applications. Nonetheless, inconsistency sometimes occurs, and it is important to be aware of its causes. Donald and Paarsch (1996), Flinn and Heckman (1982), and Heckman, Smith, and Clements (1997) describe econometric applications that have features similar to those of some of the examples, though the consistency of the bootstrap in these applications has not been investigated.

Example 2.2 (Heavy-tailed distributions): Let F_0 be the standard Cauchy distribution and $\{X_i\}$ be a random sample from this distribution. Set $T_n = \bar{X}$, the sample average. Then T_n has the standard Cauchy distribution. Let F_n be the EDF of the sample. A bootstrap analog of T_n is $T_n{*} = \bar{X}{*} - m_n$, where $\bar{X}{*}$ is the average of a bootstrap sample that is drawn randomly with replacement from the data $\{X_i\}$ and m_n is a median or trimmed mean of the data. The asymptotic normality condition of Theorem 2.2 is not satisfied, and the bootstrap estimator of the distribution of T_n is inconsistent. Athreya (1987) and Hall (1990) provide further discussion of the behavior of the bootstrap with heavy-tailed distributions. ∎

Example 2.3 (The distribution of the square of the sample average): Let $\{X_i\colon\ i = 1, \ldots, n\}$ be a random sample from a distribution with mean m and variance s^2. Let $\bar{X}$ denote the sample average. Let F_n be the EDF of the sample. Set $T_n = n^{1/2}(\bar{X}^2 - m^2)$ if $m \neq 0$ and $T_n = n\bar{X}^2$ otherwise. T_n is asymptotically normally distributed if $m \neq 0$, but T_n / s^2 is asymptotically chi-square distributed with one degree of freedom if $m = 0$. The bootstrap analog of T_n is $T_n{*} = n^a[(\bar{X}{*})^2 - \bar{X}^2]$, where $a = 1/2$ if $m \neq 0$ and $a = 1$ otherwise. The bootstrap estimator of $G_n(t, F_0) = P(T_n \leq t)$ is $G_n(t, F_n) = P_n{*}(T_n{*} \leq t)$. If $m \neq 0$, then T_n is asymptotically equivalent to a normalized sample average that satisfies the asymptotic normality condition of Theorem 2.2. Therefore, $G_n(\cdot, F_n)$ consistently estimates $G_\infty(\cdot, F_0)$ if $m \neq 0$. If $m = 0$, then T_n is not a sample average even asymptotically, so Theorem 2.2 does not apply. Condition (iii) of Theorem 2.1 is not satisfied, however, if $m = 0$, and it can be shown that the bootstrap distribution $G_n(\cdot,F_n)$ does not consistently estimate $G_\infty(\cdot, F_0)$ (Datta 1995). ∎

The following example is due to Bickel and Freedman (1981).

Example 2.4: (Distribution of the maximum of a sample): Let $\{X_i: i = 1, \ldots, n\}$ be a random sample from a distribution with absolutely continuous CDF F_0 and support $[0, q_0]$. Let $q_n = max(X_1, \ldots, X_n)$, and define $T_n = n(q_n - q_0)$. Let F_n be the EDF of the sample. The bootstrap analog of T_n is $T_n{}^* = n(q_n{}^* - q_n)$, where $q_n{}^*$ is the maximum of the bootstrap sample $\{X_i{}^*\}$ that is obtained by sampling $\{X_i\}$ randomly with replacement. The bootstrap does not consistently estimate $G_n(-t, F_0) = P_n(T_n \leq -t)$ $(t \geq 0)$. To see why, observe that $P_n{}^*(T_n{}^* = 0) = 1 - (1 - 1/n)^n \rightarrow 1 - e^{-1}$ as $n \rightarrow \infty$. It is easily shown, however, that the asymptotic distribution of T_n is $G_\infty(-t, F_0) = 1 - \exp[-tf(q_0)]$, where $f(x) = dF(x)/dx$ is the probability density function of X. Therefore, $P(T_n = 0) \rightarrow 0$, and the bootstrap estimator of $G_n(\cdot, F_0)$ is inconsistent. ∎

Example 2.5 (Parameter on a boundary of the parameter space): The bootstrap does not consistently estimate the distribution of a parameter estimator when the true parameter point is on the boundary of the parameter space. To illustrate, consider estimation of the population mean m subject to the constraint $m \geq 0$. Estimate m by $m_n = \bar{X} I(\bar{X} > 0)$, where $\bar{X}$ is the average of the random sample $\{X_i: i = 1, \ldots, n\}$. Set $T_n = n^{1/2}(m_n - m)$. Let F_n be the EDF of the sample. The bootstrap analog of T_n is $T_n{}^* = n^{1/2}(m_n{}^* - m_n)$, where $m_n{}^*$ is the estimator of m that is obtained from a bootstrap sample. The bootstrap sample is obtained by sampling $\{X_i\}$ randomly with replacement. If $m > 0$ and $Var(X) < \infty$, then T_n is asymptotically equivalent to a normalized sample average and is asymptotically normally distributed. Therefore, it follows from Theorem 2.2 that the bootstrap consistently estimates the distribution of T_n. If $m = 0$, then the asymptotic distribution of T_n is censored normal, and it can be proved that the bootstrap distribution $G_n(\cdot, F_n)$ does not estimate $G_n(\cdot, F_0)$ consistently (Andrews 1997b). ∎

The next section describes resampling methods that often are consistent when the bootstrap is not. They provide consistent estimators of $G_n(\cdot, F_0)$ in each of the foregoing examples.

2.2 *Alternative Resampling Procedures*

This section describes two resampling methods whose requirements for consistency are weaker than those of the bootstrap. Each is based on drawing subsamples of size $m < n$ from the original data. In one method, the subsamples are drawn randomly with replacement. In the other, the subsamples are drawn without replacement. These subsampling methods often estimate $G_n(\cdot, F_0)$ consistently even when the bootstrap does not. They are not perfect substitutes for the bootstrap, however, because they tend to be less accurate than the bootstrap when the bootstrap is consistent.

In the first subsampling method, which will be called *replacement subsampling*, a bootstrap sample is obtained by drawing $m < n$ observations from the estimation sample $\{X_i: i = 1, \ldots, n\}$. In other respects, it is identical to the standard bootstrap based on sampling F_n. Thus, the replacement subsampling estimator of $G_n(\cdot, F_0)$ is $G_m(\cdot, F_n)$. Swanepoel (1986) gives conditions under which the replacement bootstrap consistently estimates the distribution of T_n in Example 2.4 (the distribution of the maximum of a sample). Andrews (1997b) gives conditions under which it consistently estimates the distribution of T_n in Example 2.5 (parameter on the boundary of the parameter space). Bickel, *et al*. (1997) provide a detailed discussion of the consistency and rates of convergence of replacement bootstrap estimators. To obtain some intuition into why replacement subsampling works, let F_{mn} be the EDF of a sample of size n drawn from the empirical distribution of the estimation data. Observe that if $m \to \infty$, $n \to \infty$, and $m/n \to 0$, then the random sampling error of F_n as an estimator of F_0 is smaller than the random sampling error of F_{mn} as an estimator of F_n. This makes the subsampling method less sensitive than the bootstrap to the behavior of $G_n(\cdot, F)$ for F's in a neighborhood of F_0 and, therefore, less sensitive to violations of continuity conditions such as condition (iii) of Theorem 2.1.

The method of subsampling without replacement will be called *non-replacement subsampling*. This method has been investigated in detail by Politis and Romano (1994), who show that it consistently estimates the distribution of a statistic under very weak conditions. In particular, the conditions required for consistency of the non-replacement subsampling estimator are much weaker than those required for consistency of the bootstrap estimator. Politis *et al*. (1997) extend the subsampling method to heteroskedastic time series.

To describe the non-replacement subsampling method, let $t_n = t_n(X_1, \ldots, X_n)$ be an estimator of the population parameter q, and set $T_n = r(n)(t_n - q)$, where the normalizing factor $r(n)$ is chosen so that $G_n(t, F_0) = P(T_n \leq t)$ converges to a nondegenerate limit $G_\infty(t, F_0)$ at continuity points of the latter. In example 2.1 (estimating the distribution of the sample average), for instance, q is the population mean, $t_n = \bar{X}$, and $r(n) = n^{1/2}$. Let $\{ X_{i_j} : j = 1,\ldots,m \}$ be a subset of $m < n$ observations taken from the sample $\{X_i: i = 1, \ldots, n\}$. Define $N_{mn} = \binom{n}{m}$ to be the total number of subsets that can be formed. Let $t_{m,k}$ denote the estimator t_m evaluated at the k'th subset. The non-replacement subsampling method estimates $G_n(t, F_0)$ by

$$(2.1) \qquad G_{nm}(t) \equiv \frac{1}{N_{nm}} \sum_{k=1}^{N_{nm}} I[r(m)(t_{m,k} - t_n) \leq t].$$

The intuition behind this method is as follows. Each subsample $\{ X_{i_j} \}$ is a random sample of size m from the population distribution whose CDF is F_0. Therefore, $G_m(\cdot, F_0)$ is the exact sampling distribution of $r(m)(t_m - q)$, and

$$(2.2) \qquad G_m(t, F_0) = E\{I[r(m)(t_m - q) \leq t]\}.$$

The quantity on the right-hand side of (2.2) cannot be calculated in an application because F_0 and q are unknown. Equation (2.1) is the estimator of the right-hand side of (2.2) that is obtained by replacing the population expectation by the average over subsamples and q by t_n. If n is large but m/n is small, then random fluctuations in t_n are small relative to those in t_m. Accordingly, the sampling distributions of $r(m)(t_m - t_n)$ and $r(m)(t_m - q)$ are close. Similarly, if N_{mn} is large, the average over subsamples is a good approximation to the population average. These ideas are formalized in the following theorem of Politis and Romano (1994).

Theorem 2.3: *Assume that* $G_n(t, F_0) \to G_\infty(t, F_0)$ *as* $n \to \infty$ *at each continuity point of the latter function. Also assume that* $r(m)/r(n) \to 0$, $m \to \infty$, *and* $m/n \to 0$ *as* $n \to \infty$. *Let* t *be a continuity point of* $G_\infty(t, F_0)$. *Then*: (i) $G_{nm}(t) \to^p G_\infty(t, F_0)$; (ii) *if* $G_\infty(\cdot, F_0)$ *is continuous, then*

$$\sup_t \left| G_{nm}(t) - G_\infty(t, F_0) \right| \to^p 0;$$

(iii) *let* $c_n(1 - a) = \inf\{t : G_{nm}(t) \geq 1 - a\}$ *and* $c(1 - a, F_0) = \inf\{t : G_\infty(t, F_0) \geq 1 - a\}$. *If* $G_\infty(\cdot, F_0)$ *is continuous at* $c(1 - a, F_0)$, *then* $P[r(n)(t_n - q) \leq c_n(1-a)] \to 1 - a$, *and the asymptotic coverage probability of the confidence interval* $[t_n - r(n)^{-1}c_n(1 - a), \infty)$ *is* $1 - a$.

Essentially, this theorem states that if T_n has a well-behaved asymptotic distribution, then the non-replacement subsampling method consistently estimates this distribution. The non-replacement subsampling method also consistently estimates asymptotic critical values for T_n and asymptotic confidence intervals for t_n.

In practice, N_{nm} is likely to be very large, which makes G_{nm} hard to compute. This problem can be overcome by replacing the average over all N_{nm} subsamples with the average over a random sample of subsamples (Politis and Romano 1994). These can be obtained by sampling the data $\{X_i: i = 1, \ldots, n\}$ randomly without replacement.

It is not difficult to show that the conditions of Theorem 2.3 are satisfied in all of the statistics considered in Examples 2.1, 2.2, 2.4, and 2.5. The conditions are also satisfied by the statistic considered in Example 2.3 if the normalization constant is known. Bertail *et al*. (1995) describe a subsampling method for estimating the normalization constant $r(n)$ when it is unknown and provide Monte Carlo evidence on the numerical performance of the non-replacement subsampling method with an estimated normalization constant. In each of the foregoing examples, the replacement subsampling method works because the subsamples are random samples of the true population distribution of X, rather than an estimator of the population distribution. Therefore, replacement subsampling, in contrast to the bootstrap, does not require assumptions such as condition (iii) of Theorem 2.1 that restrict the behavior of $G_n(\cdot, F)$ for F's in a neighborhood of F_0.

The non-replacement subsampling method enables the asymptotic distributions of statistics to be estimated consistently under very weak conditions. However, the standard bootstrap is typically more accurate than non-replacement subsampling when the former is consistent. Suppose that $G_n(\cdot, F_0)$ has an Edgeworth expansion through $O(n^{-1/2})$, as is the case with the distributions of most asymptotically normal statistics encountered in applied econometrics. Then, as will be discussed in Section 3, $|G_n(t, F_n) - G_n(t, F_0)|$, the error made by the bootstrap estimator of $G_n(t, F_0)$, is at most $O(n^{-1/2})$ almost surely. In contrast, the error made by the non-replacement subsampling estimator, $|G_{nm}(t) - G_n(t, F_0)|$, is no smaller than $O_p(n^{-1/3})$ (Politis and Romano 1994).[3] Thus, the standard bootstrap estimator of $G_n(t, F_0)$ is more accurate than the non-replacement subsampling estimator in a setting that arises frequently in applications. Similar results can be obtained for statistics that are asymptotically chi-square distributed. Thus, the standard bootstrap is more attractive than the non-replacement subsampling method in most applications in econometrics. The subsampling method may be used, however, if characteristics of the sampled population or the statistic of interest cause the standard bootstrap estimator to be inconsistent. Non-replacement subsampling may also be useful in situations where checking the consistency of the bootstrap is difficult. Examples of this include inference about the parameters of certain kinds of structural search models (Flinn and Heckman 1982), auction models (Donald and Paarsch 1996), and binary-response models that are estimated by Manski's (1975, 1985) maximum score method.

3. ASYMPTOTIC REFINEMENTS

The previous section described conditions under which the bootstrap yields a consistent estimator of the distribution of a statistic. Roughly speaking, this means that the bootstrap gets the statistic's asymptotic distribution right, at least if the sample size is sufficiently large. As was discussed in Section 1, however, the bootstrap often does much more than get the asymptotic distribution right. In a large number of situations that are important in applied econometrics, it provides a higher-order asymptotic approximation to the distribution of a statistic. This section explains how the bootstrap can be used to obtain asymptotic refinements. Section 3.1 describes the use of the bootstrap to reduce the finite-sample bias of an estimator. Section 3.2 explains how the bootstrap obtains higher-order approximations to the distributions of statistics. The results of Section 3.2 are used in Sections 3.3 and 3.4 to show how the bootstrap obtains higher-order refinements to the rejection probabilities of tests and the coverage probabilities of confidence intervals. Sections 3.5-3.7 address additional issues associated with the use of the bootstrap to obtain asymptotic refinements. It is assumed throughout this section that the data are a simple random sample from some distribution. Methods for implementing the bootstrap and obtaining asymptotic refinements with time-series data are discussed in Section 4.1.

3.1 *Bias Reduction*

This section explains how the bootstrap can be used to reduce the finite-sample bias of an estimator. The theoretical results are illustrated with a simple numerical example. To minimize the complexity of the discussion, it is assumed that the inferential problem is to obtain a point estimate of a scalar parameter q that can be expressed as a smooth function of a vector of population moments. It is also assumed that q can be estimated consistently by substituting sample moments in place of population moments in the smooth function. Many important econometric estimators, including maximum-likelihood and generalized-method-of-moments estimators, are either functions of sample moments or can be approximated by functions of sample moments with an approximation error that approaches zero very rapidly as the sample size increases. Thus, the theory outlined in this section applies to a wide variety of estimators that are important in applications.

To be specific, let X be a random vector, and set $m = E(X)$. Assume that the true value of q is $q_0 = g(m)$, where g is a known, continuous function. Suppose that the data consist of a random sample $\{X_i: \ i = 1, ..., n\}$ of X. Define the vector $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Then q is estimated consistently by

(3.1) $$q_n = g(\bar{X}).$$

If q_n has a finite mean, then $E(q_n) = E[g(\bar{X})]$. However, $E[g(\bar{X})] \neq g(m)$ in general unless g is a linear function. Therefore, $E(q_n) \neq q_0$, and q_n is a biased estimator of q. In particular, $E(q_n) \neq q_0$ if q_n is any of a variety of familiar maximum likelihood or generalized method of moments estimators.

To see how the bootstrap can reduce the bias of q_n, suppose that g is four times continuously differentiable in a neighborhood of m and that the components of X have finite fourth absolute moments. Let G_1 denote the vector of first derivatives of g and G_2 denote the matrix of second derivatives. A Taylor series expansion of the right-hand side of (3.1) about $\bar{X} = m$ gives

(3.2) $$q_n - q_0 = G_1(m)'(\bar{X} - m) + \frac{1}{2}(\bar{X} - m)'G_2(m)(\bar{X} - m) + R_n,$$

where R_n is a remainder term that satisfies $E(R_n) = O(n^{-2})$. Therefore, taking expectations on both sides of (3.2) gives

(3.3) $$E(q_n - q_0) = \frac{1}{2}E[(\bar{X} - m)'G_2(m)(\bar{X} - m)] + O(n^{-2}).$$

The first term on the right-hand side of (3.3) has size $O(n^{-1})$. Therefore, through $O(n^{-1})$ the bias of q_n is

(3.4) $$B_n = \frac{1}{2}E[(\overline{X} - m)'G_2(m)(\overline{X} - m)]\,.$$

Now consider the bootstrap. The bootstrap samples the empirical distribution of the data. Let $\{X_i^*: i = 1,\ldots,n\}$ be a bootstrap sample that is obtained this way. Define $\overline{X}^* = n^{-1}\sum_{i=1}^{n} X_i^*$ to be the vector of bootstrap sample means. The bootstrap estimator of q is $q_n^* = g(\overline{X}^*)$. Conditional on the data, the true mean of the distribution sampled by the bootstrap is $\overline{X}$. Therefore, $\overline{X}$ is the bootstrap analog of m, and $q_n = g(\overline{X})$ is the bootstrap analog of q_0. The bootstrap analog of (3.2) is

(3.5) $$q_n^* - q_n = G_1(\overline{X})'(\overline{X}^* - \overline{X}) + \frac{1}{2}(\overline{X}^* - \overline{X})'G_2(\overline{X})(\overline{X}^* - \overline{X}) + R_n^*\,,$$

where R_n^* is the bootstrap remainder term. Let E^* denote the expectation under bootstrap sampling, that is, the expectation relative to the empirical distribution of the estimation data. Let $B_n^* \equiv E^*(q_n^* - q_n)$ denote the bias of q_n^* as an estimator of q_n. Taking E^* expectations on both sides of (3.5) shows that

(3.6) $$B_n^* = \frac{1}{2}E^*[(\overline{X}^* - \overline{X})'G_2(\overline{X})(\overline{X}^* - \overline{X})] + O(n^{-2})$$

almost surely. Because the distribution that the bootstrap samples is known, B_n^* can be computed with arbitrary accuracy by Monte Carlo simulation. Thus, B_n^* is a feasible estimator of the bias of q_n. The details of the simulation procedure are described below.

By comparing (3.4) and (3.6), it can be seen that the only differences between B_n and the leading term of B_n^* are that $\overline{X}$ replaces m in B_n^* and the empirical expectation, E^*, replaces the population expectation, E. Moreover, $E(B_n^*) = B_n + O(n^{-2})$ Therefore, through $O(n^{-1})$, use of the bootstrap bias estimate B_n^* provides the same bias reduction that would be obtained if the infeasible population value B_n could be used. This is the source of the bootstrap's ability to reduce the bias of q_n. The resulting bias-corrected estimator of q is $q_n - B_n^*$. It satisfies $E(q_n - B_n^*) = O(n^{-2})$. Thus, the bias of the bias-corrected estimator is $O(n^{-2})$, whereas the bias of the uncorrected estimator q_n is $O(n^{-1})$. [4]

The Monte Carlo procedure for computing B_n^* is as follows:

Monte Carlo Procedure for Bootstrap Bias Estimation

B1. Use the estimation data to compute q_n.

B2. Generate a bootstrap sample of size n by sampling the estimation data randomly with replacement. Compute $q_n^* = g(\bar{X}^*)$.

B3. Compute $E^*q_n^*$ by averaging the results of many repetitions of step B2. Set $B_n^* = E^*q_n^* - q_n$.

To implement this procedure it is necessary to choose the number of repetitions, m, of step B2. It usually suffices to choose m sufficiently large that the estimate of $E^*q_n^*$ does not change significantly if m is increased further. Andrews and Buchinsky (1997) discuss more formal methods for choosing the number of bootstrap replications.[5]

The following simple numerical example illustrates the bootstrap's ability to reduce bias. Examples that are more realistic but also more complicated are presented in Horowitz (1998a).

Example 3.1 (Horowitz 1998a): Let $X \sim N(0, 6)$ and $n = 10$. Let $g(m) = \exp(m)$. Then $q_0 = 1$, and $q_n = \exp(\bar{X})$. B_n and the bias of $q_n - B_n^*$ can be found through the following Monte Carlo procedure:

MC1. Generate an estimation data set of size n by sampling from the $N(0,6)$ distribution. Use this data set to compute q_n.

MC2. Compute B_n^* by carrying out steps B1-B3. Form $q_n - B_n^*$.

MC3. Estimate $E(q_n - q_0)$ and $E(q_n - B_n^* - q_0)$ by averaging the results of many repetitions of steps MC1-MC2. Estimate the mean square errors of q_n and $q_n - B_n^*$ by averaging the realizations of $(q_n - q_0)^2$ and $(q_n - B_n^* - q_0)^2$.

The following are the results obtained with 1000 Monte Carlo replications and 100 repetitions of step B2 at each Monte Carlo replication:

	Bias	Mean-Square Error
q_n	0.356	1.994
$q_n - B_n^*$	-0.063	1.246

In this example, the bootstrap reduces the magnitude of the bias of the estimator of q by nearly a factor of 6. The mean-square estimation error is reduced by 38 percent. ∎

3.2 *The Distributions of Statistics*

This section explains why the bootstrap provides an improved approximation to the finite-sample distribution of an asymptotically pivotal statistic. As before, the data are a random sample $\{X_i: i = 1, \ldots, n\}$ from a probability distribution whose CDF is F_0. Let $T_n = T_n(X_1, ..., X_n)$ be a statistic. Let $G_n(t, F_0) = P(T_n \le t)$ denote the exact, finite-sample CDF of T_n. As was discussed in Section 2, $G_n(t, F_0)$ cannot be calculated analytically unless T_n is pivotal. The objective of this section is to obtain an approximation to $G_n(t, F_0)$ that is applicable when T_n is not pivotal.

To obtain useful approximations to $G_n(t, F_0)$, it is necessary to make certain assumptions about the form of the function $T_n(X_1, \ldots, X_n)$. It is assumed in this section that T_n is a smooth function of sample moments of X or sample moments of functions of X (the smooth function model). Specifically, $T_n = n^{1/2}[H(\bar{Z}_1,...,\bar{Z}_J) - H(m_{Z_1},...,m_{Z_J})]$, where the scalar-valued function H is smooth in a sense that is defined precisely below, $\bar{Z}_j = n^{-1}\sum_{i=1}^{n} Z_j(X_i)$ for each $j = 1, \ldots, J$ and some nonstochastic function Z_j, and $m_{Z_j} = E(Z_j)$. After centering and normalization, most estimators and test statistics used in applied econometrics are either smooth functions of sample moments or can be approximated by such functions with an approximation error that is asymptotically negligible.[6] The ordinary least squares estimator of the slope coefficients in a linear mean-regression model and the t statistic for testing a hypothesis about a coefficient are exact functions of sample moments. Maximum-likelihood and generalized-method-of-moments estimators of the parameters of nonlinear models can be approximated with asymptotically negligible error by smooth functions of sample moments if the log-likelihood function or moment conditions have sufficiently many derivatives with respect to the unknown parameters.

Some important econometric estimators and test statistics do not satisfy the assumptions of the smooth function model. Quantile estimators, such as the least-absolute-deviations (LAD) estimator of the slope coefficients of a median-regression model do not satisfy the assumptions of the smooth function model because their objective functions are not sufficiently smooth. Nonparametric density and mean-regression estimators and semiparametric estimators that require kernel or other forms of smoothing also do not fit within the smooth function model. Bootstrap methods for such estimators are discussed in Section 4.3.

Now return to the problem of approximating $G_n(t, F_0)$. First-order asymptotic theory provides one approximation. To obtain this approximation, write $H(\bar{Z}_1,\ldots,\bar{Z}_J) = H(\bar{Z})$, where $\bar{Z}$ = $(\bar{Z}_1,\ldots,\bar{Z}_J)'$. Define $m_Z = E(\bar{Z})$, $\partial H(z) = \partial H(z)/\partial z$, and $\Omega = E[(\bar{Z} - m_Z)(\bar{Z} - m_Z)']$ whenever these quantities exist. Assume that:

SFM: (i) $T_n = n^{1/2}[H(\bar{Z}) - H(m_Z)]$, *where* $H(z)$ *is* 6 *times continuously partially differentiable with respect to any mixture of components of* z *in a neighborhood of* m_Z. (ii) $\partial H(m_Z) \neq 0$. (iii) *The expected value of the product of any* 16 *components of* Z *exists*.[7]

Under assumption SFM, a Taylor series approximation gives

(3.7) $$n^{1/2}[H(\bar{Z}) - H(m_Z)] = \partial H(m_Z)' n^{1/2}(\bar{Z} - m_Z) + o_p(1) .$$

Application of the Lindeberg-Levy central limit theorem to the right hand side of (3.7) shows that $n^{1/2}[H(\bar{Z}) - H(m_Z)] \to^d N(0,V)$, where $V = \partial H(m_Z)'\Omega\partial H(m_Z)$. Thus, the asymptotic CDF of T_n is $G_\infty(t, F_0) = \Phi(t / V^{1/2})$, where Φ is the standard normal CDF. This is just the usual result of the delta method. Moreover, it follows from the Berry-Esséen theorem that

$$\sup_t |G_n(t, F_0) - G_\infty(t, F_0)| = O(n^{-1/2}) .$$

Thus, under assumption SFM of the smooth function model, first-order asymptotic approximations to the exact finite-sample distribution of T_n make an error of size $O(n^{-1/2})$.[8]

Now consider the bootstrap. The bootstrap approximation to the CDF T_n is $G_n(\cdot, F_n)$. Under the smooth function model with assumption SFM, it follows from Theorem 3.2 that the bootstrap is consistent. Indeed, it is possible to prove the stronger result that $\sup_t |G_n(t, F_n) - G_\infty(t, F_0)| \to 0$ almost surely. This result insures that the bootstrap provides a good approximation to the asymptotic distribution of T_n if n is sufficiently large. It says nothing, however, about the accuracy of $G_n(\cdot, F_n)$ as an approximation to the exact finite-sample distribution function $G_n(\cdot, F_0)$. To investigate this question, it is necessary to develop higher-order asymptotic approximations to $G_n(\cdot, F_0)$ and $G_n(\cdot, F_n)$. The following theorem, which is proved in Hall (1992a), provides an essential result.

Theorem 3.1: *Let assumption SFM hold. Assume also that*

(3.8) $$\lim_{\|t\|\to\infty} \sup |E[\exp(it'Z)]| < 1 ,$$

where $i = \sqrt{-1}$. *Then*

(3.9) $$G_n(t, F_0) = G_\infty(t, F_0) + \frac{1}{n^{1/2}} g_1(t, F_0) + \frac{1}{n} g_2(t, F_0) + \frac{1}{n^{3/2}} g_3(t, F_0) + O(n^{-2})$$

uniformly over t *and*

(3.10) $$G_n(t, F_n) = G_\infty(t, F_n) + \frac{1}{n^{1/2}} g_1(t, F_n) + \frac{1}{n} g_2(t, F_n) + \frac{1}{n^{3/2}} g_3(t, F_n) + O(n^{-2})$$

uniformly over t almost surely. Moreover, g_1 and g_3 are even, differentiable functions of their first arguments, g_2 is an odd, differentiable, function of its first argument, and G_∞, g_1, g_2, and g_3 are continuous functions of their second arguments relative to the supremum norm on the space of distribution functions.

If T_n is asymptotically pivotal, then G_∞ is the standard normal distribution function. Otherwise, $G_\infty(\cdot, F_0)$ is the $N(0,V)$ distribution function, and $G_\infty(\cdot, F_n)$ is the $N(0,V_n)$ distribution function, where V_n is the quantity obtained from V by replacing population expectations and moments with expectations and moments relative to F_n.

Condition (3.8) is called the *Cramér condition*. It is satisfied if the random vector Z has a probability density with respect to Lebesgue measure.[9]

It is now possible to evaluate the accuracy of the bootstrap estimator $G_n(t, F_n)$ as an approximation to the exact, finite-sample CDF $G_n(t, F_0)$. It follows from (3.9) and (3.10) that

$$(3.11)\qquad G_n(t, F_n) - G_n(t, F_0) = [G_\infty(t, F_n) - G_\infty(t, F_0)] + \frac{1}{n^{1/2}}[g_1(t, F_n) - g_1(t, F_0)]$$

$$+\frac{1}{n}[g_2(t, F_n) - g_2(t, F_0)] + O(n^{-3/2})$$

almost surely uniformly over t. The leading term on the right-hand side of (3.11) is $[G_\infty(t, F_n) - G_\infty(t, F_0)]$. The size of this term is $O(n^{-1/2})$ almost surely uniformly over t because $F_n - F_0 = O(n^{-1/2})$ almost surely uniformly over the support of F_0. Thus, the bootstrap makes an error of size $O(n^{-1/2})$ almost surely, which is the same as the size of the error made by first-order asymptotic approximations. In terms of rate of convergence to zero of the approximation error, the bootstrap has the same accuracy as first-order asymptotic approximations. In this sense, nothing is lost in terms of accuracy by using the bootstrap instead of first-order approximations, but nothing is gained either.

Now suppose that T_n is *asymptotically pivotal*. Then the asymptotic distribution of T_n is independent of F_0, and $G_\infty(t, F_n) = G_\infty(t,, F_0)$ for all t. Equations (3.9) and (3.10) now yield

$$(3.12)\qquad G_n(t,F_n)-G_n(t,F_0)=\frac{1}{n^{1/2}}[g_1(t,F_n)-g_1(t,F_0)]$$

$$+\frac{1}{n}[g_2(t,F_n)-g_2(t,F_0)]+O(n^{-3/2})$$

almost surely. The leading term on the right-hand side of (3.12) is $n^{-1/2}[g_1(t, F_n) - g_1(t, F_0)]$. It follows from continuity of g_1 with respect to its second argument that this term has size $O(n^{-1})$ almost surely uniformly over t. Now the bootstrap makes an error of size $O(n^{-1})$, which is smaller as $n \to \infty$ than the error made by first-order asymptotic approximations. Thus, the bootstrap is more accurate than first-order asymptotic theory for estimating the distribution of a smooth asymptotically pivotal statistic.

If T_n is asymptotically pivotal, then the accuracy of the bootstrap is even greater for estimating the symmetrical distribution function $P(|T_n| \le t) = G_n(t, F_0) - G_n(-t, F_0)$. This quantity is important for obtaining the RP's of symmetrical tests and the coverage probabilities of symmetrical confidence intervals. Let Φ denote the standard normal distribution function. Then, it follows from (3.9) and the symmetry of g_1, g_2, and g_3 in their first arguments that

$$G_n(t,F_0)-G_n(-t,F_0)=[G_\infty(t,F_0)-G_\infty(-t,F_0)]+\frac{2}{n}g_2(t,F_0)+O(n^{-2})$$

$$(3.13)\qquad =2\Phi(t)-1+\frac{2}{n}g_2(t,F_0)+O(n^{-2}).$$

Similarly, it follows from (3.10) that

$$G_n(t,F_n)-G_n(-t,F_n)=[G_\infty(t,F_n)-G_\infty(-t,F_n)]+\frac{2}{n}g_2(t,F_n)+O(n^{-2})$$

$$(3.14)\qquad =2\Phi(t)-1+\frac{2}{n}g_2(t,F_n)+O(n^{-2})$$

almost surely. The remainder terms in (3.13) and (3.14) are $O(n^{-2})$ and not $O(n^{-3/2})$ because the $O(n^{-3/2})$ term of an Edgeworth expansion, $n^{-3/2}g_3(t,F)$, is an even function that, like g_1, cancels out in the subtractions used to obtain (3.13) and (3.14) from (3.9) and (3.10). Now subtract (3.13) from (3.14) and use the fact that F_n - F_0 = $O(n^{-1/2})$ almost surely to obtain

(3.15) $$[G_n(t,F_n) - G_n(-t,F_n)] - [G_n(t,F_0) - G_n(-t,F_0)]$$

$$= \frac{2}{n}[g_2(t,F_n) - g_2(t,F_0)] + O(n^{-2})$$

$$= O(n^{-3/2})$$

almost surely if T_n is asymptotically pivotal. Thus, the error made by the bootstrap approximation to the symmetrical distribution function $P(|T_n| \le t)$ is $O(n^{-3/2})$ compared to the error of $O(n^{-1})$ made by first-order asymptotic approximations.

In summary, when T_n is asymptotically pivotal, the error of the bootstrap approximation to a one-sided distribution function is

(3.16) $$G_n(t,F_n) - G_n(t,F_0) = O(n^{-1})$$

almost surely uniformly over t. The error in the bootstrap approximation to a symmetrical distribution function is

(3.17) $$[G_n(t,F_n) - G_n(-t,F_n)] - [G_n(t,F_0) - G_n(-t,F_0)] = O(n^{-3/2})$$

almost surely uniformly over t. In contrast, the errors made by first-order asymptotic approximations are $O(n^{-1/2})$ and $O(n^{-1})$, respectively, for one-sided and symmetrical distribution functions. Equations (3.16) and (3.17) provide the basis for the bootstrap's ability to reduce the finite-sample errors in the RP's of tests and the coverage probabilities of confidence intervals. Section 3.3 discusses the use of the bootstrap in hypothesis testing. Confidence intervals are discussed in Section 3.4.

3.3 *Bootstrap Critical Values for Hypothesis Tests*

This section shows how the bootstrap can be used to reduce the errors in the RP's of hypothesis tests relative to the errors made by first-order asymptotic approximations.

Let T_n be a statistic for testing a hypothesis H_0 about the sampled population. Assume that under H_0, T_n is asymptotically pivotal and satisfies assumptions SFM and (3.8). Consider a symmetrical, two-tailed test of H_0. This test rejects H_0 at the α level if $|T_n| > z_{n,\alpha/2}$, where $z_{n,\alpha/2}$, the exact, finite-sample, α-level critical value, is the 1 - α/2 quantile of the distribution of T_n.[10] The critical value solves the equation

$$(3.18) \qquad G_n(z_{n,\alpha/2}, F_0) - G_n(-z_{n,\alpha/2}, F_0) = 1 - \alpha \ .$$

Unless T_n is exactly pivotal, however, equation (3.18) cannot be solved in an application because F_0 is unknown. Therefore, the exact, finite-sample critical value cannot be obtained in an application if T_n is not pivotal.

First-order asymptotic approximations obtain a feasible version of (3.18) by replacing G_n with G_∞. Thus, the asymptotic critical value, $z_{\infty,\alpha/2}$, solves

$$(3.19) \qquad G_\infty(z_{\infty,\alpha/2}, F_0) - G_\infty(-z_{\infty,\alpha/2}, F_0) = 1 - \alpha \ .$$

Since G_∞ is the standard normal distribution when T_n is asymptotically pivotal, $z_{\infty,\alpha/2}$ can be obtained from tables of standard normal quantiles. Combining (3.13), (3.18), and (3.19) gives

$$[G_\infty(z_{n,\alpha/2}, F_0) - G_\infty(-z_{n,\alpha/2}, F_0)] - [G_\infty(z_{\infty,\alpha/2}, F_0) - G_\infty(-z_{\infty,\alpha/2}, F_0)] = O(n^{-1}) \ ,$$

which implies that $z_{n,\alpha/2} - z_{\infty,\alpha/2} = O(n^{-1})$. Thus, the asymptotic critical value approximates the exact finite sample critical value with an error whose size is $O(n^{-1})$.

The bootstrap obtains a feasible version of (3.18) by replacing F_0 with F_n. Thus, the bootstrap critical value, $z_{n,\alpha/2}^*$, solves

$$(3.20) \qquad G_n(z_{n,\alpha/2}^*, F_n) - G_n(-z_{n,\alpha/2}^*, F_n) = 1 - \alpha \ .$$ [11]

Equation (3.20) usually cannot be solved analytically, but $z_{n,\alpha/2}^*$ can be estimated with any desired accuracy by Monte Carlo simulation. To illustrate, suppose, as often happens in applications, that T_n is an asymptotically normal, Studentized estimator of a parameter θ whose value under H_0 is θ_0. That is,

$$T_n = \frac{n^{1/2}(q_n - q_0)}{s_n},$$

where q_n is the estimator of q, $n^{1/2}(q_n - q_0) \to^d N(0, s^2)$ under H_0 and s_n^2 is a consistent estimator of s^2. Then the Monte Carlo procedure for evaluating $z_{n,a/2}*$ is as follows:

Monte Carlo Procedure for Computing the Bootstrap Critical Value

T1. Use the estimation data to compute q_n.

T2. Generate a bootstrap sample of size n by sampling the distribution corresponding to F_n. For example, if F_n is the EDF of the data, then the bootstrap sample can be obtained by sampling the data randomly with replacement. If F_n is parametric so that $F_n(\cdot) = F(\cdot, q_n)$ for some function F, then the bootstrap sample can be generated by sampling the distribution whose CDF is $F(\cdot, q_n)$. Compute the estimators of q and s from the bootstrap sample. Call the results q_n* and s_n*. The bootstrap version of T_n is $T_n* = n^{1/2}(q_n* - q_n) / s_n*$.

T3. Use the results of many repetitions of T2 to compute the empirical distribution of $|T_n*|$. Set $z_{n,a/2}*$ equal to the 1 - a quantile of this distribution.

The foregoing procedure does not specify the number of bootstrap replications that should be carried out in step T3. In practice, it often suffices to choose a value sufficiently large that further increases have no important effect on $z_{n,a/2}*$. Hall (1986a) and Andrews and Buchinsky (1997) describe the results of formal investigations of the problem of choosing the number of bootstrap replications. Repeatedly estimating q in step T2 can be computationally burdensome if q_n is an extremum estimator. Davidson and MacKinnon (1997a) and Andrews (1998) show that the computational burden can be reduced by replacing the extremum estimator with an estimator that is obtained by taking a small number of Newton or quasi-Newton steps from the q_n value obtained in step T1.

To evaluate the accuracy of the bootstrap critical value $z_{n,a/2}*$ as an estimator of the exact finite-sample critical value $z_{n,a/2}$, combine (3.13) and (3.18) to obtain

$$2\Phi(z_{n,a/2}) - 1 + \frac{2}{n} g_2(z_{n,a/2}, F_0) = 1 - a + O(n^{-2}) . \tag{3.21}$$

Similarly, combining (3.14) and (3.20) yields,

$$2\Phi(z_{n,a/2}*) - 1 + \frac{2}{n} g_2(z_{n,a/2}*, F_n) = 1 - a + O(n^{-2}) \tag{3.22}$$

almost surely. Equations (3.21) and (3.22) can be solved to yield Cornish-Fisher expansions for $z_{n,a/2}$ and $z_{n,a/2}*$. The results are (Hall 1992a, p. 111)

$$z_{n,a/2} = z_{\infty,a/2} - \frac{1}{n} \frac{g_2(z_{\infty,a/2}, F_0)}{f(z_{\infty,a/2})} + O(n^{-2}) , \tag{3.23}$$

where f is the standard normal density function, and

$$z_{n,a/2}* = z_{\infty,a/2} - \frac{1}{n} \frac{g_2(z_{\infty,a/2}, F_n)}{f(z_{\infty,a/2})} + O(n^{-2}) \tag{3.24}$$

almost surely. It follows from (3.23) and (3.24) that

$$z_{n,a/2}* = z_{n,a/2} + O(n^{-3/2}) \tag{3.25}$$

almost surely. Thus, the bootstrap critical value for a symmetrical, two-tailed test differs from the exact, finite-sample critical value by $O(n^{-3/2})$ almost surely. The bootstrap critical value is more accurate than the asymptotic critical value, $z_{\infty,a/2}$, whose error is $O(n^{-1})$.

Now consider the rejection probability of the test based on T_n when H_0 is true. With the exact but infeasible a-level critical value, the RP is $P(|T_n| > z_{n,a/2}) = a$.. With the asymptotic critical value, the RP is

$$\begin{aligned} P(|T_n| > z_{\infty,a/2}) &= 1 - [G_n(z_{\infty,a/2}, F_0) - G_n(-z_{\infty,a/2}, F_0)] \\ &= a + O(n^{-1}), \qquad (3.26) \end{aligned}$$

where the last line follows from setting $t = z_{\infty,a/2}$ in (3.13). Thus, with the asymptotic critical value, the true and nominal RP's differ by $O(n^{-1})$.

Now consider the RP with the bootstrap critical value, $P(|T_n| \geq z_{n,a/2}^*)$. Because $z_{n,a/2}^*$ is a random variable, $P(|T_n| \geq z_{n,a/2}^*) \neq 1 - [G_n(z_{n,a/2}^*, F_0) - G_n(-z_{n,a/2}^*, F_0)]$. This fact complicates the calculation of the difference between the true and nominal RP's with the bootstrap critical value. The calculation is outlined in the Appendix of this chapter. The result is that

(3.27) $$P(|T_n| > z_{n,a/2}^*) = a + O(n^{-2}) .$$

In other words, the nominal RP of a symmetrical, two-tailed test with a bootstrap critical value differs from the true RP by $O(n^{-2})$ when the test statistic is asymptotically pivotal. In contrast, the difference between the nominal and true RP's is $O(n^{-1})$ when the asymptotic critical value is used.

The bootstrap does not achieve the same accuracy for one-tailed tests. For such tests, the difference between the nominal and true RP's with a bootstrap critical value is usually $O(n^{-1})$, whereas the difference with asymptotic critical values is $O(n^{-1/2})$. See Hall (1992a, pp. 102-103) for details. There are, however, circumstances in which the difference between the nominal and true RP's with a bootstrap critical value is $O(n^{-3/2})$. Hall (1992a, pp. 178-179) shows that this is true for a one-sided t test of a hypothesis about a slope (but not intercept) coefficient in a homoskedastic, linear, mean-regression model. Davidson and MacKinnon (1997b) show that it is true whenever T_n is asymptotically independent of $g_2(z_{\infty,a/2},F_n)$. They further show that many familiar test statistics satisfy this condition.

Tests based on statistics that are asymptotically chi-square distributed behave like symmetrical, two-tailed tests. Therefore, the differences between their nominal and true RP's under H_0 are $O(n^{-1})$ with asymptotic critical values and $O(n^{-2})$ with bootstrap critical values.

Singh (1981), who considered a one-tailed test of a hypothesis about a population mean, apparently was the first to show that the bootstrap provides a higher-order asymptotic approximation to the distribution of an asymptotically pivotal statistic. Singh's test was based on the standardized sample mean. Early papers giving results on higher-order approximations for Studentized means and for more general hypotheses and test statistics include Babu and Singh (1983, 1984), Beran (1988) and Hall (1986b, 1988).

3.4 *Confidence Intervals*

Let q be a population parameter whose true but unknown value is q_0. Let q_n be a $n^{1/2}$-consistent, asymptotically normal estimator of q, and let s_n be a consistent estimator of the standard deviation of the asymptotic distribution of $n^{1/2}(q_n - q_0)$. Then an asymptotic $1 - \alpha$ confidence interval for q_0 is $q_n - z_{\infty,a/2}s_n \leq q_0 \leq q_n + z_{\infty,a/2}s_n$. Define $T_n = n^{1/2}(q_n - q_0)/s_n$. Then the coverage probability of the asymptotic confidence interval is $P(|T_n| \leq z_{\infty,a/2})$. It follows from (3.26) that the difference between the true coverage probability of the interval and the nominal coverage probability, $1 - a$, is $O(n^{-1})$.

If T_n satisfies the assumptions of Theorem 3.1, then the difference between the nominal and true coverage probabilities of the confidence interval can be reduced by replacing the asymptotic critical value with the bootstrap critical value $z_{n,a/2}*$. With the bootstrap critical value, the confidence interval is $q_n - z_{n,a/2}*s_n \leq q_0 \leq q_n + z_{n,a/2}*s_n$. The coverage probability of this interval is $P(|T_n| \leq z_{n,a/2}*)$. By (3.27), $P(|T_n| \leq z_{n,a/2}*) = 1 - a + O(n^{-2})$, so the true and nominal coverage probabilities differ by $O(n^{-2})$ when the bootstrap critical value is used, whereas they differ by $O(n^{-1})$ when the asymptotic critical value is used.

Analogous results can be obtained for one-sided and equal-tailed confidence intervals. With asymptotic critical values, the true and nominal coverage probabilities of these intervals differ by $O(n^{-1/2})$. With bootstrap critical values, the differences are $O(n^{-1})$. In special cases such as the slope coefficients of homoskedastic, linear, mean-regressions, the differences with bootstrap critical values are $O(n^{-3/2})$.

The bootstrap's ability to reduce the differences between the true and nominal coverage probabilities of a confidence interval is illustrated by the following example, which is an extension of Example 3.1.

Example 3.2 (Horowitz 1998a): This example uses Monte Carlo simulation to compare the true coverage probabilities of asymptotic and bootstrap nominal 95% confidence intervals for q_0 in the model of Example 3.1. The Monte Carlo procedure is:

MC4: Generate an estimation data set of size $n = 10$ by sampling from the $N(0,6)$ distribution. Use this data set to compute q_n.

MC5: Compute $z_{n,\alpha/2}*$ by carrying out steps T2-T3 of Section 3.3. Determine whether q_0 is contained in the confidence intervals based on the asymptotic and bootstrap critical values.

MC6: Determine the empirical coverage probabilities of the asymptotic and bootstrap confidence intervals from the results of 1000 repetitions of steps MC4-MC5.

The empirical coverage probability of the asymptotic confidence interval was 0.886 in this experiment, whereas the empirical coverage probability of the bootstrap interval was 0.943. The asymptotic coverage probability is statistically significantly different from the nominal probability of 0.95 ($p < 0.01$), whereas the bootstrap coverage probability is not ($p > 0.10$). ∎

3.5 *The Importance of Asymptotically Pivotal Statistics*

The arguments in Sections 3.2-3.4 show that the bootstrap provides higher-order asymptotic approximations to distributions, RP's of tests, and coverage probabilities of confidence intervals based on smooth, asymptotically pivotal statistics. These include test statistics whose asymptotic distributions are standard normal or chi-square and, thus, most statistics that are used for testing hypotheses about the parameters of econometric models. Models that satisfy the required smoothness conditions include linear and nonlinear mean-regression models, error-components mean-regression models for panel data, logit and probit models that have at least one continuously distributed explanatory variable, and tobit models. The smoothness conditions are also satisfied by parametric sample-selection models in which the selection equation is a logit or probit model with at least one continuously distributed explanatory variable. Asymptotically pivotal statistics based on median-regression models do not satisfy the smoothness conditions. Bootstrap methods for such statistics are discussed in Section 4.3. The ability of the bootstrap to provide asymptotic refinements for smooth, asymptotically pivotal statistics provides a powerful argument for using them in applications of the bootstrap.

The bootstrap may also be applied to statistics that are not asymptotically pivotal, but it does not provide higher-order approximations to their distributions. Estimators of the structural parameters of econometric models (e.g., slope and intercept parameters, including regression coefficients; standard errors, covariance matrix elements, and autoregressive coefficients) usually are not asymptotically pivotal. The asymptotic distributions of centered structural parameter estimators are often normal with means of zero but have variances that depend on the unknown population distribution of the data. The errors of bootstrap estimates of the distributions of statistics that are not asymptotically pivotal converge to zero at the same rate as the errors made by first-order asymptotic approximations.[12]

Higher-order approximations to the distributions of statistics that are not asymptotically pivotal can be obtained through the use of bootstrap iteration (Beran 1987, 1988; Hall 1992a) or bias-correction methods (Efron 1987). Bias correction methods are not applicable to symmetrical tests and confidence intervals. Bootstrap iteration is discussed in Section 4.4. Bootstrap iteration is highly computationally intensive, which makes it unattractive when an asymptotically pivotal statistic is available.

3.6 *The Parametric Versus the Nonparametric Bootstrap*

The size of the error in the bootstrap estimate of a RP or coverage probability is determined by the size of $F_n - F_0$. Thus, F_n should be the most efficient available estimator. If F_0 belongs to a known parametric family $F(\cdot, q)$, $F(\cdot, q_n)$ should be used to generate bootstrap samples, rather than the EDF. Although the bootstrap provides asymptotic refinements regardless of whether $F(\cdot, q_n)$ or the EDF is used, the results of Monte Carlo experiments have shown that the numerical accuracy of the bootstrap tends to be much higher with $F(\cdot, q_n)$ than with the EDF. If the objective is to test a hypothesis H_0 about q, further gains in efficiency and performance can be obtained by imposing the constraints of H_0 when obtaining the estimate q_n.

To illustrate, consider testing the hypothesis H_0: $b_1 = 0$ in the Box-Cox regression model

(3.28) $$Y^{(l)} = b_0 + b_1 X + U ,$$

where $Y^{(l)}$ is the Box-Cox (1964) transformation of Y, X is an observed, scalar explanatory variable, U is an unobserved random variable, and b_0 and b_1 are parameters. Suppose that $U \sim N(0, s^2)$.[13] Then bootstrap sampling can be carried out in the following ways:

1. Sample (Y, X) pairs from the data randomly with replacement.

2. Estimate l, b_0, and b_1 in (3.28) by maximum likelihood, and obtain residuals $\hat{U}$. Generate Y values from $Y = [l_n(b_0 + b_1 X + U^*) + 1]^{1/l_n}$, where l_n, b_0, and b_1 are the estimates of l, b_0, and b_1; and U^* is sampled randomly with replacement from the $\hat{U}$.

3. Same as method 2 except U^* is sampled randomly from the distribution $N(0, s_n^2)$, where s_n^2 is the maximum likelihood estimate of s^2.

4. Estimate l, b_0, and s^2 in (3.28) by maximum likelihood subject to the constraint $b_1 = 0$. Then proceed as in method 2.

5. Estimate l, b_0, and s^2 in (3.28) by maximum likelihood subject to the constraint $b_1 = 0$. Then proceed as in method 3.

In methods 2-5, the values of X may be fixed in repeated samples or sampled independently of $\hat{U}$ from the empirical distribution of X.

Method 1 provides the least efficient estimator of F_n and typically has the poorest numerical accuracy. Method 5 has the greatest numerical accuracy. Method 3 will usually have greater numerical accuracy than method 2. If the distribution of U is not assumed to belong to a known parametric family, then methods 3 and 5 are not available, and method 4 will usually have greater numerical accuracy than methods 1-2. Of course, parametric maximum likelihood cannot be used to estimate b_0, b_1, and l if the distribution of U is not specified parametrically.

If the objective is to obtain a confidence interval for b_1 rather than to test a hypothesis, methods 4 and 5 are not available. Method 3 will usually provide the greatest numerical accuracy if the distribution of U is assumed to belong to a known parametric family, and method 2 if not.

One reason for the relatively poor performance of method 1 is that it does not impose the condition $E(U \mid X = x) = 0$. This problem is discussed further in Section 5.2, where heteroskedastic regression models are considered.

3.7 *Recentering*

The bootstrap provides asymptotic refinements for asymptotically pivotal statistics because, under the assumptions of the smooth function model, $\sup_t |G_n(t, F_n) - G_n(t, F_0)|$ converges to zero as $n \to \infty$ more rapidly than $\sup_t |G_\infty(t, F_0) - G_n(t, F_0)|$. One important situation in which this does not necessarily happen is generalized method of moments (GMM) estimation of an overidentified parameter when F_n is the EDF of the sample.

To see why, let q_0 be the true value of a parameter q that is identified by the moment condition $Eh(X, q) = 0$. Assume that $\dim(h) > \dim(q)$. If, as is often the case in applications, the distribution of X is not assumed to belong to a known parametric family, the EDF of X is the most obvious candidate for F_n. The sample analog of $Eh(X, q)$ is then

$$E^* h(X, q) = \frac{1}{n}\sum_{i=1}^{n} h(X_i, q),$$

where E^* denotes the expectation relative to F_n. The sample analog of q_0 is q_n, the GMM estimator of q. In general, $E^*h(X, q_n) \neq 0$ in an overidentified model, so bootstrap estimation based on the

EDF of X implements a moment condition that does not hold in the population the bootstrap samples. As a result, the bootstrap estimator of the distribution of the statistic for testing the overidentifying restrictions is inconsistent (Brown *et al*. 1997). The bootstrap does consistently estimate the distributions of $n^{1/2}(q_n - q_0)$ (Hahn 1996) and the t statistic for testing a hypothesis about a component of q. However, it does not provide asymptotic refinements for the RP of the t test or the coverage probability of a confidence interval.

This problem can be solved by basing bootstrap estimation on the recentered moment condition $E^*h^*(X, q_n) = 0$, where

$$(3.29) \qquad h^*(X,q) = h(X,q) - \frac{1}{n}\sum_{i=1}^{n} h(X_i, q_n) \, .$$

Hall and Horowitz (1996) show that the bootstrap with recentering provides asymptotic refinements for the RP's of t tests of hypotheses about components of q and the test of overidentifying restrictions. The bootstrap with recentering also provides asymptotic refinements for confidence intervals. Intuitively, the recentering procedure works by replacing the misspecified moment condition $E^*h(X, q) = 0$ with the condition $E^*h^*(X, q) = 0$, which does hold in the population that the bootstrap samples.

Freedman (1981) recognized the need for recentering residuals in regression models without intercepts. See, also, Efron (1979).

Brown *et al*. (1997) propose an alternative approach to recentering. Instead of replacing h with h^* for bootstrap estimation, they replace the empirical distribution of X with an empirical likelihood estimator that is constructed so that $E^*h(X, q_n) = 0$.[14] The empirical likelihood estimator assigns a probability mass p_{ni} to observation X_i ($i = 1, \ldots, n$). The p_{ni}'s are determined by solving the problem

$$\underset{p_{n1},\ldots,p_{nn}}{\text{maximize:}} \quad \sum_{i=1}^{n} \log p_{ni}$$

$$\text{subject to:} \quad \sum_{i=1}^{n} p_{ni} h(X_i, q_n) = 0; \quad \sum_{i=1}^{n} p_{ni} = 1; \quad p_{ni} \geq 0$$

In general, the solution to this problem yields $p_{ni} \neq n^{-1}$, so the empirical likelihood estimator of the distribution of X is not the same as the empirical distribution. Brown *et al*. (1997) implement the bootstrap by sampling $\{X_i\}$ with probability weights p_{ni} instead of randomly with replacement. They argue that the bootstrap is more accurate with empirical-likelihood recentering than with recentering by (3.29) because the empirical-likelihood estimator of the distribution of X is asymptotically efficient under the moment conditions $Eh(X,q) = 0$. With either method of recentering, however, the differences between the nominal and true RP's of symmetrical tests and between the nominal and true coverage probabilities of symmetrical confidence intervals are $O(n^{-2})$. Thus, the differences between the errors made with the two recentering methods are likely to be small with samples of the sizes typically encountered in applications.

Brown *et al*. (1997) develop the empirical-likelihood recentering method only for simple random samples. Kitamura (1997) has shown how to carry out empirical-likelihood estimation with dependent data. It is likely, therefore, that empirical-likelihood recentering can be extended to GMM estimation with dependent data. The recentering method based on (3.29) requires no modification for use with dependent data (Hall and Horowitz 1996). Section 4.1 provides further discussion of the use of the bootstrap with dependent data.

4. EXTENSIONS

This section explains how the bootstrap can be used to obtain asymptotic refinements in certain situations where the assumptions of Section 3 are not satisfied. Section 4.1 treats dependent data. Section 4.2 treats kernel density and nonparametric mean-regression estimators. Section 4.3 shows how the bootstrap can be applied to certain non-smooth estimators. Section 4.4 describes how bootstrap iteration can be used to obtain asymptotic refinements without an asymptotically pivotal statistic. Section 4.5 discusses additional special problems that can arise in implementing the bootstrap. Section 4.6 discusses the properties of bootstrap critical values for testing a hypothesis that is false.

4.1 *Dependent Data*

With dependent data, asymptotic refinements cannot be obtained by using independent bootstrap samples. Bootstrap sampling must be carried out in a way that suitably captures the dependence of the data-generation process. This section describes several methods for doing this. It also explains how the bootstrap can be used to obtain asymptotic refinements in GMM estimation with dependent data. At present, higher-order asymptotic approximations and asymptotic refinements are available only when the data-generation process is stationary and strongly geometrically mixing. Except when stated otherwise, it is assumed here that this requirement is satisfied. Non-stationary data-generation processes are discussed in Section 4.1.3.

4.1.1 *Methods for Bootstrap Sampling with Dependent Data*

Bootstrap sampling that captures the dependence of the data can be carried out relatively easily if there is a parametric model, such as an ARMA model, that reduces the data-generation process to a transformation of independent random variables. For example, suppose that the series $\{X_t\}$ is generated by the stationary, invertible, finite-order ARMA model

(4.1) $$A(L,\alpha)X_t = B(L,\beta)U_t$$

where A and B are known functions, L is the backshift operator, α and β are vectors of parameters, and $\{U_t\}$ is a sequence of independently and identically distributed (*iid*) random variables. Let α_n and β_n be $n^{1/2}$-consistent, asymptotically normal estimators of α and β, and let $\{\hat{U}_t\}$ be the centered residuals of the estimated model (4.1). Then a bootstrap sample $\{X_t^*\}$ can be generated as

$$A(L,\alpha_n)X_t^* = B(L,\beta_n)U_t^*,$$

where $\{U_t^*\}$ is a random sample from the empirical distribution of the residuals $\{\hat{U}_t\}$. If the distribution of U_t is assumed to belong to a known parametric family (e.g., the normal distribution), then $\{U_t^*\}$ can be generated by independent sampling from the estimated parametric distribution. Bose (1988) provides a rigorous discussion of the use of the bootstrap with autoregressions. Bose (1990) treats moving average models.

When there is no parametric model that reduces the data-generation process to independent sampling from some probability distribution, the bootstrap can be implemented by dividing the data into blocks and sampling the blocks randomly with replacement. The block bootstrap is important in GMM estimation with dependent data, because the moment conditions on which GMM estimation is based usually do not specify the dependence structure of the GMM residuals. The blocks may be non-overlapping (Carlstein 1986) or overlapping (Hall 1985, Künsch 1988, Politis and Romano 1994). To describe these blocking methods more precisely, let the data consist of observations $\{X_i: \ i = 1, \ldots, n\}$. With non-overlapping blocks of length l, block 1 is observations $\{X_j: \ j = 1, \ldots, l\}$, block 2 is observations $\{X_{l+j}: \ j = 1, \ldots, l\}$, and so forth. With overlapping blocks of length l, block 1 is observations $\{X_j: \ j = 1, \ldots, l\}$, block 2 is observations $\{X_{j+1}: \ j = 1, \ldots, l\}$, 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 1994). 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 fixed (non-random) 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. The asymptotically optimal block length is defined as the one that minimizes the asymptotic mean-square error of the block bootstrap estimator. The asymptotically optimal block length and its rate of increase with increasing n depend on what is being estimated. Hall *et al.* (1995) showed that with either overlapping or non-overlapping blocks with non-random lengths, the asymptotically optimal block-length is $l \sim n^r$, where r = 1/3 for estimating bias or variance, r = 1/4 for estimating a one-sided distribution function (e.g., $P(T_n \leq t)$), and r = 1/5 for estimating a two-sided distribution function (e.g., $P(|T_n| \leq t)$). Hall *et al*. (1995) also show that overlapping blocks provide somewhat higher estimation efficiency than non-overlapping ones. The efficiency difference is likely to be very small in applications, however. For estimating a two-sided distribution function, for example, the root-mean-square estimation error (RMSE) with either blocking method is $O(n^{-6/5})$. The numerical difference between the RMSE's can be illustrated by considering the case of a normalized sample average. Let $T_n = (\overline{X} - m)/s$, where $\overline{X}$ is the sample average of observations $\{X_i\}$, $m = E(\overline{X})$, and $s^2 = Var(\overline{X})$. Then the results of Hall, *et al*. (1995) imply that for estimating $P(|T_n| \leq t)$, the reduction in asymptotic RMSE from using overlapping blocks instead of nonoverlapping ones is less than 10 percent.

Lahiri (1997) has investigated the asymptotic efficiency of the stationary bootstrap. He showed that the asymptotic relative efficiency of the stationary bootstrap compared to the block bootstrap with non-random block lengths is always less than one and can be arbitrarily close to zero. More precisely, let $RMSE_{SB}$ and $RMSE_{NR}$, respectively, denote the asymptotic RMSE's of the stationary bootstrap and the block bootstrap with overlapping or non-overlapping blocks with non-random lengths. Then $RMSE_{NR}/\ RMSE_{SB} < 1$ always and can be arbitrarily close to zero. Thus, at least in terms of asymptotic RMSE, the stationary bootstrap is unattractive relative to the block bootstrap with fixed-length blocks.

Implementation of the block bootstrap in an application requires a method for choosing the block length with a finite sample. Hall, *et al.* (1995) describe a subsampling method for doing this when the block lengths are non-random. The idea of the method is to use subsamples to create an empirical analog of the mean-square error of the bootstrap estimator of the quantity of interest. Let

y denote this quantity (e.g., a two-sided distribution function). Let y_n be the bootstrap estimator of y that is obtained using a preliminary block-length estimate. Let $m < n$. Let $y_{m,i}(l')$ ($i = 1, \ldots, n - m$) denote the bootstrap estimates of y that are computed using all the $n - m$ runs of length m in the data and block length l'. Let l_m be the value of l' that minimizes $\sum_i [y_{m,i}(l') - y_n]^2$. The estimator of the asymptotically optimal block length is $(n/m)^r l_m$, where $r = 1/3$ for estimating bias or variance, $r = 1/4$ for estimating a one-sided distribution function, and $r = 1/5$ for estimating a two-sided distribution function

Bühlmann (1997) has proposed an alternative to blocking for use when the data-generation process can be represented as an infinite-order autoregression. In this method, called the *sieve bootstrap*, the infinite-order autoregression is replaced by an approximating autoregression with a finite-order that increases at a suitable rate as $n \to \infty$. The coefficients of the finite-order autoregression are estimated, and the bootstrap is implemented by sampling the centered residuals from the estimated finite-order model. Bühlmann (1997) gives conditions under which this procedure yields consistent estimators of variances and distribution functions. Bühlmann (1998) shows that the sieve bootstrap provides an asymptotic refinement for estimating the CDF of the t statistic for testing a one-sided hypothesis about the trend function in an AR(∞) process with a deterministic trend.

4.1.2 *Asymptotic Refinements in GMM Estimation with Dependent Data*

This section discusses the use of the block bootstrap to obtain asymptotic refinements in GMM estimation with dependent data. Lahiri (1992) showed that the block bootstrap provides asymptotic refinements through $O(n^{-1/2})$ for normalized sample moments and for a Studentized sample moment with m-dependent data. Hall and Horowitz (1996) showed that the block bootstrap provides asymptotic refinements through $O(n^{-1})$ for symmetrical tests and confidence intervals based on GMM estimators. Their methods can also be used to show that the bootstrap provides refinements through $O(n^{-1/2})$ for one-sided tests and confidence intervals. Hall and Horowitz (1996) do not assume that the data-generation process is *m*-dependent[15].

Regardless of whether overlapping or nonoverlapping blocks are used, block bootstrap sampling does not exactly replicate the dependence structure of the original data-generation process. For example, if nonoverlapping blocks are used, bootstrap observations that belong to the same block are deterministically related, whereas observations that belong to different blocks are independent. This dependence structure is unlikely to be present in the original data-generation process. As a result, the finite-sample covariance matrices of the asymptotic forms of parameter estimators obtained from the original sample and from the bootstrap sample are different. The practical consequence of this difference is that asymptotic refinements through $O(n^{-1})$ cannot be obtained by applying the "usual" formulae for test statistics to the block-bootstrap sample. It is necessary to develop special formulae for the bootstrap versions of test statistics. These formulae contain factors that correct for the differences between the asymptotic covariances of the original-sample and bootstrap versions of test statistics without distorting the higher-order terms of asymptotic expansions that produce refinements.

Lahiri (1992) derived the bootstrap version of a Studentized sample mean for *m*-dependent data. Hall and Horowitz (1996) derived formulae for the bootstrap versions of the GMM symmetrical, two-tailed *t* statistic and the statistic for testing overidentifying restrictions. As an illustration of the form of the bootstrap statistics, consider the GMM *t* statistic for testing a hypothesis about a component of a parameter q that is identified by the moment condition $Eh(X, q) = 0$. Hall and Horowitz (1996) showed that the corrected formula for the bootstrap version of the GMM *t* statistic is

$$T_n^* = (S_n / S_b)\tilde{T}_n ,$$

where $\tilde{T}_n$ is the "usual" GMM *t* statistic applied to the bootstrap sample, S_n is the "usual" GMM standard error of the estimate of the component of q that is being tested, and S_b is the exact standard deviation of the asymptotic form of the bootstrap estimate of this component. S_n is computed from the original estimation sample, not the bootstrap sample. Hansen (1982) gives formulae for the usual GMM t statistic and standard error. S_b can be calculated because the process generating bootstrap data is known exactly. An analogous formula is available for the bootstrap version of the statistic for testing overidentifying restrictions but is much more complicated algebraically than the formula for the t statistic. See Hall and Horowitz (1996) for details.

At present, the block bootstrap is known to provide asymptotic refinements for symmetrical tests and confidence intervals based on GMM estimators only if the residuals $\{h(X_i, q_0): \ i = 1,2,...\}$ at the true parameter point, q_0, are uncorrelated after finitely many lags. That is,

(4.2) $$E[h(X_i, q_0)h(X_j, q_0)'] = 0 \quad \text{if } |i - j| > M$$

for some $M < \infty$.[16] This restriction is not equivalent to m-dependence because it does not preclude correlations among higher powers of components of h that persist at arbitrarily large lags (e.g., stochastic volatility). Although the restriction is satisfied in many econometric applications (see, e.g., Hansen 1982, Hansen and Singleton 1982), there are others in which relaxing it would be useful. The main problem in doing so is that without (4.2), it is necessary to use a kernel-type estimator of the GMM covariance matrix (see, e.g., Newey and West 1987, 1994; Andrews 1991, Andrews and Monahan 1992). Kernel-type estimators are not functions of sample moments and converge at rates that are slower than $n^{-1/2}$. However, present results on the existence of asymptotic expansions that achieve $O(n^{-1})$ accuracy with dependent data apply only to functions of sample moments that have $n^{-1/2}$ rates of convergence (Götze and Hipp 1983, 1994). It will be necessary to extend existing theory of asymptotic expansions with dependent data before (4.2) can be relaxed for symmetrical tests and confidence intervals.

Condition (4.2) is not needed for one-sided tests and confidence intervals, where the bootstrap provides only $O(n^{-1/2})$ refinements. Götze and Künsch (1996) and Lahiri (1996) give conditions under which the moving-block-bootstrap approximation to the distribution of a statistic that is Studentized with a kernel-type variance estimator is accurate through $O_p(n^{-1/2})$. When the conditions are satisfied,

$$(4.3) \qquad \sup_t |P(T_n \le t) - P^*(T_n^* \le t)| = o_p(n^{-1/2}),$$

where T_n^* is the bootstrap analog of the Studentized statistic T_n, and the moving block bootstrap is used to generate bootstrap samples. In Götze and Künsch (1996), T_n is the Studentized form of a smooth function of sample moments. In Lahiri (1996), T_n is a Studentized statistic for testing a hypothesis about a slope coefficient in a linear mean-regression model. Achieving the result (4.3) requires, among other things, use of a suitable kernel or weight function in the variance estimator. Götze and Künsch (1996) show that (4.3) holds with a rectangular or quadratic kernel but not with a triangular one.

4.1.3 *The Bootstrap with Non-Stationary Processes*

The foregoing results assume that the data-generation process is stationary. Most research to date on using the bootstrap with non-stationary data has been concerned with establishing consistency of bootstrap estimators of distribution functions, not with obtaining asymptotic refinements. An exception is Lahiri (1992), who gives conditions under which the bootstrap estimator of the distribution of the normalized sample average of non-stationary data differs from the true distribution by $o(n^{-1/2})$ almost surely. Thus, under Lahiri's conditions, the bootstrap is more accurate than first-order asymptotic approximations. Lahiri's result requires *a priori* knowledge of the covariance function of the data and does not apply to Studentized sample averages. Moreover Lahiri assumes the existence of the covariance function, so his result does not apply to unit-root processes.

The consistency of the bootstrap estimator of the distribution of the slope coefficient or Studentized slope coefficient in a simple unit-root model has been investigated by Basawa *et al.* (1991a, 1991b), Datta (1996), and Ferretti and Romo (1996). The model is

(4.4) $X_i = bX_{i-1} + U_i; \quad i = 1,2,\ldots,n$,

where $X_0 = 0$ and $\{U_i\}$ is an *iid* sequence with $E(U_i) = 0$ and $E(U_i^2) = s^2 < \infty$. Let b_n denote the ordinary least squares estimator of b in (4.4):

$$b_n = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_{i-1}^2}. \tag{4.5}$$

Let b_0 denote the true but unknown value of b. Consider using the bootstrap to estimate the sampling distribution of (b_n - b_0) or the t statistic for testing H$_0$: $b = b_0$. It turns out that when b_0 = 1 is possible, the consistency of the bootstrap estimator is much more sensitive to how the bootstrap sample is drawn than when it is known that $|b_0| < 1$.

Basawa *et al*. (1991a) investigate the consistency of a bootstrap estimator of the distribution of the t statistic in the special case that $U \sim N(0,1)$. In this case, the t statistic is

$$t_n = \left[\sum_{i=1}^{n} X_{i-1}^2\right]^{1/2} (b_n - b_0). \tag{4.6}$$

In Basawa *et al*. (1991a), the bootstrap sample $\{X_i^*: \ i = 1, \ldots, n\}$ is generated recursively from the estimated model

(4.7) $X_i^* = b_n X_{i-1}^* + U_i^*$,

where $X_0^* = 0$ and $\{U_i^*\}$ is an independent random sample from the $N(0,1)$ distribution. The bootstrap version of the t statistic is

$$t^* = \left[\sum_{i=1}^{n} (X_{i-1}^*)^2\right]^{1/2} (b_n^* - b_n),$$

where b_n^* is obtained by replacing X_i with X_i^* in (4.5). Basawa *et al*. (1991a) show that the bootstrap distribution function $P_n^*(t^* \le t)$ does not consistently estimate the population distribution function $P_n(t \le t)$. This result is not surprising. The asymptotic distribution of t is discontinuous at

b_0 = 1. Therefore, condition (iii) of Theorem 2.1 is not satisfied if the set of data-generation processes under consideration includes ones with and without $b_0 = 1$.

This problem can be overcome by specifying that $b_0 = 1$, thereby removing the source of the discontinuity. Basawa *et al*. (1991b) investigate the consistency of the bootstrap estimator of the distribution of the statistic $Z_n \equiv n(b_n - 1)$ for testing the unit-root hypothesis H$_0$: $b_0 = 1$ in (4.4). The bootstrap sample is generated by the recursion

(4.8) $$X_i^* = X_{i-1}^* + U_i^* ,$$

where $X_0^* = 0$ and $\{U_i^*\}$ is a random sample from the centered residuals of (4.4) under H$_0$. The centered residuals are $\hat{U}_i = X_i - X_{i-1} - \overline{U}$, where $\overline{U} = n^{-1}\sum_{i=1}^{n}(X_i - X_{i-1})$. The bootstrap analog of Z_n is $Z_n^* = n(b_n^* - 1)$, where b_n^* is obtained by replacing X_i with X_i^* in (4.5). Basawa *et al*. (1991b) show that if H$_0$ is true, then $|P_n^*(Z_n^* \leq z) - P_n(Z_n \leq z)| = o_p(1)$ uniformly over z.

The discontinuity problem can be overcome without the restriction $b_0 = 1$ by using bootstrap samples consisting of $m < n$ observations (Datta 1996). This approach has the advantage of yielding a confidence interval for b_0 that is valid for any $b_0 \in (-\infty, \infty)$. Consider model (4.4) with the additional assumption that $E|U_i|^{2+d} < \infty$ for some $d > 0$. Let b_n be the ordinary least squares estimator of b, and define t_n as in (4.6). Let $\hat{U}_i = X_i - b_n X_{i-1} - n^{-1}\sum_{i=1}^{n}(X_i - b_n X_{i-1})$ ($i = 1, \ldots, n$) denote the centered residuals from the estimated model, and let $\{U_i^*: i = 1, \ldots, m\}$ be a random sample of $\{\hat{U}_i\}$ for some $m < n$. The bootstrap sample is generated by the recursion (4.7) but with $i = 1, \ldots, m$ instead of $i = 1, \ldots, n$. Let b_m^* denote the ordinary least squares estimator of b that is obtained from the bootstrap sample. Define the bootstrap version of t_n by

$$t_m^* = \left[\sum_{i=1}^{m}(X_{i-1}^*)^2\right]^{1/2}(b_m^* - b_n) .$$

Datta (1996) proves that if $[m(\log\log n)^2]/n \to 0$ as $n \to \infty$, then $|P_m^*(t_m^* \leq t) - P_n(t_n \leq t)| = o(1)$ almost surely as $n \to \infty$ uniformly over z for any $b_0 \in (-\infty, \infty)$.

Ferretti and Romo consider a test of H_0: $b_0 = 1$ in (4.4). Let b_n be the ordinary least squares estimator of b, and let

(4.9) $$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - b_n X_{i-1})^2 .$$

The test statistic is

(4.10) $$\tilde{t}_n = \frac{1}{s_n}\left(\sum_{i=1}^{n} X_{i-1}^2\right)^{1/2}(b_n - 1) .$$

The bootstrap sample is generated from the centered residuals of the estimated model by using the recursion (4.8). Let b_n* denote the ordinary least squares estimator of b that is obtained from the bootstrap sample. The bootstrap version of the test statistic, $\tilde{t}_n*$, is obtained by replacing X_i and b_n with X_i* and b_n* in (4.9) and (4.10). Ferretti and Romo (1996) show that $|P_n*(\tilde{t}_n* \leq t) - P_n(\tilde{t}_n \leq t)| = o(1)$ almost surely as $n \to \infty$. Ferretti and Romo (1996) also show how this result can be extended to the case in which $\{U_i\}$ in (4.4) follows an AR(1) process.

The results of Monte Carlo experiments (Li and Maddala 1996, 1997) suggest that the differences between the true and nominal RP's of tests of hypotheses about integrated or cointegrated data-generation processes are smaller with bootstrap-based critical values than with asymptotic ones. At present, however, there are no theoretical results on the ability of the bootstrap to provide asymptotic refinements for tests or confidence intervals when the data are integrated or cointegrated.

4.2 *Kernel Density and Regression Estimators*

This section describes the use of the bootstrap to carry out inference about kernel nonparametric density and mean-regression estimators. These are not smooth functions of sample moments, even approximately, so the results of Section 3 do not apply to them. In particular, kernel density and mean-regression estimators converge more slowly than $n^{-1/2}$, and their distributions have unconventional asymptotic expansions that are not in powers of $n^{-1/2}$. Consequently, the sizes of the asymptotic refinements provided by the bootstrap are also not powers of $n^{-1/2}$. Sections 4.2.1-4.2.3 discuss bootstrap methods for nonparametric density estimation. Nonparametric mean regression is discussed in Section 4.2.4.

4.2.1 *Nonparametric Density Estimation*

Let f denote the probability density function (with respect to Lebesgue measure) of the scalar random variable X. The problem addressed in this section is inferring f from a random sample of X, $\{X_i: i = 1, \ldots, n\}$, without assuming that f belongs to a known, finite-dimensional family of functions. Point estimation of f can be carried out by the kernel method. The kernel estimator of $f(x)$ is

$$f_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where K is a *kernel* function with properties that are discussed below and $\{h_n: n = 1, 2, \ldots\}$ is a strictly positive sequence of bandwidths.

The properties of kernel density estimators are described by Silverman (1986), among others. To state the properties that are relevant here, let $r \geq 2$ be an even integer. Assume that f has r bounded, continuous derivatives in a neighborhood of x. Let K be a bounded function that is symmetrical about 0 and has support [-1,1].[17] In addition, let K satisfy

$$\int_{-1}^{1} u^j K(u)du = \begin{cases} 1 \text{ if } j = 0 \\ 0 \text{ if } 1 \leq j \leq r-1 \\ A_K \neq 0 \text{ if } j = r. \end{cases} \tag{4.11}$$

Define

$$B_K = \int_{-1}^{1} K(u)^2 du.$$

Also define $b_n(x) = E[f_n(x) - f(x)]$ and $s_n^2(x) = Var[f_n(x)]$. Then

$$b_n(x) = h_n^r \frac{A_K}{r!} f^{(r)}(x) + o(h_n^r)$$

and

(4.12) $$s_n^2(x) = \frac{B_K}{nh_n} f(x) .$$

Moreover, if nh_n^{2r+1} is bounded as $n \to \infty$, then

$$Z_n(x) \equiv \frac{f_n(x) - f(x) - b_n(x)}{s_n(x)}$$

(4.13) $$= \frac{f_n(x) - E[f_n(x)]}{s_n(x)} \to^d N(0,1).$$

The fastest possible rate of convergence of $f_n(x)$ to $f(x)$ is achieved by setting $h_n \propto n^{-1/(2r+1)}$. When this happens, $f_n(x) - f(x) = O_p[n^{-r/(2r+1)}]$, $b_n(x) \propto n^{-r/(2r+1)}$, and $s_n(x) \propto n^{-r/(2r+1)}$.

A Studentized statistic that is asymptotically pivotal and can be used to test a hypothesis about $f(x)$ or form a confidence interval for $f(x)$ can be obtained from (4.13) if suitable estimators of $s_n^2(x)$ and $b_n(x)$ are available. The need for estimating an asymptotic variance is familiar. An estimator of $s_n^2(x)$ can be formed by replacing $f(x)$ with $f_n(x)$ on the right-hand side of (4.12). However, the asymptotic expansions required to obtain asymptotic refinements are simpler if $s_n^2(x)$ is estimated by a sample analog of the exact, finite-sample variance of $f_n(x)$ instead of a sample analog of (4.12), which is the variance of the asymptotic distribution of $f_n(x)$. A sample analog of the exact finite-sample variance of $f_n(x)$ is given by

$$s_n^2(x) = \frac{1}{(nh_n)^2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)^2 - \frac{f_n(x)^2}{n} .$$

If $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$, then $(nh_n)[s_n^2(x) - s_n^2(x)] = O_p(1)$ as $n \to \infty$. Define the Studentized form of Z_n by

(4.14) $$t_n = \frac{f_n(x) - E[f_n(x)]}{s_n(x)}.$$

Then t_n is the asymptotic t statistic for testing a hypothesis about $E[f_n(x)]$ or forming a confidence interval for $E[f_n(x)]$. The asymptotic distribution of t_n is $N(0,1)$. However, unless the asymptotic bias $b_n(x)$ is negligibly small, t_n cannot be used to test a hypothesis about $f(x)$ or form a confidence interval for $f(x)$. Because $s_n^{-1}(x) = O[(nh_n)^{1/2}]$ and $s_n^{-1}(x) = O_p[(nh_n)^{1/2}]$, $b_n(x)$ is negligibly small only if $(nh_n)^{1/2} b_n(x) = o(1)$ as $n \to \infty$. The problem of asymptotic bias cannot be solved by replacing $E[f_n(x)]$ with $f(x)$ on the right-hand side of (4.14) because the asymptotic distribution of the resulting version of t_n is not centered at 0 unless $b_n(x)$ is negligibly small. Section 4.2.2 discusses ways to deal with asymptotic bias.

4.2.2 *Asymptotic Bias and Methods for Controlling It*

Asymptotic bias is a characteristic of nonparametric estimators that is not shared by estimators that are smooth functions of sample moments. As has just been explained, asymptotic bias may prevent t_n from being suitable for testing a hypothesis about $f(x)$ or constructing a confidence interval for $f(x)$. Asymptotic bias also affects the performance of the bootstrap. To see why, let $\{X_i^*:\ i = 1, \ldots, n\}$ be a bootstrap sample that is obtained by sampling the data $\{X_i\}$ randomly with replacement. Then the bootstrap estimator of f is

(4.14) $$f_n^*(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i^*}{h_n}\right).$$

The bootstrap analog of $s_n^2(x)$ is

$$s_n^{2*}(x) = \frac{1}{(nh_n)^2} \sum_{i=1}^{n} K\left(\frac{x - X_i^*}{h_n}\right)^2 - \frac{f_n^*(x)^2}{n}.$$

Define the bootstrap analog of t_n by

$$t_n^* = \frac{f_n^*(x) - f_n(x)}{s_n^*(x)}.$$

It is clear from (4.14) that $E^*[f_n^*(x) - f_n(x)] = 0$. Thus, $f_n^*(x)$ an unbiased estimator of $f_n(x)$ in a finite sample as well as asymptotically, whereas $f_n(x)$ is an asymptotically biased estimator of $f(x)$. It can be shown that the bootstrap distribution of t_n^* converges in probability to $N(0,1)$. Therefore, despite the unbiasedness of $f_n^*(x)$, t_n^* is a bootstrap t statistic for testing a hypothesis about $E[f_n(x)]$ or forming a confidence interval for $E[f_n(x)]$. It is not a bootstrap t statistic for testing a hypothesis about $f(x)$ or forming a confidence interval for $f(x)$ unless $b_n(x)$ is negligibly small.

There are two ways to overcome the difficulties posed by asymptotic bias so that t_n and t_n^* become statistics for testing hypotheses about $f(x)$ and forming confidence intervals for $f(x)$ instead of $E[f_n(x)]$. One is the method of explicit bias removal. It consists of forming an estimator of $b_n(x)$, say $\hat{b}_n(x)$, that can be subtracted from $f_n(x)$ to form the asymptotically unbiased estimator $f_n(x) - \hat{b}_n(x)$. The other method is undersmoothing. This consists of setting $h_n \propto n^{-k}$ with $k > 1/(2r + 1)$. With undersmoothing, $(nh_n)^{1/2}b_n(x) = o_p(1)$ as $n \to \infty$, so that $b_n(x)$ is asymptotically negligible. Neither method is compatible with achieving the fastest rate of convergence of a point-estimator of $f(x)$. With undersmoothing, the rate of convergence of $f_n(x)$ is that of $s_n(x)$. This is $n^{-(1 - k)/2)}$, which is slower than $n^{-r/(2r + 1)}$. Explicit bias removal with $h_n \propto n^{-1/(2r + 1)}$ and rate of convergence $n^{-r/(2r + 1)}$ for $f_n(x)$ requires $f(x)$ to have more than r derivatives. When $f(x)$ has the required number of derivatives, the fastest possible rate of convergence of $f_n(x)$ is $n^{-s/(2s + 1)}$ for some $s > r$. This rate is achieved with $h_n \propto n^{-1/(2s + 1)}$, but the resulting estimator of $f(x)$ is asymptotically biased. Thus, regardless of the method that is used to remove asymptotic bias, testing a hypothesis about $f(x)$ or forming a confidence interval requires using a bandwidth sequence that converges more rapidly than the one that maximizes the rate of convergence of a point estimator of $f(x)$. Nonparametric point estimation and nonparametric interval estimation or testing of hypotheses are different tasks that require different degrees of smoothing.

Hall (1992b) compares the errors in the coverage probabilities of bootstrap confidence intervals with undersmoothing and explicit bias removal. He shows that when the number of derivatives of $f(x)$ is held constant, undersmoothing achieves a smaller error in coverage probability than does explicit bias removal. This conclusion also applies to the rejection probabilities of hypothesis tests; the difference between true and nominal rejection probabilities can be made smaller with undersmoothing than with explicit bias removal. Thus, undersmoothing is the better method for handling asymptotic bias when the aim is to minimize differences between true and nominal rejection and coverage probabilities of bootstrap-based hypothesis tests and confidence intervals. Accordingly, undersmoothing is used for bias removal in the remainder of this section.

4.2.3 *Asymptotic Refinements*

The argument showing that the bootstrap provides asymptotic refinements for tests of hypotheses and confidence intervals in nonparametric density estimation is similar to that made in Section 3 for the smooth function model. The main step is proving that the distributions of t_n and t_n^* have Edgeworth expansions that are identical up to a sufficiently small remainder. The result is stated in Theorem 4.1, which is proved in Hall (1992a, pp. 268-282).

Theorem 4.1: *Assume that f has r bounded, continuous derivatives in a neighborhood of x. Let $h_n \to 0$ and $(nh_n)/(\log n) \to \infty$ as $n \to \infty$. Let K be a bounded function that is symmetrical about* 0, *has support* [-1,1], *and satisfies* (4.11) *for some $r \geq 2$. Also, assume that there is a partition of* [-1,1], $u_0 = -1 < u_1 < \ldots < u_m = 1$ *such that K' exists, is bounded, and is either strictly positive or strictly negative on each interval (u_j, u_{j+1}). Then there are even functions q_1 and q_3 and an odd function q_2 such that*

$$(4.15)\qquad P(t_n \leq t) = \Phi(t) + \frac{1}{(nh_n)^{1/2}} q_1(t) + \frac{1}{nh_n} q_2(t) + \left(\frac{h_n}{n}\right)^{1/2} q_3(t) + O[(nh_n)^{-3/2} + n^{-1}]$$

uniformly over t. Moreover, there are even functions q_{n1} and q_{n3} and an odd function q_{n2} such that $q_{nj}(t) - q_j(t) \to 0$ as $n \to \infty$ uniformly over t almost surely ($j = 1, \ldots, 3$), *and*

$$P^*(t_n^* \leq t) = \Phi(t) + \frac{1}{(nh_n)^{1/2}} q_{n1}(t) + \frac{1}{nh_n} q_{n2}(t) + \left(\frac{h_n}{n}\right)^{1/2} q_{n3}(t) + O[(nh_n)^{-3/2} + n^{-1}]$$

uniformly over t *almost surely*.

Hall (1992a, pp. 211-216) gives explicit expressions for the functions q_j and q_{nj}.

To see the implications of Theorem 4.1, consider a symmetrical test of a hypothesis about $f(x)$. The results that will be obtained for this test also apply to symmetrical confidence intervals. Let the hypothesis be H_0: $f(x) = f_0$. A symmetrical test rejects H_0 if $|f_n(x) - f_0|$ is large. Suppose that $nh_n^{r+1} \to 0$ as $n \to \infty$. This rate of convergence of h_n insures that the asymptotic bias of $f_n(x)$ has a negligibly small effect on the error made by the higher-order approximation to the distribution of t_n that is used to obtain asymptotic refinements.[18] It also makes the effects of asymptotic bias sufficiently small that t_n can be used to test H_0. Rejecting H_0 if $|f_n(x) - f_0|$ is large is then equivalent to rejecting H_0 if $|t_n|$ is large, thereby yielding a symmetrical t test of H_0.

Now suppose that the critical value of the symmetrical t test is obtained from the asymptotic distribution of t_n, which is $N(0,1)$. The asymptotic α-level critical value of the symmetrical t test is $z_{\alpha/2}$, the $1 - \alpha/2$ quantile of the standard normal distribution. Theorem 4.1 shows that $P(|t_n| > z_{\alpha/2}) = \alpha + O[(nh_n)^{-1}]$. In other words, when the asymptotic critical value is used, the difference between the true and nominal rejection probabilities of the symmetrical t test is $O[(nh_n)^{-1}]$.

Now consider the symmetrical t test with a bootstrap critical value. The bootstrap α-level critical value, $z_{n,\alpha/2}^*$, satisfies $P^*(|t_n^*| \geq z_{n,\alpha/2}^*) = \alpha$. By Theorem 4.1,

$$(4.16) \qquad P^*(|t_n^*| > t) - P(|t_n| > t) = o[(nh_n)^{-1}]$$

almost surely uniformly over t. It can also be shown that $P(|t_n| > z_{n,\alpha/2}^*) = \alpha + o[(nh_n)^{-1}]$. Thus, with the bootstrap critical value, the difference between the true and nominal rejection probabilities of the symmetrical t test is $o[(nh_n)^{-1}]$. The bootstrap reduces the difference between the true and nominal rejection probabilities because it accounts for the effects of the $O[(nh_n)^{-1}]$ term of the Edgeworth expansion of the distribution of t_n. First-order asymptotic approximations ignore this term. Thus, the bootstrap provides asymptotic refinements for hypothesis tests and confidence intervals based on a kernel nonparametric density estimator provided that the bandwidth h_n converges sufficiently rapidly to make the asymptotic bias of the density estimator negligibly small.

The conclusion that first-order asymptotic approximations make an error of size $O[(nh_n)^{-1}]$ assumes that $nh_n^{r+1} \to 0$. If this condition is not satisfied, the error made by first-order approximations is dominated by the effect of asymptotic bias and is larger than $O[(nh_n)^{-1}]$ This result is derived at the end of this section.

The bootstrap can also be used to obtain asymptotic refinements for one-sided and equal-tailed tests and confidence intervals. For one-sided tests and confidence intervals with bootstrap critical values, the differences between the true and nominal rejection and coverage probabilities are $O[(nh_n)^{-1} + (nh_n)^{1/2}h_n^r]$. These are minimized by setting $h_n \propto n^{-3/(2r+3)}$, in which case the errors are $O[n^{-2r/(2r+3)}]$. For equal-tailed tests and confidence intervals with bootstrap critical values, the differences between the true and nominal rejection probabilities and coverage probabilities are $O[(nh_n)^{-1} + nh_n^{2r+1} + h_n^r]$. These are minimized by setting $h_n \propto n^{-1/(r+1)}$, in which case the errors are $O[n^{-r/(r+1)}]$. In contrast, the error made by first-order asymptotic approximations is $O[(nh_n)^{-1/2}]$ in both the one-sided and equal-tailed cases. Hall (1992a, pp. 220-224) provides details and a discussion of certain exceptional cases in which smaller errors can be achieved. In contrast to the situation with the smooth function model, the orders of refinement achievable in nonparametric density estimation are different for one-sided and equal-tailed tests and confidence intervals.

The Error Made by First-Order Asymptotics when nh_n^{r+1} Does Not Converge to 0: The effects of having $h_n \to 0$ too slowly are most easily seen by assuming that $s_n(x)$ is known so that t_n is replaced by

$$Z_n = \frac{f_n(x) - f(x) - b_n(x)}{s_n(x)}.$$

A symmetrical test of H_0 rejects if $|f_n(x) - f_0|/s_n(x)$ is large. If H_0 is true, then

$$P\left\{\frac{f_n(x)-f_0}{s_n(x)} \leq z\right\} = P\left\{Z_n \leq z - \frac{b_n(x)}{s_n(x)}\right\}$$

for any z, and

$$P\left[\frac{|f_n(x)-f_0|}{s_n(x)} \leq z\right] = P\left[Z_n \leq z - \frac{b_n(x)}{s_n(x)}\right] - P\left[Z_n \leq -z - \frac{b_n(x)}{s_n(x)}\right]. \tag{4.17}$$

Each term on the right-hand side of (4.17) has an asymptotic expansion of the form (4.15) except without the q_3 term and the $O(n^{-1})$ remainder term, which arise from random sampling error in $s_n^2(x)$. Specifically,

$$\begin{aligned} P\left[\frac{|f_n(x)-f_0|}{s_n(x)} \leq z\right] &= \Phi\left[z - \frac{b_n(x)}{s_n(x)}\right] - \Phi\left[-z - \frac{b_n(x)}{s_n(x)}\right] \\ &+ \frac{1}{(nh_n)^{1/2}}\left\{p_1\left[z - \frac{b_n(x)}{s_n(x)}\right] - p_1\left[-z - \frac{b_n(x)}{s_n(x)}\right]\right\} \\ &+ \frac{1}{nh_n}\left\{p_2\left[z - \frac{b_n(x)}{s_n(x)}\right] - p_2\left[-z - \frac{b_n(x)}{s_n(x)}\right]\right\} + O[(nh_n)^{-3/2}], \end{aligned} \tag{4.18}$$

where p_1 is an even function and p_2 is an odd function. Hall (1992a, p. 212) provides a proof and the details of p_1 and p_2. A Taylor series expansion of the right-hand side of (4.18) combined with $b_n(x) = O(h_n^r)$ and $s_n(x) = O[(nh_n)^{-1/2}]$ yields

$$P\left[\frac{|f_n(x)-f_0|}{s_n(x)} \leq z\right] = \Phi(z) - \Phi(-z) + O[h_n^r + (nh_n)h_n^{2r} + (nh_n)^{-1}]. \tag{4.19}$$

The remainder term on the right-hand side of (4.19) is dominated by h_n^r, which is the effect of asymptotic bias, unless $nh_n^{r+1} \to 0$. Thus, the error made by first-order asymptotic approximations exceeds $O[(nh_n)^{-1}]$ unless $f_n(x)$ is sufficiently undersmoothed to make the asymptotic bias $b_n(x)$ negligible, which is equivalent to requiring $nh_n^{r+1} \to 0$ as $n \to \infty$.

4.2.4 *Kernel Nonparametric Mean Regression*

In nonparametric mean-regression, the aim is to infer the mean of a random variable Y conditional on a covariate X without assuming that the conditional mean function belongs to a known finite-dimensional family of functions. Define $G(x) = E(Y|X = x)$ to be the conditional mean function. Let X be a scalar random variable whose distribution has a probability density function f. This section explains how the bootstrap can be used to obtain asymptotic refinements for tests of hypotheses about $G(x)$ and confidence intervals that are based on kernel estimation of G..

Let the data consist of a random sample, $\{Y_i, X_i\colon\ i = 1, \ldots, n\}$, of the joint distribution of (Y, X). The kernel nonparametric estimator of $G(x)$ is

$$G_n(x) = \frac{1}{nh_n f_n(x)} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right),$$

where

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

K is a kernel function and $\{h_n\}$ a sequence of bandwidths. The properties of $G_n(x)$ are discussed by Härdle (1990). To state the ones that are relevant here, let $r \geq 2$ be an even integer. Assume that G and f each have r bounded, continuous derivatives in a neighborhood of x. Let K be a bounded function that is symmetrical about 0, has support [-1,1], and satisfies (4.11). Define B_K and A_K as in Section 4.2.1. Set $V(z) = Var(Y|X = z)$, and assume that this quantity is finite and continuous in a neighborhood of $z = x$. Also define

$$b_n(x) = h_n^r \frac{A_K}{r!\, f(x)} \left\{ \frac{\partial^r}{\partial x^r}[G(x) f(x)] - f^{(r)}(x) \right\}$$

and

$$\text{(4.20)} \qquad s_n^2(x) = \frac{B_K}{nh_n} \frac{V(x)}{f(x)}.$$

If nh_n^{2r+1} is bounded as $n \to \infty$, then

$$Z_n(x) \equiv \frac{G_n(x) - G(x) - b_n(x)}{s_n(x)} \to^d N(0,1)$$

The fastest possible rate of convergence of $G_n(x)$ to $G(x)$ is achieved by setting $h_n \propto n^{-1/(2r+1)}$. When this happens, $G_n(x) - G(x) = O_p[n^{-r/(2r+1)}]$, $b_n(x) \propto n^{-r/(2r+1)}$, and $s_n(x) \propto n^{-r/(2r+1)}$.

The issues involved in converting Z_n into an asymptotically pivotal statistic that can be used to test a hypothesis about $G(x)$ or form a confidence interval for $G(x)$ are the same as in kernel density estimation. It is necessary to replace $s_n(x)$ with a suitable estimator and to remove the asymptotic bias $b_n(x)$. As in kernel density estimation, asymptotic bias can be removed to sufficient order by undersmoothing. Undersmoothing for a symmetrical test or confidence interval consists of choosing h_n so that $nh_n^{r+1} \to 0$ as $n \to \infty$.[19]

Now consider estimation of $s_n^2(x)$. One possibility is to replace $f(x)$ with $f_n(x)$ and $V(x)$ with a consistent estimator on the right-hand side of (4.20). The higher-order asymptotics of $G_n(x)$ are simpler, however, if $s_n^2(x)$ is estimated by a sample analog of the exact finite-sample variance of the asymptotic form of $G_n(x) - G(x)$. With asymptotic bias removed by undersmoothing, the asymptotic form of $G_n(x) - G(x)$ is

(4.21) $$G_n(x) - G(x) = \frac{1}{nh_n f(x)} \sum_{i=1}^{n} [Y_i - G(x)] K\left(\frac{x - X_i}{h_n}\right) + o_p(1) .$$

The variance of the first term on the right-hand side of (4.21) is then estimated by the following sample analog, which will be used here to estimate $s_n^2(x)$[20]:

$$s_n^2(x) = \frac{1}{[nh_n f_n(x)]^2} \sum_{i=1}^{n} [Y_i - G_n(x)]^2 K\left(\frac{x - X_i}{h_n}\right)^2 .$$

Now define

$$t_n = \frac{G_n(x) - G(x)}{s_n(x)} .$$

With asymptotic bias removed through undersmoothing, t_n is asymptotically distributed as $N(0,1)$ and is an asymptotically pivotal statistic that can be used to test a hypothesis about $G(x)$ and to form a confidence interval for $G(x)$. The bootstrap version of t_n is

$$t_n^* = \frac{G_n^*(x) - G_n(x)}{s_n^*(x)},$$

where $G_n^*(x)$ is obtained from $G_n(x)$ by replacing the sample $\{Y_i, X_i\}$ with the bootstrap sample $\{Y_i^*, X_i^*\}$, and $s_n^*(x)$ is obtained from $s_n(x)$ by replacing the sample with the bootstrap sample, $f_n(x)$ with $f_n^*(x)$, and $G_n(x)$ with $G_n^*(x)$.[21]

The Edgeworth expansions of the distributions of t_n and t_n^* are similar in structure to those of the analogous statistic for kernel density estimators. The result for symmetrical tests and confidence intervals can be stated as follows. Let $E(Y^4 | X = z)$ be finite and continuous for all z in a neighborhood of x. Let K satisfy the conditions of Theorem 4.1. Then there are functions q and q_n such that $q_n - q = o(1)$ uniformly and almost surely as $n \to \infty$,

$$(4.22)\qquad P(|t_n| \leq t) = 2\Phi(t) - 1 + \frac{1}{nh_n} q(t) + o[(nh_n)^{-1}]$$

uniformly over t, and

$$P^*(|t_n^*| \leq t) = 2\Phi(t) - 1 + \frac{1}{nh_n} q_n(t) + o[(nh_n)^{-1}]$$

uniformly over t almost surely. It follows that the bootstrap estimator of the distribution of $|t_n|$ is accurate through $O[(nh_n)^{-1}]$, whereas first-order asymptotic approximations make an error of this size. Let $z_{n,a/2}^*$ be the bootstrap a-level critical value of for testing the hypothesis H_0: $G(x) = G_0$. Then $P^*(|t_n^*| > z_{n,a/2}^*) = a$, and it can be shown that $P(|t_n| > z_{n,a/2}^*) = a + o[(nh_n)^{-1}]$. Hall (1992, Section 4.5) discusses the mathematical details. Thus, with the bootstrap critical value, the true and nominal rejection probabilities of a symmetrical t test of H_0 differ by $o[(nh_n)^{-1}]$. In contrast, it follows from (4.22) that the difference is $O[(nh_n)^{-1}]$ if first-order asymptotic approximations are used to obtain the critical value. The same conclusions hold for the coverage probabilities of symmetrical confidence intervals for $G(x)$.

4.3 *Non-Smooth Estimators*

Some estimators are obtained by maximizing or minimizing a function that is discontinuous or whose first derivative is discontinuous. Two important examples are Manski's (1975, 1985) maximum-score (MS) estimator of the slope coefficients of a binary-response model and the least-absolute deviations (LAD) estimator of the slope coefficients of a linear median-regression model. The objective function of the MS estimator and the first derivative of the objective function of the LAD estimator are step functions and, therefore, discontinuous. The LAD and MS estimators cannot be approximated by smooth functions of sample moments, so they do not satisfy the assumptions of the smooth function model. Moreover, the Taylor-series methods of asymptotic distribution theory do not apply to the LAD and MS estimators, which greatly complicates the analysis of their asymptotic distributional properties. As a consequence, little is known about the ability of the bootstrap to provide asymptotic refinements for hypothesis tests and confidence intervals based on these estimators. Indeed it is not known whether the bootstrap even provides a consistent approximation to the asymptotic distribution of the MS estimator.

This section explains how the LAD and MS estimators can be smoothed in a way that greatly simplifies the analysis of their asymptotic distributional properties. The bootstrap provides asymptotic refinements for hypothesis tests and confidence intervals based on the smoothed LAD and MS estimators. In addition, smoothing accelerates the rate of convergence of the MS estimator and simplifies even its first-order asymptotic distribution. Smoothing does not change the rate of convergence or first-order asymptotic distribution of the LAD estimator. The LAD estimator is treated in Section 4.3.1, and the MS estimator is treated in Section 4.3.2

4.3.1 *The LAD Estimator for a Linear Median-Regression Model*

A linear median-regression model has the form

(4.23) $Y = Xb + U$,

where Y is an observed scalar, X is an observed $1\times q$ vector, b is a $q\times 1$ vector of constants, and U is an unobserved random variable that satisfies median($U|X = x$) = 0 almost surely. Let $\{Y_i, X_i: i = 1, \ldots, n\}$ be a random sample from the joint distribution of (Y, X) in (4.23). The LAD estimator of b, $\tilde{b}_n$, solves

$$\underset{b\in B}{\text{minimize}}: \tilde{H}_n(b) \equiv \frac{1}{n}\sum_{i=1}^{n}|Y_i - X_i b|$$

$$= \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i b)[2I(Y_i - X_i b > 0) - 1], \tag{4.24}$$

where B is the parameter set and $I(\cdot)$ is the indicator function. Bassett and Koenker (1978) and Koenker and Bassett (1978) give conditions under which the LAD estimator is $n^{1/2}$-consistent and $n^{1/2}(\tilde{b}_n - b)$ is asymptotically normal.

$\tilde{H}_n(b)$ has cusps and, therefore, a discontinuous first derivative, at points b such that $Y_i = X_i b$ for some i. This non-smoothness causes the Edgeworth expansion of the LAD estimator to be non-standard and very complicated (De Angelis *et al*. 1993). The bootstrap is known to estimate the distribution of $n^{1/2}(\tilde{b}_n - b)$ consistently (De Angelis *et al*. 1993, Hahn 1995), but it is not known whether the bootstrap provides asymptotic refinements for hypothesis tests and confidence intervals based on $\tilde{b}_n$.[22]

Horowitz (1998b) suggests removing the cusps in $\tilde{H}_n$ by replacing the indicator function with a smooth function, thereby producing a modified objective function whose derivatives are continuous. The resulting smoothed LAD (SLAD) estimator is first-order asymptotically equivalent to the unsmoothed LAD estimator but has much simpler higher-order asymptotics.

Specifically, let K be a bounded, differentiable function satisfying $K(v) = 0$ if $v \leq -1$ and $K(v) = 1$ if $v \geq 1$. Let $\{h_n\}$ be a sequence of bandwidths that converges to 0 as $n \to \infty$. The SLAD estimator solves

$$\text{(4.25)} \qquad \underset{b \in B}{\text{minimize}}: H_n(b) \equiv \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i b)\left[2K\left(\frac{Y_i - X_i b}{h_n}\right) - 1\right].$$

K is analogous to the integral of a kernel function for nonparametric density estimation. K is not a kernel function itself.

Let b_n be a solution to (4.25). Horowitz (1998b) gives conditions under which $n^{1/2}(b_n - \tilde{b}_n) = o_p(1)$. Thus, the smoothed and unsmoothed LAD estimators are first-order asymptotically equivalent. It follows from this asymptotic equivalence and the asymptotic normality of LAD estimators that $n^{1/2}(b_n - b) \to^d N(0,V)$, where $V = D^{-1}E(X'X)D^{-1}$, $D = 2E[X'Xf(0|x)]$, and $f(\cdot|x)$ is the probability density function of U conditional on $X = x$.

A t statistic for testing a hypothesis about a component of b or forming a confidence interval can be constructed from consistent estimators of D and $E(X'X)$. D can be estimated consistently by $D_n(b_n)$, where

$$\text{(4.26)} \qquad D_n(b) = \frac{2}{nh_n}\sum_{i=1}^{n} X_i' X_i K'\left(\frac{Y_i - X_i b}{h_n}\right).$$

$E(X'X)$ can be estimated consistently by the sample average of $X'X$. However, the asymptotic expansion of the distribution of the t statistic is simpler if $E(X'X)$ is estimated by the sample analog of the exact finite-sample variance of $\partial H_n(b)/\partial b$ at $b = b$. This estimator is $T_n(b_n)$, where

$$\text{(4.27)} \qquad T_n(b) = \frac{1}{n}\sum_{i=1}^{n} X_i' X_i \left\{\left[2K\left(\frac{Y_i - X_i b}{h_n}\right) - 1\right] + 2\left(\frac{Y_i - X_i b}{h_n}\right)K'\left(\frac{Y_i - X_i b}{h_n}\right)\right\}^2.$$

It is not difficult to show that V is estimated consistently by $V_n \equiv D_n(b_n)^{-1}T_n(b_n)D_n(b_n)^{-1}$. Now let b_{nj} and b_j, respectively, be the j'th components of b_n and b ($j = 1, \ldots, q$). Let V_{nj} be the (j, j)

component of V_n. The t statistic for testing H_0: $b_j = b_{j0}$ is $t_n = n^{1/2}(b_{nj} - b_{j0})/V_{nj}^{1/2}$. If H_0 is true, then $t_n \to^d N(0,1)$, so t_n is asymptotically pivotal.

To obtain a bootstrap version of t_n, let $\{Y_i^*, X_i^*: i = 1, \ldots, n\}$ be a bootstrap sample that is obtained by sampling the data $\{Y_i, X_i\}$ randomly with replacement. Let b_n^* be the estimator of b that is obtained by solving (4.25) with $\{Y_i^*, X_i^*\}$ in place of $\{Y_i, X_i\}$. Let V_{nj}^* be the version of V_{nj} that is obtained by replacing b_n and $\{Y_i, X_i\}$, respectively, with b_n^* and $\{Y_i^*, X_i^*\}$ in (4.26) and (4.27). Then the bootstrap analog of t_n is $t_n^* = n^{1/2}(b_{nj}^* - b_{nj})/(V_{nj}^*)^{1/2}$.

By using methods similar to those used with kernel density and mean-regression estimators, it can be shown that under regularity conditions, t_n and t_n^* have Edgeworth expansions that are identical almost surely through $O[(nh_n)^{-1}]$. Horowitz (1998b) gives the details of the argument. In addition, reasoning similar to that used in Section 4.2.3 shows that the bootstrap provides asymptotic refinements for hypothesis tests and confidence intervals based on the SLAD estimator. For example, consider a symmetrical t test of H_0. Let $z_{n,a/2}^*$ be the bootstrap a-level critical value for this test. That is, $z_{n,a/2}^*$ satisfies $P^*(|t_n^*| > z_{n,a/2}^*) = a$. Then $P(|t_n| > z_{n,a/2}^*) = a + o[(nh_n)^{-1}]$. In contrast, first-order asymptotic approximations make an error of size $O[(nh_n)^{-1}]$. This is because first-order approximations ignore a term in the Edgeworth expansion of the distribution of $|t_n|$ whose size is $O[(nh_n)^{-1}]$, whereas the bootstrap captures the effects of this term.

The conditions under which this result holds include: (1) for almost every x and every u in a neighborhood of 0, $f(u|x)$ is r -1 times continuously differentiable with respect to u; (2) K satisfies (4.11) and has four bounded, Lipschitz continuous derivatives everywhere; and (3) $h_n \propto n^{-k}$, where $2/(2r + 1) < k < 1/3$. Complete regularity conditions are given in Horowitz (1998b). Condition (3) implies that $r \geq 4$. Therefore, the size of the refinement obtained by the bootstrap is $O(n^{-c})$, where $7/9 < c < 1$.

The bootstrap also provides asymptotic refinements for one-sided tests and confidence intervals and for asymptotic chi-square tests of hypotheses about several components of b. In addition, it is possible to construct a smoothed version of Powell's (1984, 1986) censored LAD estimator and to show that the bootstrap provides asymptotic refinements for tests and confidence intervals based on the smoothed censored LAD estimator. Horowitz (1998b) provides details, a method for choosing h_n in applications, and Monte Carlo evidence on the numerical performance of the t test with bootstrap critical values.

4.3.2 *The Maximum Score Estimator for a Binary-Response Model*

The most frequently used binary-response model has the form $Y = I(Xb + U \geq 0)$, where X is an observed random vector, b is a conformable vector of constants, and U is an unobserved random variable. The parameter vector b is identified only up to scale, so a scale normalization is needed. Here, scale normalization will be accomplished by assuming that $|b_1| = 1$, where b_1 is the first component of b. Let $\tilde{b}$ and $\tilde{b}$ denote the vectors consisting of all components of b and b except the first. The maximum-score estimator of b, $b_n \equiv (b_{n1}, \tilde{b}_n{}')'$, solves

$$(4.28) \qquad \underset{b \in B}{\text{maximize}}: \tilde{H}_n(b) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)I(X_i b \geq 0),$$

where $\{Y_i, X_i: \ i = 1, \ldots, n\}$ is a random sample from the joint distribution of (Y, X), and B is a compact parameter set in which the scale normalization holds

Manski (1975, 1985) shows that if median$(U|X = x) = 0$ almost surely, the first component of X is continuously distributed with a non-zero coefficient, and certain other conditions are satisfied, then $(b_{n1}, \tilde{b}_n{}')' \to b$ almost surely. Because $b_{n1} = \pm 1$, b_{n1} converges to b_1 faster than any power of n. Cavanagh (1987) and Kim and Pollard (1990) show that $\tilde{b}_n$ converges in probability at the rate $n^{-1/3}$ and that $n^{1/3}(\tilde{b}_n - \tilde{b})$ has a complicated, non-normal asymptotic distribution. The MS estimator is important despite its slow rate of convergence and complicated limiting distribution because it is semiparametric (that is, it does not require the distribution of U to belong to a known, finite-dimensional family) and it permits the distribution of U to have arbitrary heteroskedasticity of unknown form provided that the centering assumption median($U|X = x$) = 0 holds.

The asymptotic distribution of the MS estimator is too complex for use in testing hypotheses about b or constructing confidence intervals. Manski and Thompson (1986) suggested using the bootstrap to estimate the mean-square error of the MS estimator and presented Monte Carlo evidence suggesting that the bootstrap works well for this purpose. However, it is not known whether the bootstrap consistently estimates the asymptotic distribution of the MS estimator.

The MS estimator converges slowly and has a complicated limiting distribution because it is obtained by maximizing a step function. Horowitz (1992) proposed replacing the indicator function on the right-hand side of (4.28) by a differentiable function. The resulting estimator is called the *smoothed maximum score* (SMS) estimator. It solves

$$\text{(4.29)} \qquad \underset{b\in B}{\text{maximize}}: H_n(b) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)K\left(\frac{X_i b}{h_n}\right),$$

where K is a bounded, differentiable function satisfying $K(v) = 0$ if $v \leq -1$ and $K(v) = 1$ if $v \geq 1$, and $\{h_n\}$ is a sequence of bandwidths that converges to 0 as $n \to \infty$. As in SLAD estimation, K is analogous to the integral of a kernel function. Let $\tilde{b}$ again be the vector of all components of b but the first. Let $b_n \equiv (b_{n1}, \tilde{b}_n')'$ be the SMS estimator of $(b_1, \tilde{b}')'$. Horowitz (1992) gives conditions under which $(nh_n)^{1/2}(\tilde{b}_n - \tilde{b} - h_n^r l) \to^d N(0,V)$, where $r \geq 2$ is an integer that is related to the number of times that the CDF of U and the density function of Xb are continuously differentiable, nh_n^{2r+1} is bounded as $n \to \infty$, l is an asymptotic bias, and V is a covariance matrix.

The rate of convergence of the SMS estimator of $\tilde{b}$ is at least $n^{-2/5}$ and can be arbitrarily close to $n^{-1/2}$ if the CDF of U and density function of Xb have sufficiently many derivatives. Thus, smoothing increases the rate of convergence of the MS estimator.

To obtain an asymptotically pivotal t statistic for testing a hypothesis about a component of $\tilde{b}$ or forming a confidence interval, it is necessary to remove the asymptotic bias of $\tilde{b}_n$ and construct a consistent estimator of V. Asymptotic bias can be removed by undersmoothing. For first-order asymptotic approximations, undersmoothing consists of choosing h_n so that $nh_n^{2r+1} \to 0$ as $n \to \infty$. However, for the reasons explained in the discussion of equation (4.19), the stronger condition $nh_n^{r+1} \to 0$ is needed to obtain asymptotic refinements through $O[(nh_n)^{-1}]$. V can be estimated consistently by $V_n = Q_n(b_n)^{-1}D_n(b_n)Q_n(b_n)^{-1}$, where for any $b \in B$

(4.30) $$Q_n(b) = \frac{1}{nh_n^2}\sum_{i=1}^{n}(2Y_i - 1)\tilde{X}_i'\tilde{X}_i K''\left(\frac{X_i b}{nh_n}\right),$$

(4.31) $$D_n(b) = \frac{1}{nh_n}\sum_{i=1}^{n}\tilde{X}_i'\tilde{X}_i\left[K'\left(\frac{X_i b}{h_n}\right)\right]^2,$$

and $\tilde{X}$ consists of all components of X but the first.

Now let $\tilde{b}_{nj}$ and $\tilde{b}_j$, respectively, be the j'th components of $\tilde{b}_n$ and $\tilde{b}$. Let V_{nj} be the (j, j) component of V_n. The t statistic for testing H$_0$: $\tilde{b}_j = \tilde{b}_{j0}$ is $t_n = (nh_n)^{1/2}(\tilde{b}_{nj} - \tilde{b}_{j0})/V_{nj}^{1/2}$. If H$_0$ is true, then $t_n \to^d N(0,1)$, so t_n is asymptotically pivotal.

To obtain a bootstrap version of t_n, let $\{Y_i^*, X_i^*: i = 1, \ldots, n\}$ be a bootstrap sample that is obtained by sampling the data $\{Y_i, X_i\}$ randomly with replacement. Let b_n^* be the estimator of b that is obtained by solving (4.29) with $\{Y_i^*, X_i^*\}$ in place of $\{Y_i, X_i\}$. Let V_{nj}^* be the version of V_{nj} that is obtained by replacing b_n and $\{Y_i, X_i\}$, respectively, with b_n^* and $\{Y_i^*, X_i^*\}$ in (4.30) and (4.31). Then the bootstrap analog of t_n is $t_n^* = (nh_n)^{1/2}(\tilde{b}_{nj}^* - \tilde{b}_{nj})/(V_{nj}^*)^{1/2}$.

By using methods similar to those used with kernel density and mean-regression estimators, it can be shown that t_n and t_n* have Edgeworth expansions that are identical almost surely through $O[(nh_n)^{-1}]$. See Horowitz (1998c) for the details of the argument. It follows that the bootstrap provides asymptotic refinements for hypothesis tests and confidence intervals based on the SMS estimator. For a symmetrical t test or confidence interval, the true and nominal rejection or coverage probabilities differ by $o[(nh_n)^{-1}]$ when bootstrap critical values are used, whereas they differ by $O[(nh_n)^{-1}]$ when first-order asymptotic critical values are used. First-order approximations ignore a term in the Edgeworth expansion of the distribution of $|t_n|$ whose size is $O[(nh_n)^{-1}]$, whereas the bootstrap captures the effects of this term.

The conditions under which this result holds include: (1) the CDF of U conditional on X and the density of Xb conditional on X have sufficiently many derivatives; (2) K satisfies (4.11) for some $r \geq 8$; and (3) $h_n \propto n^{-k}$, where $1/(r + 1) < k < 1/7$. Complete regularity conditions are given in Horowitz (1998c). Conditions (2) and (3) imply that the size of the refinement obtained by the bootstrap is $O(n^{-c})$, where $6/7 < c < 1$. The bootstrap also provides asymptotic refinements for one-sided tests and confidence intervals and for asymptotic chi-square tests of hypotheses about several components of $\tilde{b}$. Horowitz (1998c) discusses methods for choosing h_n in applications and gives Monte Carlo evidence on the numerical performance of the t test with bootstrap critical values.

4.4 *Bootstrap Iteration*

The discussion of asymptotic refinements in this chapter has emphasized the importance of applying the bootstrap to asymptotically pivotal statistics. This section explains how the bootstrap can be used to create an asymptotic pivot when one is not available. Asymptotic refinements can be obtained by applying the bootstrap to the bootstrap-generated asymptotic pivot. The computational procedure is called *bootstrap iteration* or *prepivoting* because it entails drawing bootstrap samples from bootstrap samples as well as using the bootstrap to create an asymptotically pivotal statistic.

The discussion here concentrates on the use of prepivoting to test hypotheses (Beran 1988). Beran (1987) explains how to use prepivoting to form confidence regions. Hall (1986b) describes an alternative approach to bootstrap iteration.

Let T_n be a statistic for testing a hypothesis H_0 about a sampled population whose CDF is F_0. Assume that under H_0, T_n satisfies assumptions SFM and (3.8) of the smooth function model. Define $F = F_0$ if H_0 is true, and define F to be the CDF of a distribution that satisfies H_0 otherwise. Let $G_n(t, F) \equiv P_F(T_n \leq t)$ denote the exact, finite-sample CDF of T_n under sampling from the population whose CDF is F. Suppose that H_0 is rejected if T_n is large. Then the exact α-level critical value of T_n, $z_{n\alpha}$, is the solution to $G_n(z_{n\alpha}, F) = 1 - \alpha$ under H_0. An exact α-level test based on T_n can be obtained by rejecting H_0 if $G_n(T_n, F) > 1 - \alpha$. Thus, if F were known, $g_n \equiv G_n(T_n, F)$ could be used as a statistic for testing H_0. Prepivoting is based on the idea of using g_n as a test statistic.

A test based on g_n cannot be implemented in an application unless T_n is pivotal because F and, therefore, g_n are unknown. A feasible test statistic can be obtained by replacing F with an estimator F_n that imposes the restrictions of H_0 and is $n^{1/2}$-consistent for F_0 if H_0 is true. Replacing F with F_n produces the bootstrap statistic $g_n{}^* = G_n(T_n, F_n)$. $G_n(\cdot, F_n)$ and, therefore, $G_n(T_n, F_n)$ can be estimated with arbitrary accuracy by carrying out a Monte Carlo simulation in which random samples are drawn from F_n. Given any t, let $H_n(t, F_0) = P_{F_0}(g_n{}^* \leq t)$ = $P_{F_0}[G_n(T_n, F_n) \leq t]$. An exact test based on $g_n{}^*$ rejects H_0 at the α level if $H_n(g_n{}^*, F_0) > 1 - \alpha$. This test cannot be implemented because F_0 is unknown. If the bootstrap is consistent, however, the asymptotic distribution of $g_n{}^*$ is uniform on [0,1]. Therefore, H_0 is rejected at the asymptotic α level if $g_n{}^* > 1 - \alpha$. Now observe that $g_n{}^*$ is asymptotically pivotal even if T_n is not; the asymptotic distribution of $g_n{}^*$ is $U[0,1]$ regardless of F_0. This suggests that asymptotic refinements can be obtained by carrying out a second stage of bootstrap sampling in which the bootstrap is used to estimate the finite-sample distribution of $g_n{}^*$.

The second stage of bootstrapping consists of drawing samples from each of the first-stage bootstrap samples that are used to compute g_n*. Suppose that there are M first-stage samples. The m'th such sample yields a bootstrap version of T_n, say T_{nm}, and an estimator F_{nm} of F_n that is consistent with H_0. F_{nm} can be sampled repeatedly to obtain $G_n(\cdot, F_{nm})$, the EDF of T_n under sampling from F_{nm}, and $g_{nm} \equiv G_n(T_{nm}, F_{nm})$. Now estimate $H_n(\cdot, F_0)$ by $H_n(\cdot, F_n)$, which is the EDF of g_{nm} ($m = 1, \ldots, M$). The iterated bootstrap test rejects H_0 at the α level if $H_n(g_n^*, F_n) > 1 - \alpha$.

Beran (1988) shows that when prepivoting and bootstrap iteration are applied to a statistic T_n, the true and nominal probabilities of rejecting a correct null hypothesis differ by $o(n^{-1/2})$ for a one-sided test and $o(n^{-1})$ for a symmetrical test even if T_n is not asymptotically pivotal. By creating an asymptotic pivot in the first stage of bootstrapping, prepivoting and bootstrap iteration enable asymptotic refinements to be obtained for a non-asymptotically-pivotal T_n. The same conclusions apply to the coverage probabilities of confidence intervals. Beran (1988) presents the results of Monte Carlo experiments that illustrate the numerical performance of this procedure.

The computational procedure for carrying out prepivoting and bootstrap iteration is given by Beran (1988) and is as follows:

1. Obtain T_n and F_n from the estimation data $\{X_i: i = 1, \ldots, n\}$, which are assumed to be a random sample of a possibly vector-valued random variable X.

2. Let $c_1, \ldots, c_M$ be M bootstrap samples of size n that are drawn from the population whose distribution is F_n. Let F_{nm} denote the estimate of F_n that is obtained from c_m. Let T_{nm} be the version of T_n that is obtained from c_m. The EDF of $\{T_{nm}: m = 1, \ldots, M\}$ estimates $G_n(\cdot, F_n)$. Set $g_n^* = M^{-1}\sum_{m=1}^{M} I(T_{nm} \le T_n)$.

3. For each m, let $c_{m,1}, \ldots, c_{m,K}$ be K further bootstrap samples of size n, each drawn from the population whose CDF is F_{nm}. Let T_{nmk} be the version of T_n that is obtained from c_{mk}. Set $G_n(T_{nm}, F_{nm}) = K^{-1}\sum_{k=1}^{K} I(T_{nmk} \le T_{nm})$. Each of the $G_n(T_{nm}, F_{nm})$ ($m = 1, \ldots, n$} is a second-

stage estimate of g_n. Estimate $H_n(g_n^*, F_0)$ by $H_n(g_n^*, F_n) = M^{-1}\sum_{m=1}^{M} I[G_n(T_{nm}, F_{nm}) \leq g_n^*]$. Reject H_0 at the α level if $H_n(g_n^*, F_n) > 1 - \alpha$.

4.5 *Special Problems*

The bootstrap provides asymptotic refinements because it amounts to a one-term Edgeworth expansion. The bootstrap cannot be expected to perform well when an Edgeworth expansion provides a poor approximation to the distribution of interest. An important case of this is instrumental-variables estimation with poorly correlated instruments and regressors. It is well known that first-order asymptotic approximations are especially poor in this situation (Hillier 1985, Nelson and Startz 1990ab, Phillips 1983). The bootstrap does not offer a solution to this problem. With poorly correlated instruments and regressors, Edgeworth expansions of estimators and test statistics involve denominator terms that are close to zero. As a result, the higher-order terms of the expansions may dominate the lower-order ones for a given sample size, in which case the bootstrap may provide little improvement over first-order asymptotic approximations. Indeed, with small samples the numerical accuracy of the bootstrap may be even worse than that of first-order asymptotic approximations.

The bootstrap also does not perform well when the variance estimator used for Studentization has a high variance itself. This problem can be especially severe when the parameters being estimated or tested are variances or covariances of a distribution. This happens, for example, in estimation of covariance structures of economic processes (Abowd and Card 1987, 1988; Behrman *et al*. 1994; Griliches 1979; Hall and Mishkin 1982). In such cases Studentization is carried out with an estimator of the variance of an estimated variance. Imprecise estimation of a variance also affects the finite-sample performance of asymptotically efficient GMM estimators because the asymptotically optimal weight matrix is the inverse of the covariance matrix of the GMM residuals. The finite-sample mean-square error of the asymptotically efficient estimator can greatly exceed the mean-square error of an asymptotically inefficient estimator that is obtained with a non-stochastic weight matrix. Horowitz (1998a) shows that in the case of estimating covariance structures, this problem can be greatly mitigated by using a trimmed version of the covariance estimator that excludes "outlier" observations. See Horowitz (1998a) for details. Section 5.5 presents a numerical illustration of the effects of trimming.

4.6 *The Bootstrap when the Null Hypothesis is False*

To understand the power of a test based on a bootstrap critical value, it is necessary to investigate the behavior of the bootstrap when the null hypothesis being tested, H_0, is false. Suppose that bootstrap samples are generated by a model that satisfies a false H_0 and, therefore, is misspecified relative to the true data-generation process. If H_0 is simple, meaning that it completely specifies the data-generation process, then the bootstrap amounts to Monte Carlo estimation of the exact finite-sample critical value for testing H_0 against the true data-generation process. Indeed, the bootstrap provides the exact critical value, rather than a Monte Carlo estimate, if $G(\cdot, F_n)$ can be calculated analytically. Tests of simple hypotheses are rarely encountered in econometrics, however.

In most applications, H_0 is composite. That is, it does not specify the value of a finite- or infinite-dimensional "nuisance" parameter y. In the remainder of this section, it is shown that a test of a composite hypothesis using a bootstrap-based critical value is a higher-order approximation to a certain exact test. The power of the test with a bootstrap critical value is a higher-order approximation to the power of the exact test.

Except in the case of a test based on a pivotal statistic, the exact finite-sample distribution of the test statistic depends on y. Therefore, except in the pivotal case, it is necessary to specify the value of y to obtain exact finite-sample critical values. The higher-order approximation to power provided by the bootstrap applies to a value of y that will be called the *pseudo-true value*. To define the pseudo-true value, let y_n be an estimator of y that is obtained under the incorrect assumption that H_0 is true. Under regularity conditions (see, e.g., Amemiya 1985, White 1982), y_n converges in probability to a limit y^*, and $n^{1/2}(y_n - y^*) = O_p(1)$. y^* is the pseudo-true value of y.

Now let T_n be a statistic that is asymptotically pivotal under H_0. Suppose that its exact CDF with an arbitrary value of y is $G_n(\cdot, y)$, and that under H_0 its asymptotic CDF is $G_0(\cdot)$. Suppose that bootstrap sampling is carried out subject to the constraints of H_0. Then the bootstrap generates samples from a model whose parameter value is y_n, so the exact distribution of the bootstrap version of T_n is $G_n(\cdot, y_n)$. Under H_0 and subject to regularity conditions, $G_n(\cdot, y_n)$ has an asymptotic expansion of the form

$$(4.32) \qquad G_n(z, y_n) = G_0(z) + n^{-j/2} g_j(z, y^*) + o_p(n^{-j/2})$$

uniformly over z, where $j = 1$ or 2 depending on the symmetry of T_n. Usually $j = 1$ if T_n is a statistic for a one-tailed test and $j = 2$ if T_n is a statistic for a symmetrical, two-tailed test. $G_n(z, y^*)$ has an expansion identical to (4.32) through $O(n^{-j/2})$. Therefore, through $O_p(n^{-j/2})$, bootstrap sampling when H_0 is false is equivalent to generating data from a model that satisfies H_0 with pseudo-true values of the parameters not specified by H_0. It follows that when H_0 is false, bootstrap-based critical values are equivalent through $O_p(n^{-j/2})$ to the critical values that would be obtained if the model satisfying H_0 with pseudo-true parameter values were correct. Moreover, the power of a test of H_0 using a bootstrap-based critical value is equal through $O(n^{-j/2})$ to the power against the true data-generation process that would be obtained by using the exact finite-sample critical value for testing H_0 with pseudo-true parameter values.

5. MONTE CARLO EXPERIMENTS

This section presents the results of some Monte Carlo experiments that illustrate the numerical performance of the bootstrap as a means of reducing differences between the true and nominal rejection probabilities of tests of statistical hypotheses.

5.1 *The Information-Matrix Test*

White's (1982) information-matrix (IM) test 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 Monte Carlo experiments carried out by many investigators have shown that the asymptotic distribution is a very poor approximation to the true, finite-sample distribution. With sample sizes in the range found in applications, the true and nominal probabilities that the IM test with asymptotic critical values rejects a correct model can differ by a factor of 10 or more (Horowitz 1994, Kennan and Neumann 1988, Orme 1990, Taylor 1987).

Horowitz (1994) reports the results of Monte Carlo experiments that investigate the ability of the bootstrap to provide improved finite-sample critical values for the IM test, thereby reducing the distortions of RP's that occur with asymptotic critical values. Three forms of the test were used: the Chesher (1983) and Lancaster (1984) form, White's (1982) original form, and Orme's (1990) w_3. The Chesher-Lancaster form is relatively easy to compute because, in contrast to the other forms, it does not require third derivatives of the log-density function or analytic expected values of derivatives of the log-density. However, first-order asymptotic theory gives an especially poor approximation to its finite-sample distribution. Orme (1990) found through Monte Carlo experimentation that the distortions of RP's are smaller with w_3 than with many other forms of the IM test statistic. Orme's w_3 uses expected values of third derivatives of the log-density, however, so it is relatively difficult to compute.

Horowitz's (1994) experiments consisted of applying the three forms of the IM test to Tobit and binary probit models. Each model had either one or two explanatory variables X that were obtained by sampling either the $N(0,1)$ or the $U[0,1]$ distribution. There were 1000 replications in each experiment. Other details of the Monte Carlo procedure are described in Horowitz (1994).

Table 1 summarizes the results of the experiments. As expected, the differences between empirical and nominal RP's are very large when asymptotic critical values are used. This is especially true for the Chesher-Lancaster form of the test. When bootstrap critical values are used, however, the differences between empirical and nominal RP's are very small. The bootstrap essentially eliminates the distortions of the RP's of the three forms of the IM test.

5.2 *The t Test in a Heteroskedastic Regression Model*

In this section, the heteroskedasticity-consistent covariance matrix estimator (HCCME) of Eicker (1963,1967) and White (1980) is used to carry out a t test of a hypothesis about b in the model

(5.1) $Y = Xb + U$.

In this model, U is an unobserved random variable whose probability distribution is unknown and that may have heteroskedasticity of unknown form. It is assumed that $E(U|X = x) = 0$ and $Var(U|X = x) < \infty$ for all x in the support of X.

Let b_n be the ordinary least squares (OLS) estimator of b in (5.1), b_{ni} and b_i be the i'th components of b_n and b, and s_{ni} be the square root of the (i,i) element of the HCCME. The t statistic for testing H_0: $b_i = b_{i0}$ is $T_n = (b_{ni} - b_{i0})/s_{ni}$. Under regularity conditions, $T_n \to^d N(0,1)$ as $n \to \infty$. However, Chesher and Jewitt (1987) have shown that s_{ni}^2 can be seriously biased downward. Therefore, the true RP of a test based on T_n is likely to exceed the nominal RP. As is shown later in this section, the differences between the true and nominal RP's can be very large when n is small.

The bootstrap can be implemented for model (5.1) by sampling observations of (Y,X) randomly with replacement. The resulting bootstrap sample is used to estimate b by OLS and compute T_n*, the t statistic for testing H_0*: $b_i = b_{ni}$. The empirical distribution of T_n* is obtained by repeating this process many times, and the a-level bootstrap critical value for T_n* is estimated from this distribution. Since U may be heteroskedastic, the bootstrap cannot be implemented by resampling OLS residuals independently of X. Similarly, one cannot implement the bootstrap by sampling U from a parametric model because (5.1) does not specify the distribution of U or the form of any heteroskedasticity.

Randomly resampling (Y,X) pairs does not impose the restriction $E(U | X = x) = 0$ on the bootstrap sample. As will be seen later in this section, the numerical performance of the bootstrap can be improved greatly through the use of an alternative resampling procedure, called the *wild bootstrap*, that imposes this restriction. The wild bootstrap was introduced by Liu (1988) following a suggestion of Wu (1986). Mammen (1993) establishes the ability of the wild bootstrap to provide asymptotic refinements for the model (5.1). Cao-Abad (1991), Härdle and Mammen (1993), and Härdle and Marron (1991) use the wild bootstrap in nonparametric regression.

To describe the wild bootstrap, write the estimated form of (5.1) as

$$Y_i = X_i b_n + U_{ni}; \quad i = 1, 2, \ldots, n$$

where Y_i and X_i are the i'th observed values of Y and X, and U_{ni} is the i'th OLS residual. For each $i = 1, \ldots, n$, let F_i be the unique 2-point distribution that satisfies $E(Z|F_i) = 0$, $E(Z^2|F_i) = U_{ni}^2$, and $E(Z^3|F_i) = U_{ni}^3$, where Z is a random variable with the CDF F_i. Then, $Z = (1-\sqrt{5})U_{ni} / 2$ with probability $(1+\sqrt{5})/(2\sqrt{5})$, and $Z = (1+\sqrt{5})U_{ni} / 2$ with probability $1-(1+\sqrt{5})/(2\sqrt{5})$. The wild bootstrap is implemented as follows:

1. For each $i = 1, \ldots, n$, sample U_i^* randomly from F_i. Set $Y_i^* = X_i b_n + U_i^*$.

2. Estimate (5.1) by OLS using the bootstrap sample $\{Y_i^*, X_i: i = 1, \ldots, n\}$. Compute the resulting t statistic, T_n^*.

3. Obtain the empirical distribution of the wild-bootstrap version of T_n^* by repeating steps 1 and 2 many times. Obtain the wild-bootstrap critical value of T_n^* from the empirical distribution.

Horowitz (1997) reports the results of a Monte Carlo investigation of the ability of the bootstrap and wild bootstrap to reduce the distortions in the RP of a symmetrical, two-tailed t test that occur when asymptotic critical values are used. The bootstrap was implemented by resampling

(Y,X) pairs, and the wild bootstrap was implemented as described above. The experiments also investigate the RP of the t test when the HCCME is used with asymptotic critical values and when a jackknife version of the HCCME is used with asymptotic critical values (MacKinnon and White 1985). MacKinnon and White (1985) found through Monte Carlo experimentation that with the jackknife HCCME and asymptotic critical values, the t test had smaller distortions of RP than it did with several other versions of the HCCME.

The experiments use $n = 25$. X consists of an intercept and either 1 or 2 explanatory variables. In experiments in which X has an intercept and one explanatory variable, $\boldsymbol{b} = (1, 0)'$. In experiments in which X has an intercept and two explanatory variables, $\boldsymbol{b} = (1,0,1)'$. The hypothesis tested in all experiments is H_0: $\boldsymbol{b}_2 = 0$. The components of X were obtained by independent sampling from a mixture of normal distributions in which $N(0,1)$ was sampled with probability 0.9 and $N(2,9)$ was sampled with probability 0.1. The resulting distribution of X is skewed and leptokurtotic. Experiments were carried out using homoskedastic and heteroskedastic U's. When U was homoskedastic, it was sampled randomly from $N(0,1)$. When U was heteroskedastic, the U value corresponding to $X = x$ was sampled from $N(0,\Omega_x)$, where $\Omega_x = 1 + x^2$ or $\Omega_x = 1 + x_1^2 + x_2^2$, depending on whether X consists of 1 or 2 components in addition to an intercept. Ω_x is the covariance matrix of U corresponding to the random-coefficients model $Y = X\boldsymbol{b} + X\boldsymbol{d} + V$, where V and the components of $\boldsymbol{d}$ are independently distributed as $N(0,1)$. There were 1000 Monte Carlo replications in each experiment.

Table 2 shows the empirical RP's of nominal 0.05-level t tests of H_0. The differences between the empirical and nominal RP's using the HCCME and asymptotic critical values are very large. Using the jackknife version of the HCCME or critical values obtained from the bootstrap greatly reduces the differences between the empirical and nominal RP's, but the empirical RP's are still 2-3 times the nominal ones. With critical values obtained from the wild bootstrap, the differences between the empirical and nominal RP's are very small. In these experiments, the wild bootstrap essentially removes the distortions of RP that occur with asymptotic critical values.

5.3 *The t Test in a Box-Cox Regression Model*

The *t* statistic for testing a hypothesis about a slope coefficient in a linear regression model with a Box-Cox (1964) transformed dependent variable is not invariant to changes in the measurement units, or scale, of the dependent variable (Spitzer 1984). The numerical value of the *t* statistic and the finite-sample RP's of the *t* test with asymptotic critical values vary according to the measurement units or scale that is used. As a result, the finite-sample RP's of the *t* test with asymptotic critical values can be far from the nominal RP's. The bootstrap provides a better approximation to the finite-sample distribution and, therefore, better finite-sample critical values.

Horowitz (1997) reports the results of a Monte Carlo investigation of the finite-sample RP of a symmetrical *t* test of a hypothesis about a slope coefficient in a linear regression model with a Box-Cox transformed dependent variable. The model generating the data is

$$Y^{(l)} = b_0 + b_1 X + U \ ,$$

where $Y^{(l)}$ is the Box-Cox transformed value of the dependent variable Y, $U \sim N(0, s^2)$, $b_0 = 2$, $b_1 = 0$ and $s^2 = 0.0625$. X was sampled from $N(4,4)$ and was fixed in repeated samples. The hypothesis being tested is H_0: $b_1 = 0$. The value of l is either 0.01 or 1, depending on the experiment, and the scale of Y was 0.2, 1, or 5. The sample sizes were $n = 50$ and 100. There were 1000 replications in each experiment.

The results of the experiments are summarized in Table 3. The empirical critical value of the *t* test tends to be much smaller than the asymptotic critical value of 1.96, especially in the experiments with a scale factor of 5. As a result, the empirical RP of the *t* test is usually much smaller than its nominal RP. The mean bootstrap critical values, however, are very close to the empirical critical values, and the RP's based on bootstrap critical values are very close to the nominal ones.

5.5 *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 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 been found to have poor finite-sample properties 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. Altonji and Segal also found that the true coverage probabilities of asymptotic confidence intervals based on the OMD estimator tend to be much lower than the nominal coverage probabilities. Thus, estimation and inference based on the OMD estimator can be highly misleading with finite samples.

Horowitz (1998a) reports the results of a Monte Carlo investigation the ability of the bootstrap to reduce the bias and RMSE of the OMD estimator and reduce the differences between true and nominal coverage probabilities of nominal 95% confidence intervals based on this estimator. The data-generation processes used in the Monte Carlo experiments were taken from Altonji and Segal (1994). In each experiment, X has 10 components, and the sample size is $n =$

500. The j'th component of X, X_j (j = 1, …, 10) is generated by $X_j = (Z_j + rZ_{j+1})/(1 + r^2)^{1/2}$, where $Z_1, \ldots, Z_{11}$ are *iid* random variables with means of 0 and variances of 1, and $r = 0.5$. The Z's are sampled from five different distributions depending on the experiment. These are $U[0,1]$, $N(0,1)$, Student t with 10 degrees of freedom, exponential, and lognormal. It is assumed that r 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 q that is identified by the moment conditions $Var(X_j) = q$ (j = 1, …, 10) and $Cov(X_j, X_{j-1}) = rq/(1 + r^2)$ (j = 2, …, 10). Experiments were carried out with the EWMD and OMD estimators as well as a version of the OMD estimator that uses a trimmed estimator of the asymptotically optimal weight matrix. See Horowitz (1998a) for an explanation of the trimming procedure.

The results of the experiments are summarized in Table 4. The OMD estimator, $q_{n,OMD}$ is biased and its RMSE exceeds that of the EWMD estimator, $q_{n,EWMD}$ for all distributions of Z except the uniform. Moreover, the coverage probabilities of confidence intervals based on $q_{n\ OMD}$ with asymptotic critical values are far below the nominal value of 0.95 except in the experiment with uniform Z's. Bootstrap bias reduction greatly reduces both the bias and RMSE of $q_{n,OMD}$. In addition, the use of bootstrap critical values greatly reduces the errors in the coverage probabilities of confidence intervals based on $q_{n,OMD}$. In the experiments with normal, Student t, or uniform Z's, the bootstrap essentially eliminates the bias of $q_{n,OMD}$ and the errors in the coverage probabilities of the confidence intervals. Moreover, the RMSE of the bias-corrected $q_{n,OMD}$ in these experiments is 12-50% less than that of $q_{n,EWMD}$.

When Z is exponential or lognormal, the bootstrap reduces but does not eliminate the bias of $q_{n,OMD}$ and the errors in the coverage probabilities of confidence intervals. Horowitz (1998a) shows that the poor performance of the bootstrap in these cases is caused by imprecise estimation of the OMD weight and covariance matrices. This problem is largely eliminated through the use of the trimmed estimator of these matrices. With trimming, $q_{n,OMD}$ with exponential or lognormal

Z's has a RMSE that is the same as or less than that of the EWMD estimator, and the empirical coverage probabilities of confidence intervals are close to the nominal values.

6. CONCLUSIONS

The bootstrap consistently estimates the asymptotic distributions of econometric estimators and test statistics under conditions that are sufficiently general to accommodate most applications. Subsampling methods usually can be used in place of the standard bootstrap when the latter is not consistent. Together, the bootstrap and subsampling methods provide ways to substitute computation for mathematical analysis if analytical calculation of the asymptotic distribution of an estimator or test statistic is difficult or impossible.

Under conditions that are stronger than those required for consistency but still general enough to accommodate a wide variety of econometric applications, the bootstrap reduces the finite-sample biases of estimators and provides a better approximation to the finite-sample distribution of an estimator or test statistic than does first-order asymptotic theory. The approximations of first-order asymptotic theory are often quite inaccurate with samples of the sizes encountered in applications. As a result, the true and nominal probabilities that a test rejects a correct hypothesis can be very different when critical values based on first-order approximations are used. Similarly, the true and nominal coverage probabilities of confidence intervals based on asymptotic critical values can be very different. The bootstrap can provide dramatic reductions in the differences between true and nominal rejection and coverage probabilities of tests and confidence intervals. In many cases of practical importance, the bootstrap essentially eliminates finite-sample errors in rejection and coverage probabilities.

This chapter has also emphasized the need for care in applying the bootstrap. The importance of asymptotically pivotal statistics for obtaining asymptotic refinements has been stressed. Proper attention also must be given to matters such as recentering, correction of test statistics in the block bootstrap for dependent data, smoothing, and choosing the distribution from which bootstrap samples are drawn. These qualifications do not, however, detract from the importance of the bootstrap as a practical tool for improving inference in applied econometrics.

APPENDIX: Informal Derivation of (3.27)

To derive (3.27), write $P(|T_n| \geq z_{n,a/2}*)$ in the form

(A.1) $$P(|T_n| > z_{n,a/2}*) = 1 - [P(T_n \leq z_{n,a/2}*) - P(T_n \leq -z_{n,a/2}*)]$$

$$= 1 - \{P[T_n - (z_{n,a/2}* - z_{\infty.a/2}) \leq z_{\infty,a/2}] - P[T_n + (z_{n,a/2}* - z_{\infty,a/2}) \leq -z_{\infty,a/2}]\}.$$

With an error whose size is almost surely $O(n^{-2})$, $(z_{n,a/2}* - z_{\infty,a/2})$ on the right-hand side of (A.1) can be replaced with a Cornish-Fisher expansion that retains terms through $O(n^{-3/2})$. This expansion can be obtained by applying the delta method to the difference between (3.23) and (3.24). The result is

(A.2) $$z_{n,a/2}* - z_{\infty,a/2} = -\frac{1}{n}\frac{g_2(z_{\infty,a/2}, F_0)}{f(z_{\infty,a/2})} + \frac{1}{n^{3/2}} n^{1/2} r_3(\bar{Z}) + O(n^{-2}),$$

where r_3 is a smooth function, $r_3(m_Z) = 0$, and $n^{1/2} r_3(\bar{Z}) = O_p(1)$ as $n \to \infty$. Substituting (A.2) into (A.1) yields

(A.3) $$P(|T_n| > z_{n,a/2}*) = 1 - \{P[T_n - n^{-3/2} n^{1/2} r_3(\bar{Z}) \leq z_{\infty,a/2} + n^{-1} r_2(z_{\infty,a/2})]$$

$$- P[T_n + n^{-3/2} n^{1/2} r_3(\bar{Z}) \leq -z_{\infty,a/2} - n^{-1} r_2(z_{\infty,a/2})]\} + O(n^{-2}).$$

where

(A.4) $$r_2(z) = -\frac{g_2(z_{\infty,a/2}, F_0)}{f(z_{\infty,a/2})}.$$

The next step is to replace the right-hand side of (A.3) with an Edgeworth approximation. To do this, it is necessary to provide a detailed specification of the function g_2 in (3.9) and (3.13). Let $k_{j,n}$ denote the j'th cumulant of T_n.[23] Under assumption SFM, $k_{j,n}$ can be expanded in a power series. For a statistic such as T_n whose asymptotic distribution has a variance of 1,

$$k_{1,n} = \frac{k_{12}}{n^{1/2}} + \frac{k_{13}}{n^{3/2}} + O(n^{-5/2}),$$

$$k_{2,n} = 1 + \frac{k_{22}}{n} + O(n^{-2}),$$

$$k_{3,n} = \frac{k_{31}}{n^{1/2}} + \frac{k_{32}}{n^{3/2}} + O(n^{-5/2}),$$

and

$$k_{4,n} = \frac{k_{41}}{n} + O(n^{-2}),$$

where the coefficients k_{jk} are functions of moments of products of components of Z. The function g_2 is then

(A.5) $$g_2(t, F_0) = -t\left\{\frac{1}{2}(k_{22} + k_{12}^2) + \frac{1}{24}(k_{41} + 4k_{12}k_{31})(t^2 - 3) + \frac{1}{72}k_{31}^2(t^4 - 10t^2 + 15)\right\}f(t).$$

See Hall (1992a, pp. 46-56) for details. Denote the quantity on the right-hand side of (A.5) by $\tilde{g}_2(t, k_0)$, where k_0 denotes the k_{jk} coefficients that are associated with cumulants of the distribution of T_n. Let $\hat{k}_n$ denote the k_{jk} coefficients that are associated with cumulants of $T_n + n^{-3/2}n^{1/2}r_3(\bar{Z})$, and let $\tilde{g}_2(t, \hat{k}_n)$ denote the version of $\tilde{g}_2$ that is obtained by replacing k_0 with $\hat{k}_n$. Now replace $g_2(t, F_0)$ in (3.13) with $\tilde{g}_2(t, \hat{k}_n)$. Also, replace t with $z_{\infty,a/2} + n^{-1}r_2(z_{\infty,a/2})$ in (3.13). Substituting the result into the right-hand side of (A.3) gives the following Edgeworth approximation to $P(|T_n| > z_{n,a/2}{*})$:

(A.6) $$P(|T_n| > z_{n,a/2}{*}) = 2\{1 - \Phi[z_{\infty,a/2} + n^{-1}r_2(z_{\infty,a/2})]\}$$
$$-2n^{-1}\tilde{g}_2[z_{\infty,a/2} + n^{-1}r_2(z_{\infty,a/2}), \hat{k}_n] + O(n^{-2}).$$

A Taylor-series expansion of the right-hand side of (A.6) combined with (A.4) and the fact that $2[1 - \Phi(z_{\infty,a/2})] = a$ gives

(A.7) $$P(|T_n| > z_{n,a/2}{*}) = a + \frac{2}{n}[\tilde{g}_2(z_{\infty,a/2}, k_0) - \tilde{g}_2(z_{\infty,a/2}, \hat{k}_n)] + O(n^{-2}).$$

It is not difficult to show that $\tilde{g}_2(z_{\infty,a/2}, k_0) - \tilde{g}_2(z_{\infty,a/2}, \hat{k}_n) = o(n^{-1})$. (Roughly speaking, this is because $n^{-1} r_3(\bar{Z}) = o(n^{-1})$ almost surely.) Therefore, the second term on the right-hand side of (A.7) is $o(n^{-2})$, which yields (3.27).

FOOTNOTES

[1] There is not general agreement on the name that should be given to the probability that a test rejects a true null hypothesis (that is, the probability of a Type I error). The source of the problem is that if the null hypothesis is composite, then the rejection probability can be different for different probability distributions in the null. Hall (1992, p. 148) uses the word *level* to denote the rejection probability at the distribution that was, in fact, sampled. Beran (1988, p. 696) defines *level* to be the supremum of rejection probabilities over all distributions in the null hypothesis. Other authors (Lehmann 1959, p. 61; Rao 1973, p. 456) use the word *size* for the supremum. Lehmann defines *level* as a number that exceeds the rejection probability at all distributions in the null hypothesis. In this chapter, the term *rejection probability* or *RP* will be used to mean the probability that a test rejects a true null hypothesis with whatever distribution generated the data. The RP of a test is the same as Hall's definition of level. The RP is different from the size of a test and from Beran's and Lehmann's definitions of *level*.

[2] The Mallows metric is defined by $r(P,Q)^2 = \inf\{E\|Y - X\|^2 : Y \sim P,\ X \sim Q\}$. The infimum is over all joint distributions of (Y, X) whose marginals are *P* and *Q*. . Weak convergence of a sequence of distributions in the Mallows metric implies convergence of the corresponding sequences of first and second moments. See Bickel and Freedman (1981) for a detailed discussion of this metric.

[3] Hall and Jing (1996) show how certain types of asymptotic refinements can be obtained through non-replacement subsampling. The rate of convergence of resulting error is, however, slower than the rate achieved with the standard bootstrap.

[4] If $E(q_n)$ does not exist, then the "bias reduction" procedure described here centers a higher-order approximation to the distribution of q_n - q_0.

[5] It is not difficult to show that the bootstrap provides bias reduction even if $m = 1$. However, the bias-corrected estimator of q may have a large variance if m is too small. The asymptotic distribution of the bias-corrected estimator is the same as that of the uncorrected estimator if m increases sufficiently rapidly as n increases. See Brown (1996) for further discussion.

[6] The meaning of asymptotic negligibility in this context may be stated precisely as follows. Let

$\tilde{T}_n = \tilde{T}_n(X_1,...,X_n)$ be a statistic, and let $T_n = n^{1/2}[H(\bar{Z}_1,...,\bar{Z}_J) - H(m_{Z1},...,m_{ZJ})]$. Then the error made by approximating $\tilde{T}_n$ with T_n is asymptotically negligible if there is a constant $c > 0$ such that $n^2 P[n^2 |\tilde{T}_n - T_n| > c] = O(1)$ as $n \to \infty$.

[7] The proof that the bootstrap provides asymptotic refinements is based on an Edgeworth expansion of a sufficiently high-order Taylor-series approximation to T_n. Assumption SFM insures that H has derivatives and Z has moments of sufficiently high order to obtain the Taylor series and Edgeworth expansions that are used to obtain a bootstrap approximation to the distribution of T_n that has an error of size $O(n^{-2})$. . SFM may not be the weakest condition needed to obtain this result. It certainly assumes the existence of more derivatives of H and moments of Z than needed to obtain less accurate approximations. For example, asymptotic normality of T_n can be proved if H has only one continuous derivative and Z has only two moments. See Hall (1992a, pp. 52-56 and 238-259) for a statement of the regularity conditions needed to obtain various levels of asymptotic and bootstrap approximations.

[8] Some statistics that are important in econometrics have asymptotic chi-square distributions. Such statistics often satisfy the assumptions of the smooth function model but with $\partial H(m_Z) = 0$ and $\partial^2 H(z)/\partial z \partial z' \big|_{z=m_Z} \neq 0$. Versions of the results described here for asymptotically normal statistics are also available for asymptotic chi-square statistics. First-order asymptotic approximations to the finite-sample distributions of asymptotic chi-square statistics typically make errors of size $O(n^{-1})$. Chandra and Ghosh (1979) give a formal presentation of higher-order asymptotic theory for asymptotic chi-square statistics.

[9] More generally, (3.8) is satisfied if the distribution of Z has a non-degenerate absolutely continuous component in the sense of the Lebesgue decomposition. There are also circumstances in which (3.8) is satisfied even when the distribution of Z does not have a non-degenerate absolutely continuous component. See Hall (1992a, pp. 66-67) for examples. In addition, (3.8) can be modified to deal with econometric models that have a continuously distributed dependent variable but discrete covariates. See Hall (1992a, p. 266).

[10] Another form of two-tailed test is the equal-tailed test. An equal tailed test rejects H_0 if $T_n > z_{n,\alpha/2}$ or $T_n < z_{n,(1-\alpha/2)}$, where $z_{n,(1-\alpha/2)}$ is the $\alpha/2$-quantile of the finite-sample distribution of T_n. If the distribution of T_n is symmetrical about 0, then equal-tailed and symmetrical tests are the same.

Otherwise, they are different. Most test statistics used in econometrics have symmetrical asymptotic distributions, so the distinction between equal-tailed and symmetrical tests is not relevant when the RP is obtained from first-order asymptotic theory. Many test statistics have asymmetrical finite-sample distributions, however. Higher-order approximations to these distributions, such as the approximation provided by the bootstrap, are also asymmetrical. Therefore, the distinction between equal-tailed and symmetrical tests is important in the analysis of asymptotic refinements. Note that "symmetrical" in a symmetrical test refers to the way in which the critical value is obtained, not to the finite-sample distribution of T_n, which is asymmetrical in general.

[11] The empirical distribution of the data is discrete, so (3.20) may not have a solution if F_n is the EDF of the data. However, Hall (1992a, pp. 283-286) shows that there is a solution at a point α_n whose difference from α decreases exponentially fast as $n \to \infty$. The error introduced into the analysis by ignoring the difference between α_n and α is $o(n^{-2})$ and, therefore, negligible for purposes of the discussion in this chapter.

[12] Under mild regularity conditions, the constant the multiplies the rate of convergence of the error of the bootstrap estimate of the distribution function of a non-asymptotically-pivotal statistic is smaller than the constant that multiplies the rate of convergence of the error that is made by the normal approximation. This need not happen, however, with the errors in the RP's of tests and coverage probabilities of confidence intervals. See Beran (1982) and Liu and Singh (1985).

[13.] Strictly speaking, U cannot be normally distributed unless l = 0 or 1, but the error made by assuming normality is negligibly small if the right-hand side of the model has a negligibly small probability of being negative. Amemiya and Powell (1981) discuss ways to avoid assuming normality.

[14] The empirical-likelihood estimator is one of a larger class of estimators of F that are described by Brown *et al*. (1997) and that impose the restriction $E^*h(X, q_n) = 0$. All estimators in the class are asymptotically efficient.

[15] The regularity conditions required to achieve asymptotic refinements in GMM estimation with dependent data include the existence of considerably more higher-order moments than are needed with *iid* data as well as a modified version of the Cramér condition that takes account of the dependence. See Hall and Horowitz (1996) for a precise statement of the conditions.

[16] Tests and confidence regions based on asymptotic chi-square statistics, including the test of overidentifying restrictions, are symmetrical. Therefore, restriction (4.2) also applies to them.

[17] The results stated in this section do not require assuming that r is even or that K is a symmetrical function, but these assumptions simplify the exposition and are not highly restrictive in applications.

[18] The asymptotic bias contributes a term of size $[(nh_n)^{1/2}b_n(x)]^2 = O(nh_n^{2r+1})$ to the Edgeworth expansion of the distribution of $|t_n|$. Because t_n^* is unbiased, this term is not present in the expansion of the distribution of $|t_n^*|$. Therefore, the expansions of the distributions of $|t_n|$ and $|t_n^*|$ agree through $O[(nh_n)^{-1}]$ only if $nh_n^{r+1} \rightarrow 0$ as $n \rightarrow \infty$.

[19] It is also possible to carry out explicit bias removal in kernel mean-regression. Härdle *et al* (1995) compare the methods of explicit bias removal and undersmoothing for a one-sided confidence interval. They show that for a one-sided interval, there are versions of the bootstrap and explicit bias removal that give better coverage accuracy than the bootstrap with undersmoothing.

[20] Hall (1992a, p. 226) proposes an estimator of $s_n^2(x)$ that is $n^{1/2}$-consistent when Y is homoskedastic (that is, $Var(Y|X = x)$ is independent of x). The estimator used here is consistent (but not $n^{1/2}$-consistent) when Y has heteroskedasticity of unknown form.

[21] The discussion here assumes that the bootstrap sample is obtained by randomly sampling the empirical distribution of (Y, X). If $V(z)$ is a constant (that is, the model is homoskedastic), then bootstrap sampling can also be carried out by sampling centered regression residuals conditional on the observed values of X. See Hall (1992a, Section 4.5).

[22] Janas (1993) shows that a smoothed version of the bootstrap provides asymptotic refinements for a symmetrical t test of a hypothesis about a population median (no covariates).

[23] The cumulants of a distribution are coefficients in a power-series expansion of the logarithm of its characteristic function. The first three cumulants are the mean, variance, and third moment about the mean. The fourth cumulant is the fourth moment about the mean minus three times the square of the variance.

TABLE 1

EMPIRICAL REJECTION PROBABILITIES OF NOMINAL 0.05-LEVEL INFORMATION-MATRIX TESTS OF PROBIT AND TOBIT MODELS[1]

N	Distr. of X	RP Using Asymptotic Critical Values: White	Chesh.-Lan.	Orme	RP Using Bootstrap-Based Crit. Values: White	Chesh.-Lan.	Orme
				Binary Probit Models			
50	N(0,1)	0.385	0.904	0.006	0.064	0.056	0.033
	U(-2,2)	0.498	0.920	0.017	0.066	0.036	0.031
100	N(0,1)	0.589	0.848	0.007	0.053	0.059	0.054
	U(-2,2)	0.632	0.875	0.027	0.058	0.056	0.049
				Tobit Models			
50	N(0,1)	0.112	0.575	0.038	0.083	0.047	0.045
	U(-2,2)	0.128	0.737	0.174	0.051	0.059	0.054
100	N(0,1)	0.065	0.470	0.167	0.038	0.039	0.047
	U(-2,2)	0.090	0.501	0.163	0.046	0.052	0.039

[1] Source: Horowitz (1994).

TABLE 2

EMPIRICAL REJECTION PROBABILITIES OF t TESTS USING HETEROSKEDASTICITY-CONSISTENT COVARIANCE MATRIX ESTIMATORS[1]

n = 25

Form of Test	Empirical RP at Nominal 0.05 Level: 1-Variable Homoskedastic Model	1-Variable Random Coeff. Model	2-Variable Homoskedastic Model	2-Variable Random Coeff. Model
Asymptotic	0.156	0.306	0.192	0.441
Jackknife	0.096	0.140	0.081	0.186
Bootstrap (Y,X) Pairs	0.100	0.103	0.114	0.124
Wild Bootstrap	0.050	0.034	0.062	0.057

[1] Source: Horowitz (1997).

TABLE 3[1]

EMPIRICAL REJECTION PROBABILITIES OF t TESTS FOR BOX-COX REGRESSION MODEL[1]

Nominal RP = 0.05

n	λ	Scale Fac.	RP Using Crit. Val. from Asymp.	RP Using Crit. Val. from Boot.	Empirical Crit. Val.	Mean Bootstrap Crit. Val.
50	0.01	0.2	0.048	0.066	1.930	1.860
		1.0	0.000	0.044	0.911	0.909
		5.0	0.000	0.055	0.587	0.571
100	0.01	0.2	0.047	0.053	1.913	1.894
		1.0	0.000	0.070	1.201	1.165
		5.0	0.000	0.056	0.767	0.759
50	1.0	0.2	0.000	0.057	1.132	1.103
		1.0	0.000	0.037	0.625	0.633
		5.0	0.000	0.036	0.289	0.287
100	1.0	0.2	0.000	0.051	1.364	1.357
		1.0	0.000	0.044	0.836	0.835
		5.0	0.000	0.039	0.401	0.391

[1] Source: Horowitz (1997).

TABLE 4: RESULTS OF MONTE CARLO EXPERIMENTS WITH ESTIMATORS OF COVARIANCE STRUCTURES[1]

	EWMD	OMD without Bootstrap			OMD with Bootstrap			Trimmed OMD with Bootstrap		
Distr.	RMSE	Bias	RMSE	Coverage Prob. with Asymptotic Critical Value	Bias	RMSE	Coverage Prob. with Bootstrap Critical Value	Bias	RMSE	Coverage Prob. with Bootstrap Critical Value
Uniform	0.019	0.005	0.015	0.93	0.002	0.014	0.96			
Normal	0.024	0.016	0.025	0.85	0.0	0.021	0.95			
Student t	0.029	0.024	0.034	0.79	0.002	0.026	0.95			
Exponential	0.042	0.061	0.073	0.54	0.014	0.048	0.91	0.004	0.042	0.96
Lognormal	0.138	0.274	0.285	0.03	0.136	0.173	0.76	0.046	0.126	0.91

[1] Source: Horowitz (1998a). Nominal coverage probability is 0.95. Based on 1000 replications.

REFERENCES

Abowd, J.M and D. Card (1987). Intertemporal labor supply and long-term employment contracts, *American Economic Review*, 77, 50-68.

Abowd, J.M. and D. Card (1989). On the covariance of earnings and hours changes, *Econometrica*, 57, 411-445.

Altonji, J.G. and L.M. Segal (1994). Small sample bias in GMM estimation of covariance structures, NBER Technical Working Paper no. 156, National Bureau of Economic Research, Cambridge, MA.

Altonji, J.G. and L.M. Segal (1996). Small sample bias in GMM estimation of covariance structures, *Journal of Business and Economic Statistics*, 14, 353-366.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Amemiya, T. and J.L. Powell (1981). A comparison of the Box-Cox maximum likelihood estimator and the non-linear two-stage least squares estimator, *Journal of Econometrics*, 17, 351-381.

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59. 817-858.

Andrews, D.W.K. (1997a). A conditional Kolmogorov test, *Econometrica*, 65, 1097-1128.

Andrews, D.W.K. (1997b). A simple counterexample to the bootstrap, Cowles Foundation discussion paper no. 1157, Cowles Foundation for Research in Economics, Yale University, New Haven, CT.

Andrews, D.W.K. (1998). A computaionally attractive k-step bootstrap procedure for extremum estimators, unpublished manuscript, Cowles Foundation for Research in Economics, Yale University, New Haven, CT.

Andrews, D.W.K. and M. Buchinsky (1997). On the Number of Bootstrap Repetitions for Bootstrap Standard Errors, Confidence Intervals, Confidence Regions, and Tests, working paper, Department of Economics, Brown University, Providence, RI.

Andrews, D.W.K. and J.C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 59, 817-858.

Athreya, K. (1987). Bootstrap of the mean in the infinite variance case, *Annals of Statistics*, 15, 724-731.

Babu, G.J. and K. Singh (1983). Inference on means using the bootstrap, *Annals of Statistics*, 11, 999-1003.

Babu, G.J. and K. Singh (1984). On one term correction by Efron's bootstrap, *Sankhya*, *Series A*, 46, 219-232.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991a). Bootstrapping unstable first-order autoregressive processes, *Annals of Statistics*, 19, 1098-1101.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991b). Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes, *Communications in Statistics - Theory and Methods*, 20, 1015-1026.

Bassett, G. and R. Koenker (1978). Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association*, 73, 618-621.

Behrman, J., M. Rosenzweig, and P. Taubman (1994). Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment, *Journal of Political Economy*, 102, 1131-1174.

Beran, R. (1982). Estimated sampling distributions: The bootstrap and competitors, *Annals of Statistics*, 10, 212-225.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets, *Biometrika*, 74, 457-468.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of aymptotic refinements, *Journal of the American Statistical Association*, 83, 687-697.

Beran, R. and G.R. Ducharme (1991). *Asymptotic Theory for Bootstrap Methods in Statistics*, Les Publications CRM, Centre de recherches mathematiques, Universite de Montreal, Montreal, Canada.

Bertail, P., D.N. Politis, and J.P. Romano (1995). On subsampling estimators with unknown rate of convergence, Technical report 95-20, Department of Statistics, Purdue University, West Lafayette, IN.

Bickel, P.J. and D.A. Freedman (1981). Some asymptotic theory for the bootstrap, *Annals of Statistics*, 9, 1196-1217.

Bickel, P..J., F. Götze, and W.R. van Zwet (1997). Resampling fewer than *n* observations: gains, losses, and remedies for losses, *Statistica Sinica*, 7, 1-32.

Blanchard, O.J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review*, 79, 655-673.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions, *Annals of Statistics*, 16, 1709-1722.

Bose, A. (1990). Bootstra in moving average models, *Annals of the Institute of Statistical Mathematics*, 42, 753-768.

Box, G.E.P. and D.R. Cox (1964). An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, 211-243.

Breusch, T. and P. Schmidt (1988). Alternative forms of the Wald test: how long is a piece of string? *Communications in Statistics A, Theory and Methods*, 17, 2789-2795.

Brown, B.W. (1999). Simulation variance reduction for bootstrapping, in R. Mariano, T. Schuermann, and M. Weeks, eds., *Simulation-Based Econometrics: Methods and Applications*, New York: Cambridge University Press, forthcoming.

Brown, B., W.K. Newey, and S. May (1997). Efficient bootstrapping for GMM. Unpublished manuscript, Department of Economics, Massachusetts Institute of Technology.

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

Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series, *Annals of Statistics*, 26, 48-83.

Cao-Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression, *Annals of Statistics*, 19, 2226-2231.

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.

Cavanagh, C.L. (1987). Limiting behavior of estimators defined by optimization, unpublished manuscript, Department of Economics, Harvard University.

Chandra, T.K. and J.K. Ghosh (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables, *Sankhya*, 41, Series A, 22-47.

Chesher, A. (1983). The information matrix test, *Economics Letters*, 13, 45-48.

Chesher, A. and I. Jewitt (1987). The bias of a heteroskedasticity consistent covariance matrix estimator, *Econometrica*, 55, 1217-1222.

Dagenais, M.G. and J.-M. Dufour (1991). Invariance, nonlinear models, and asymptotic tests, *Econometrica*, 59, 1601-1615.

Datta, S. (1995). On a modified bootstrap for certain asymptotically non-normal statistics, *Statistics and Probability Letters*, 24, 91-98.

Datta, S. (1996). On asymptotic properties of bootstrap for AR(1) processes, *Journal of Statistical Planning and Inference*, 53, 361-374.

Davidson, R. and J.G. MacKinnon (1997a). Bootstrap testing in nonlinear models, *International Economic Review*, forthcoming.

Davidson, R. and J.G. MacKinnon (1997b). The size distortion of bootstrap tests, unpublished working paper, Department of Economics, Queen's University, Kingston, Ontario.

Davison, A.C. and D.V. Hinkley (1997). *Bootstrap Methods and Their Application*, Cambridge, U.K.: Cambridge University Press.

De Angelis, D., P. Hall, and G.A. Young (1993). Analytical and bootstrap approximations to estimator distributions in L^1 regression, *Journal of the American Statistical Association*, 88, 1310-1316.

Donald, S.G. and H.J. Paarsch (1996). Identification, estimation, and testing in empirical models of auctions within the independent private values paradigm, *Econometric Theory*, 12, 517-567.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics*, 7, 1-26.

Efron, B. (1987). Better bootstrap confidence intervals, *Journal of the American Statistical Association*, 82, 171-185.

Efron, B. and R.J. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions, *Annals of Mathematical Statistics*, 34, 447-456.

Eicker, F. (1967). Limit theorems for regression with unequal and dependent errors, in L. LeCam and J. Neyman (eds.), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press, pp. 59-82.

Ferretti, N. and J. Romo (1996). Unit root bootstrap tests for AR(1) models, *Biometrika*, 83, 849-860.

Flinn, C. and J. Heckman (1982). New methods for analyzing structural models of labor force dynamics, *Journal of Econometrics*, 18, 115-168.

Freedman, D.A. (1981). Bootstrapping regression models, *Annals of Statistics*, 9, 1218-1228.

Götze, F. and C. Hipp (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Warscheinlichkeitstheorie 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.

Gregory, A.W. and M.R. Veall (1985). Formulating Wald tests of nonlinear restrictions, *Econometrica*, 53, 1465-1468.

Griliches, Z. (1979). Sibling models and data in economics: beginnings of a survey, *Journal of Political Economy*, 87, S37-S64.

Hahn, J. (1995). Bootstrapping the quantile regression estimators, *Econometric Theory*, 11, 105-121.

Hahn, J. (1996). A note on bootstrapping generalized method of moments estimators, *Econometric Theory*, 12, 187-197.

Hall, P. (1985). Resampling a coverage process, *Stochastic Process Applications*, 19, 259-269.

Hall, P. (1986a). On the number of bootstrap simulations required to construct a confidence interval, *Annals of Statistics*, 14, 1453-1462.

Hall, P. (1986b). On the bootstrap and confidence intervals, *Annals of Statistics*, 14, 1431-1452.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals, *Annals of Statistics*, 16, 927-953.

Hall, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions, *Annals of Probability*, 18, 1342-1360.

Hall, P. (1992a). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P. (1992b). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Annals of Statistics*, 20, 675-694.

Hall, P. (1994). Methodology and theory for the bootstrap, in R.F. Engle and D.F. McFadden, eds., *Handbook of Econometrics*, vol. 4, Ch. 39. Amsterdam: Elsevier Science, B.V.

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.

Hall, R.E. and F. Mishkin (1982). The sensitivity of consumption to transitive income: estimates from panel data on households, *Econometrica*, 50, 461-481.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029-1054.

Hansen, L.P. and K. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica*, 50, 1269-1286.

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge, U.K.: Cambridge University Press.

Härdle, W. and J.D. Hart (1992). A bootstrap test for positive definiteness of income effect matrices, *Econometric Theory*, 8, 276-290.

Härdle, W., W. Hildenbrand, and M. Jerison (1991). Empirical evidence on the law of demand, *Econometrica*, 59, 1525-1550.

Härdle, W., S. Huet, and E. Jolivet (1995). Better bootstrap confidence intervals for regression curve estimation, *Statistics*, 26, 287-306.

Härdle, W. and J.S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression, *Annals of Statistics*, 19, 778-796.

Härdle, W. and E. Mammen (1993). Comparing nonparametric versus parametric regression fits, *Annals of Statistics*, 21, 1926-1947.

Heckman, J.J, J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts, *Review of Economic Studies*, 64, 487-535.

Hillier, G.H. (1985). On the joint and marginal densities of instrumental variables estimators in a general structural equation, *Econometric Theory*, 1, 53-72.

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

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

Horowitz, J.L. (1997). Bootstrap methods in econometrics: theory and numerical performance, in D.M. Kreps and K.F. Wallis, eds. *Advances in Economics and Econometrics: Theory and Applications*, Seventh World Congress, vol. 3, Ch. 7. Cambridge, U.K.: Cambridge University Press.

Horowitz, J.L. (1998a): Bootstrap methods for covariance structures, *Journal of Human Resources*, 33, 39-61.

Horowitz, J.L. (1998b). Bootstrap methods for median regression models, *Econometrica*, 66, 1327-1351.

Horowitz, J.L. (1998c). Bootstrap critical values for tests based on the smoothed maximum score estimator, *Journal of Econometrics*, forthcoming.

Horowitz, J.L. and N.E. Savin (1992). Noninvariance of the Wald test: the bootstrap to the rescue, working paper no. 92-04, Department of Economics, University of Iowa.

Janas, D. (1993). A smoothed bootstrap estimator for a studentized sample quantile, *Annals of the Institute of Statistical Mathematics*, 45, 317-329.

Kennan, J. and G.R. Neumann (1988). Why does the information matrix test reject so often? Working paper no. 88-4, Department of Economics, University of Iowa.

Kim, J. and D. Pollard (1990). Cube root asymptotics, *Annals of Statistics*, 18, 191-219.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes, *Annals of Statistics*, 25, 2084-2102.

Koenker, R. and G. Bassett (1978). Regression quantiles, *Econometrica*, 46, 33-50.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, 17, 1217-1241.

Lafontaine, F. and K.J. White (1986). Obtaining any Wald statistic you want, *Economics Letters*, 21, 35-48.

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*. New York: Wiley, pp. 183-214.

Lahiri, S. (1996). 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. (1997). Theoretical comparisons of block bootstrap methods, working paper, Department of Statistics, Iowa State University, Ames, IA.

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

Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley & Sons.

Li, H. and G.S. Maddala (1996). Bootstrapping time series models, *Econometric Reviews*, 15, 115-158.

Li, H. and G.S. Maddala (1997). Bootstrapping cointegrating regressions, *Journal of Econometrics*, 80, 297-318.

Liu, R.Y. (1988). Bootstrap procedures under some non-i.i.d. models, *Annals of Statistics*, 16, 1696-1708.

Liu, R.Y. and K. Singh (1987). On a partial correction by the bootstrap, *Annals of Statistics*, 15, 1713-1718.

MacKinnon, J.G. and H. White (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, *Journal of Econometrics*, 29, 305-325.

Maddala, G.S. and J. Jeong (1993). A perspective on application of bootstrap methods in econometrics, in G.S. Maddala, C.R. Rao, and H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11. Amsterdam: North-Holland, pp. 573-610.

Mammen, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations*. New York: Springer-Verlag.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics*, 21, 255-285.

Manski, Charles F. (1975). Maximum score estimation of the stochastic utility model of choice, *Journal of Econometrics* 3, 205-228.

Manski, Charles F.. (1985). Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator, *Journal of Econometrics* 27, 313-334.

Manski, C.F. and T.S. Thompson (1986). Operational characteristics of maximum score estimation, *Journal of Econometrics*, 32, 85-108.

Nankervis, J.C. and N.E. Savin (1996). The level and power of the bootstrap t-test in the trend model with AR(1) errors, *Journal of Business and Economic Statistics*, 14, 161-168.

Nelson, C.R. and R. Startz (1990a). The distribution of the instrumental variable estimator and its t ratio when the instrument is a poor one, *Journal of Business*, 63, S125-S140.

Nelson, C.R. and R. Startz (1990b). Some further results on the exact small sample properties of the instrumental variable estimator, *Econometrica*, 58, 967-976.

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.

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

Phillips, P.C.B. (1983). Exact small sample theory in the simultaneous equations model, in *Handbook of Econometrics, Vol. 1*, Z. Griliches and M.D. Intriligator (eds.), Amsterdam: North-Holland Publishing Co., Ch. 8.

Phillips, P.C.B. and J.Y. Park (1988). On the formulation of Wald tests of nonlinear restrictions, *Econometrica*, 56, 1065-1083.

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., J.P.Romano, and M. Wolf (1997). Subsampling for heteroskedastic time series, *Journal of Econometrics*, 81, 281-317.

Powell, J.L. (1984). Least absolute deviations estimation for the censored regression model, *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986). Censored regression quantiles, *Journal of Econometrics*, 32, 143-155.

Runkle, D.E. (1987). Vector autoregressions and reality, *Journal of Business and Economic Statistics*, 5, 437-442.

Shao, U. and D. Tu (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

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

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *Annals of Statistics*, 9, 1187-1195.

Spitzer, J.J. (1984). Variance estimates in models with the Box-Cox transformation: implications for estimation and hypothesis testing, *Review of Economics and Statistics*, 66, 645-652.

Stute, W. (1997). Nonparametric model checks for regression, *Annals of Statistics*, 25, 613-641.

Stute, W., W. Gonzalez-Manteiga, and M. Presedo Quindimil (1998). Bootstrap approximations in model checks for regression, *Journal of the American Statistical Association*, 93, 141-149.

Swanepoel, J.W.H. (1986). A note on proving that the (modified) bootstrap works, *Communications in Statistics Theory and Methods*, 15, 3193-3203.

Taylor, L.W. (1987). The size bias of White's information matrix test, *Economics Letters*, 24, 63-67.

Vinod, H.D. (1993). Bootstrap methods: applications in econometrics, in G.S. Maddala, C.R. Rao, and H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11. Amsterdam: North-Holland, pp. 629-661.

West, K.D. (1990). The sources of fluctuations in aggregate inventories and GNP, *Quarterly Journal of Economics*, 105, 939-971.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 48, 817-838.

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

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics*, 14, 1261-1295.