BOOTSTRAP METHODS IN ECONOMETRICS: THEORY AND NUMERICAL PERFORMANCE

by

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

November 1995

Prepared for presentation to the Invited Symposium on Bootstrap Methods in Econometrics, 7th World Congress of the Econometric Society, Tokyo, August 1995. I thank Narayana Kocherlakota, Charles F. Manski, John Nankervis, N.E. Savin, and Kenneth D. West for helpful comments and suggestions. Research supported in part by NSF grant SBR-9307677.

BOOTSTRAP METHODS IN ECONOMETRICS: THEORY AND NUMERICAL PERFORMANCE

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 maximum score estimator Manski (1975, 1985), the statistic developed by Härdle *et al.* (1991) for testing positive-definiteness of income-effect matrices, 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.[1]

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. First-order asymptotic theory often gives a poor approximation to the distributions of test statistics with the sample sizes available in applications. As a result, the nominal levels of tests based on asymptotic critical values can be very different from the true levels. The information matrix test of White (1982) is a well-known example of a test in which large finite-sample distortions of level 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 distortions of the levels of statistical tests.

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), Efron and Tibshirani (1993), Hall (1992) and Mammen (1992). Maddala and Jeong (1993) and Vinod (1993) provide reviews with an econometric orientation. Nonetheless, the bootstrap's ability to improve upon the approximations of first-order asymptotic theory appears not to be widely appreciated or used in econometrics.

The purpose of this chapter is to explain and illustrate the usefulness and limitations of the bootstrap for improving upon first-order asymptotic approximations in contexts of interest in econometrics. The discussion is informal and expository. I hope it will give readers a feeling for the practical value of the bootstrap in econometrics. Mathematically rigorous treatments of the theory of the bootstrap are available in the books by Beran and Ducharme (1991) and Hall (1992) as well as in journal articles that are cited later in this chapter.

The discussion concentrates on the use of the bootstrap to obtain improved finite-sample critical values for test statistics. Particular emphasis is placed on the importance of applying the bootstrap to statistics whose asymptotic distributions are independent of unknown population parameters. Such statistics are called "asymptotically pivotal." Simple bootstrap procedures provide improved approximations to the distributions of asymptotically pivotal statistics but not to the distributions of statistics that lack this property.

The problem of obtaining critical values for test statistics is closely related to that of obtaining confidence intervals. As is discussed in more detail in Section 2c, many of the methods that are described here for obtaining critical values of test statistics can also be used to obtain confidence intervals with improved finite-sample coverage probabilities.

It is not possible in a single chapter to provide a thorough treatment of all aspects of the bootstrap that may be useful in econometrics. Accordingly, the discussion in this chapter is selective. It focusses on methods that I believe are especially likely to be useful to applied researchers. Topics that are not discussed but that may be useful in some settings include bootstrap iteration and prepivoting, bias-correction methods, and bootstrap methods for semi- and nonparametric estimators that converge at slower than root-n rates. Discussion of the first two topics can be found in Beran and Ducharme (1991), Efron and Tibshirani (1993), and Hall (1992). Hall (1992) also discusses the theory of the bootstrap for kernel nonparametric density estimation and regression. The theory of the bootstrap for many semiparametric estimators of interest in econometrics (e.g., single-index models, sample-selection models, partially linear models) is still largely undeveloped.

It should be borne in mind throughout this chapter that although the bootstrap often provides better finite-sample critical values for test statistics than does first-order asymptotic theory, bootstrap critical values are still approximations and are 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 level right as a factor in selecting a test statistic. Among other benefits, this enables one to direct attention to choosing a statistic that has high power against alternatives of interest. This chapter does not treat the problem of choosing statistics to maximize power in applications. It does, however, provide some numerical examples illustrating that achieving high power and getting the level right are different problems. Tests whose finite-sample distortions of level are small when asymptotic critical values are used may have lower power than tests whose level distortions with asymptotic critical values are large but removable by the bootstrap. Thus, in terms of power, one may do better by using a statistic with large finite-sample level distortions than one with small level distortions if the distortions of the former statistic can be made small through the use of bootstrap critical values.

The remainder of this chapter is divided into three sections. Section 2 presents the theory of the bootstrap, Section 3 presents Monte Carlo evidence on the numerical performance of the bootstrap in a variety of settings that are relevant to econometrics, and Section 4 presents concluding comments.

2. THE THEORY OF THE BOOTSTRAP

This section explains why the bootstrap provides an improved approximation to the distributions of asymptotically pivotal test statistics. It also discusses the performance of the bootstrap when the hypothesis being tested is false. In addition, several special problems are discussed. These include bootstrapping overidentified estimators derived from moment conditions and the use of the bootstrap with dependent data.

a. Why the Bootstrap Provides Asymptotic Refinements

This section gives a heuristic explanation of why the bootstrap provides an improved approximation to the finite-sample distributions of test statistics. Let the data be a random sample of size n from a probability distribution whose cumulative distribution function (CDF) is F. Denote the data by $\{X_i: i = 1,\ldots,n\}$. F may belong to a finite- or infinite-dimensional family of distribution functions. If F belongs to a finite-dimensional family indexed by the parameter θ whose population value is θ_0, write $F(x,\theta_0)$ for $P(X \leq x)$ and $F(x,\theta)$ for a general member of the parametric family. The empirical distribution function (EDF) based on the sample is denoted by F_n. F_n will also denote a generic estimator of F when this can be done without confusion.

Let $T_n = T_n(X_1,\ldots,X_n)$ be a statistic for testing a hypothesis H_0 about the distribution from which $\{X_i\}$ is drawn. Let $G_n(z,F) \equiv P(T_n \leq z)$ denote the exact, finite-sample CDF of T_n when H_0 is true. Consider a symmetrical, two-tailed test of H_0. With such a test, H_0 is rejected at the α level if $|T_n| > z_{n\alpha}$, where the critical value, $z_{n\alpha}$, satisfies $G_n(z_{n\alpha},F) - G_n(-z_{n\alpha},F) = 1 - \alpha$. Usually, $z_{n\alpha}$ cannot be evaluated in applications because F is unknown. An exception occurs if G_n 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, 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). Thus, it is usually necessary to find an approximation for $z_{n\alpha}$.

First-order asymptotic theory provides one approximation. Most test statistics used in econometrics are asymptotically pivotal; their asymptotic distributions do not depend on unknown population parameters. If n is sufficiently large, one can approximate $G_n(\bullet,F)$ by the asymptotic distribution of T_n. This does not depend on F if T_n is asymtptotically pivotal. Therefore, approximate critical values for T_n can be obtained from the asymptotic distribution without having to know F.

Another possibility is to replace the unknown F in $G_n(\bullet,F)$ with a consistent estimator F_n. $G_n(\bullet,F)$ is then approximated by $G_n(\bullet,F_n)$. If F belongs to a known finite-dimensional parametric family $F(x,\theta)$, one can set $F_n(x) = F(x,\theta_n)$, where θ_n is a consistent estimator of θ. Otherwise, one can use the EDF

$$F(x) = n^{-1} \sum^{n} I(X \leq x),$$

$n \qquad {}_{i=1} \; {}^{i}$

where I(•) is the indicator function. With the approximation $G_n(\bullet,F_n)$, the approximate α-level critical value for $|T_n|$, $z_{n\alpha}^*$, solves $G_n(z_{n\alpha}^*,F_n) - G_n(-z_{n\alpha}^*,F_n) = 1 - \alpha$.

The bootstrap consists of approximating $G_n(\bullet,F)$ with $G_n(\bullet,F_n)$. Usually, $G_n(\bullet,F_n)$ and $z_{n\alpha}^*$ cannot be evaluated analytically. They 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 essential characteristic of the bootstrap, though, is the use of F_n to approximate F in $G_n(\bullet,F)$, not the method that is used to evaluate $G_n(\bullet,F_n)$.

Under mild regularity conditions, $\sup_x |F_n(x) - F(x)|$ and $\sup_z |G_n(z,F_n) - G_n(z,F)|$ converge to 0 in probability or almost surely. This guarantees that the bootstrap provides a good approximation to $G_n(z,F)$ and $z_{n\alpha}$ if n is sufficiently large. Of course, first-order asymptotic theory also provides good approximations to these quantities if n is sufficiently large. It turns out, however, that under conditions that are explained below, the bootstrap provides approximations that are more accurate than those of first-order asymptotic theory.

To see why, it is necessary to develop a higher-order approximation to $G_n(z,F)$. Under regularity conditions, $G_n(z,F)$ has an asymptotic expansion of the form

$$G_n(z,F) = G(z,F) + n^{-1/2} g_1(z,F) + n^{-1} g_2(z,F) + o(n^{-1}) \qquad (2.1)$$

uniformly over z, where G(z,F) is the asymptotic CDF of T_n, g_1 and g_2 are functionals of (z,F), $g_1(z,F)$ is an even function of z for each F, $g_2(z,F)$ is an odd function of z, and $g_2(z,F_n) \rightarrow g_2(z,F)$ almost surely or in probability as $n \rightarrow \infty$ uniformly over z. It follows from (2.1) and the symmetry of g_1 and g_2 that

$$P(|T_n| > z) = 1 - [G(z,F) - G(-z,F)] - 2n^{-1} g_2(z,F) + o(n^{-1}) \quad (2.2)$$

uniformly over $z \geq 0$.

The bootstrap replaces F with F_n and samples F_n conditional on the original sample $\{X_i\}$. Let T_n^* be the bootstrap version of T_n and P* be the probability measure induced by bootstrap sampling. Then under bootstrap sampling

$$P^*(|T_n^*| > z) = 1 - [G(z,F_n) - G(-z,F_n)]$$

$$- 2n^{-1} g_2(z,F_n) + o_p(n^{-1}), \qquad (2.3)$$

uniformly over $z \geq 0$. It follows from (2.2) and (2.3) that if G is sufficiently smooth,

$$P^*(|T_n^*| > z) - P(|T_n| > z) = O\{[G(z,F_n) - G(z,F)]$$

$$- [G(-z,F_n) - G(-z,F)]\}$$

$$= O[F_n(z) - F(z)] = O_p(n^{-1/2})$$

uniformly over $z \geq 0$. Thus, in general the bootstrap makes an error of size $O_p(n^{-1/2})$, which is the same as the size of the error made by first-order asymptotic approximations.

Now suppose that T_n is asymptotically pivotal. Then its asymptotic distribution is independent of F, and $G(z,F_n) = G(z,F)$ for all z. Equations (2.2) and (2.3) yield

$$P^*(|T_n^*| > z) - P(|T_n| > z) = 2n^{-1} [g_2(z,F) - g_2(z,F_n)] + o_p(n^{-1})$$

$$= o_p(n^{-1}) \qquad (2.4)$$

uniformly over $z \geq 0$. Now the bootstrap is accurate through $O_p(n^{-1})$, which is more accurate than 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.

It follows from (2.4) that

$$P(|T_n| > z_{n\alpha}^*) = \alpha + o(n^{-1}).$$

Thus, with the bootstrap critical value $z_{n\alpha}^*$, the level of a symmetrical, two-tailed, test based on an asymptotically pivotal statistic is correct through $O(n^{-1})$. In contrast, first-order asymptotic theory ignores all but the leading term in (2.1). Therefore, when a critical value based on first-order asymptotic theory is used, the error in the level of the test is $O(n^{-1})$.

In fact, symmetry arguments that are more refined than those given above show that with a bootstrap critical value, the error in the level of a symmetrical test based on an asymptotically pivotal statistic is usually of size $O(n^{-2})$. For a one-tailed test, the error with the bootstrap critical value is usually of size $O(n^{-1})$, and the error with the asymptotic critical value is of size $O(n^{-1/2})$. See Hall (1992) for details.

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 (1986, 1988).

b. The Importance of Asymptotically Pivotal Statistics

The arguments in Section 2a show that the bootstrap provides higher-order asymptotic approximations to the distributions and critical values of "smooth" asymptotically pivotal statistics. These include test statistics whose asymptotic distributions are standard normal or chi-square. The ability of the bootstrap to provide asymptotic refinements for such 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, such as regression coefficients, but it does not provide a higher-order approximation to their distribution. Bootstrap estimates of the distributions of statistics that are not asymptotically pivotal have the same accuracy as first-order asymptotic approximations.

Higher-order approximations to the distributions of statistics that are not asymptotically pivotal can be obtained through the use of prepivoting or bootstrap iteration (Beran 1987,1988) or bias-correction methods (Efron 1987). Bootstrap iteration is highly computationally intensive, however, which makes it unattractive when an asymptotically pivotal statistic is available.

c. Confidence Intervals

Let θ be a population parameter, θ_n be a $n^{1/2}$-consistent, asymptotically normal estimator of θ, and s_n be an estimate of the asymptotic standard deviation of $n^{1/2}(\theta_n - \theta)$. Then an asymptotic $1 - \alpha$ confidence interval for θ is $\theta_n - z_{\alpha/2} s_n \leq \theta \leq \theta_n + z_{\alpha/2} s_n$, where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. The error in the coverage probability of this confidence interval is $O(n^{-1})$. The bootstrap can be used to reduce the error in the coverage probability.

To do this, let θ_n^* and s_n^* be the bootstrap analogs of θ_n and s_n. That is, θ_n^* and s_n^* are the estimators that are obtained by sampling the distribution whose CDF is F_n, rather than the population distribution. Let T_n^* be the t statistic for testing the hypothesis $H_0^*: \theta = \theta_n$ using the bootstrap sample, and let $z_{n\alpha/2}^*$ be the $1 - \alpha/2$ quantile of the bootstrap distribution of T_n^*. Then $\theta_n - z_{\alpha/2}^* s_n \leq \theta \leq \theta_n + z_{\alpha/2}^* s_n$ is a $1 - \alpha$ confidence interval for θ based on the bootstrap critical value. It follows from the arguments made in Section 2a that the coverage probability of this confidence interval is correct through $O(n^{-1})$. Usually, the error in its coverage probability is $O(n^{-2})$. See Hall (1992) for details in the case of a confidence interval for a population mean.

c. The Parametric Versus the Nonparametric Bootstrap

The size of the error in the bootstrap estimate of a distribution or critical value is determined by the size of $F_n - F$. Thus, F_n should be the most efficient available estimator. If F belongs to a known parametric family $F(\bullet,\theta)$, $F(\bullet,\theta_n)$ should be used to generate bootstrap samples, rather than the EDF. Although the bootstrap provides asymptotic refinements regardless of whether $F(\bullet,\theta_n)$ 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 $F(\bullet,\theta_n)$ than with the EDF. If the objective is to test a hypothesis H_0 about θ, further gains in efficiency and performance can be obtained by imposing the constraints of H_0 when obtaining the estimate θ_n.

To illustrate, consider testing the hypothesis H_0: $\beta_1 = 0$ in the Box-Cox regression model

$$Y^{(\lambda)} = \beta_0 + \beta_1 X + U, \qquad (2.5)$$

where $Y^{(\lambda)}$ is the Box-Cox (1964) transformation of Y. Suppose that $U \sim N(0,\sigma^2)$.[2] Then bootstrap sampling can be carried out in the following ways:

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

2. Estimate λ, β_0, and β_1 in (2.5) by maximum likelihood, and obtain residuals $\hat{U}$. Generate Y values from $Y = [\lambda_n(b_0 + b_1X + U^*) + 1]^{1/\lambda n}$, where λ_n, b_0, and b_1 are the estimates of λ, β_0, and β_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 σ^2.

4. Estimate λ, β_0, and σ^2 in (2.5) by maximum likelihood subject to the constraint $\beta_1 = 0$. Then proceed as in method 2.

5. Estimate λ, β_0, and σ^2 in (2.5) by maximum likelihood subject to the constraint $\beta_1 = 0$. 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 β_0, β_1, and λ if the distribution of U is not specified parametrically.

If the objective is to obtain a confidence interval for β_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 $E(U|X) = 0$. This problem is discussed further in Section 3b, where heteroskedastic regression models are considered.

d. Recentering

By replacing F with F_n in (2.1), it can be seen that the bootstrap will not obtain even the correct asymptotic distribution of a statistic T_n unless $G_n(\bullet,F_n)$ converges weakly in probability to $G(z,F)$. One important situation in which this does not happen is generalized method of moments (GMM) estimation of an overidentified parameter when F_n is the EDF of the sample.

To see why, suppose that θ is identified by the moment condition $Eh(X,\theta) = 0$, where $\dim(h) > \dim(\theta)$. 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 $Eh(X,\theta)$ is then

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

where E* denotes the expectation relative to F_n, and θ_n is the GMM estimate of θ. In general, $E^*h(X,\theta_n) \neq 0$ 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.

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

$$h^*(X,\theta_n) = h(X,\theta_n) - n^{-1}\sum_{i=1}^{n} h(X_i,\theta_n).$$

Freedman (1981) recognized the need for recentering residuals in regression models without intercepts. See, also, Efron (1979). Brown and Newey (1992) show that recentering also can be accomplished by replacing F_n with an empirical likelihood estimate of F.

e. 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 can be done relatively easily if one has 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 $\{Y_t\}$ is generated by the ARMA model

$$A(L,\alpha)Y_t = B(L,\beta)U_t, \qquad (2.6)$$

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 random variables. Let α_n and β_n be $n^{1/2}$-consistent estimators of α and β, and let $\{\hat{U}_t\}$ be the centered residuals of the estimated model (2.6). Then a bootstrap sample $\{Y_t^*\}$ can be generated as

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

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 $\{U{*}\}$ 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.

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. This approach to bootstrap sampling is important in GMM estimation with dependent data, since the moment conditions on which GMM estimation is based usually do not specify the dependence structure of the GMM residuals. Lahiri (1992) gives conditions under which the block bootstrap provides asymptotic refinements through $O_p(n^{-1/2})$ for normalized sample moments and for a Studentized sample moment with m-dependent data. Hall and Horowitz (1994) give conditions under which the block bootstrap provides asymptotic refinements through $O(n^{-1})$ for test statistics associated with GMM estimation. Hall and Horowitz (1994) do not assume that the data-generation process is m-dependent.

The blocks into which the data are divided for purposes of block-bootstrap sampling may be nonoverlapping (Carlstein 1986) or overlapping (Künsch 1988). See, also, Hall (1985). Overlapping blocks provide somewhat higher bootstrap estimation efficiency than nonoverlapping ones, but available evidence indicates that the efficiency gain from using overlapping blocks is quite small. Hall *et al.* (1995) report the results of an analytic comparison of the estimation efficiencies of overlapping and nonoverlapping blocks for bootstrap estimation of the distribution of a sample mean. Let $\bar{X}$, μ, and s denote the sample mean, population mean, and standard deviation of $\bar{X}$, respectively. For estimating $P(|\bar{X} - \mu| / s \leq z)$ the reduction in asymptotic root-mean-square error from using overlapping blocks instead of nonoverlapping ones is less than 10 percent.

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 covariance matrices of the asymptotic forms of parameter estimators obtained from the original sample and from the bootstrap sample are not the same. The practical consequence of this difference is that asymptotic refinements 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) derives the bootstrap version of a Studentized sample mean for m-dependent data. Hall and Horowitz (1994) derive 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 θ that is identified by the moment condition $Eh(X,\theta) = 0$. Hall and Horowitz (1994) show that the correct 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 θ 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 (1994) for details.

At present, the block bootstrap is known to provide asymptotic refinements in GMM estimation only if the residuals $\{h(X_i,\theta_0): i = 1,2,...\}$ at the true parameter point, θ_0, are uncorrelated after finitely many lags. That is,

$$Eh(X_i,\theta_0)h(X_j,\theta_0) = 0 \quad \text{if } |i - j| > M \qquad (2.7)$$

for some $M < \infty$. This restriction is not equivalent to m-dependence since 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 (2.7), 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 root-n. However, present results on the existence of higher-order asymptotic expansions with dependent data (Götze and Hipp 1983) apply only to functions of sample moments that have root-n rates of convergence. Thus, it is necessary to extend existing theory of asymptotic expansions with dependent data before (2.7) can be relaxed.

f. Special Problems

The discussion in Section 2a shows that 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.

Other examples of bootstrap failure that are relevant to econometrics include the estimating the distribution of the maximum of a sample from the uniform distribution (Bickel and Freedman 1981), estimating the distribution of the mean of a sample from a population with infinite variance Athreya (1987), and unit-root models with certain resampling procedures (Basawa *et al.* 1991). Politis and Romano (1994) describe an alternative to the bootstrap that provides the correct asymptotic distribution in some of these cases.

g. 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, 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 $G(\bullet,F_n)$ 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 ψ. 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 an exact test of a certain simple hypothesis. The power of the test with a bootstrap critical value is a higher-order approximation to the power of the exact test of the simple hypothesis.

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

Now let T_n be a statistic that is asymptotically pivotal under H_0. Suppose that its exact CDF with an arbitrary value of ψ is $G_n(\bullet,\psi)$, and that under H_0 its asymptotic CDF is $G(\bullet)$. 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 ψ_n, so the exact distribution of the bootstrap version of T_n is $G_n(\bullet,\psi_n)$. Under H_0 and subject to regularity conditions, $G_n(\bullet,\psi_n)$ has an asymptotic expansion of the form

$$G_n(z,\psi_n) = G(z) + n^{-j/2} g_j(z,\psi_n) + o_p(n^{-j/2}) \qquad (2.8)$$

uniformly over z, where j = 1 or 2 depending on the symmetry of T_n. Usually j = 1 if T_n is the statistic for a one-tailed test and j = 2 if T_n is the statistic for a symmetrical, two-tailed test. It follows from (2.8) and the convergence of ψ_n to ψ^* that

$$G_n(z,\psi_n) = G_n(z,\psi^*) + o_p(n^{-j/2})$$

uniformly over z. 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 a model that satisfies H_0 with pseudo-true parameter values.

3. MONTE CARLO EVIDENCE ON NUMERICAL PERFORMANCE

The bootstrap provides a higher-order asymptotic approximation to critical values for tests based on "smooth" asymptotically pivotal statistics. When a bootstrap-based critical value is used for such a test, the difference between the test's true and nominal levels decreases more rapidly with increasing sample size than it does when the critical value is obtained from first-order asymptotic theory. Given a sufficiently large sample, the nominal level of the test will be closer to the true level when a bootstrap critical value is used than when a critical value based on first-order asymptotic theory is used. However, nothing in the theory guarantees that the numerical difference between the true and nominal levels of a test using a bootstrap critical value will be small in a specific application with a fixed sample size.

This section provides Monte Carlo evidence on the numerical performance of the bootstrap as a means of reducing differences between the true and nominal levels of tests. In the examples that are presented, the bootstrap often, though not always, essentially eliminates the distortions of level that occur when critical values are obtained from first-order asymptotic theory. In cases where the bootstrap does not remove distortions of level, it provides an indication that first-order asymptotic approximations are inaccurate.

a. 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 levels of the IM test with asymptotic critical values 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 level 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) ω_3. The Chesher-Lancaster form is relatively easy to compute because 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 level distortions of ω_3 are smaller than those of many other forms of the IM test statistic. Orme's ω_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. The Monte Carlo procedure consisted of repeating the following steps 1000 times for each form of the IM test:

1. Generate an estimation data set of size n = 50 or 100 by random sampling from the model under consideration. X was fixed in repeated samples. Estimate the unknown parameters of the model by maximum likelihood and compute the IM test statistic using the full vector of indicators. Call its value IM_0.

2. Generate a bootstrap sample of size n by random sampling from the model under consideration but using the parameter values estimated in step 1 instead of the true values. Using this sample, re-estimate the model's parameters by maximum likelihood and compute the IM test statistic. Call its value IM_B. Estimate the 0.05-level critical value of the IM test from the empirical distribution of IM_B that is obtained by repeating this step 100 times.[3] Let z_n* denote the estimated critical value.

3. Reject the model being tested at the nominal 0.05 level based on the bootstrap critical value if $IM_0 > z_n*$. Reject the model at the nominal 0.05 level based on the asymptotic critical value if IM_0 exceeds the 0.95 quantile of the chi-square distribution with degrees of freedom equal to the number of indicators.

Table 1 summarizes the results of the experiments. As expected, the differences between empirical and nominal levels 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 levels are very small. The bootstrap essentially eliminates the level distortions of the three forms of the IM test.

Horowitz (1994) also carried out a Monte Carlo investigation of the power of the IM test with bootstrap critical values. This investigation, like the investigation of levels, was carried out using Tobit and binary probit models. However, the models used to generate data in step 1 above were different from those estimated in steps 1 and 2. Data were generated from models that either included interaction terms among the components of X or were heteroskedastic. The estimated models did not have either interactions or heteroskedasticity.

Table 2 summarizes the results of the power experiments. The powers of Chesher-Lancaster and original White forms of test are similar and larger than those of ω_3 when bootstrap critical values are used. Since the Chesher-Lancaster form has larger distortions of level with asymptotic critical values than do the other forms, these results show that getting the level of a test right and getting high power are different tasks. A test with severe distortions of level when asymptotic critical values are used may have higher power once the distortions are removed than a test whose true and nominal levels with asymptotic critical values are similar to one another. In the examples shown here, the bootstrap eliminates the problem of getting level right and permits concentration on choosing a form of the IM test with high power.

b. 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 β in the model

$$Y = X\beta + U. \tag{3.1}$$

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 $\Omega(x) \equiv Var(U|X = x) < \infty$ for all x in the support of X.

Let b_n be the ordinary least squares (OLS) estimator of β in (3.1), b_{ni} and β_i be the i'th components of b_n and β, and s_{ni} be the square root of the (i,i) element of the HCCME. The t statistic for testing H_0: $\beta_i = \beta_{i0}$ is

$$T_n = (b_{ni} - \beta_{i0})/s_{ni} .$$

Under regularity conditions, T_n is asymptotically distributed as N(0,1). However, Chesher and Jewitt (1987) have shown that s_{ni}^2 can be seriously biased downward. Therefore, the true level of a test based on T_n is likely to exceed the nominal level. As is shown later in this section, the differences between the true and nominal levels can be very large when n is small.

The bootstrap can be implemented for model (3.1) by resampling observations of (Y,X) randomly with replacement. The resulting bootstrap sample is used to estimate β by OLS and compute T_n*, the t statistic for testing H_0*: $\beta_i = b_{ni}$. The bootstrap empirical distribution of T_n* is obtained by repeating this process many times, and the α-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, $\hat{U}$, independently of X. Similarly, one cannot implement the bootstrap by sampling U from a parametric model because (3.1) 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) established the ability of the wild bootstrap to provide asymptotic refinements for the model (3.1). 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 (3.1) as

$$Y_i = X_i b_n + \hat{U}_i, \quad i = 1,\ldots,n \qquad (3.2)$$

where Y_i and X_i are the i'th observed values of Y and X, and $\hat{U}_i$ is the i'th OLS residual. For each i = 1,...,n, let F_i be the unique 2-point distribution that satisfies

$$E(Z_i|F_i) = 0$$

$$E(Z_i^2|F_i) = \hat{U}_i^2$$

$$E(Z_i^3|F_i) = \hat{U}_i^3 ,$$

where Z is a random variable with the CDF F_i. In this distribution, $Z = (1 - \sqrt{5})\hat{U}_i/2$ with probability $(1 + \sqrt{5})/(2\sqrt{5})$, and $Z = (1 + \sqrt{5})\hat{U}_i/2$ with probability $1 - (1 + \sqrt{5})/(2\sqrt{5})$. The wild bootstrap is implemented as follows:

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

2. Estimate (3.1) 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.

I have carried out a small Monte Carlo investigation of the ability of the bootstrap and wild bootstrap to reduce the distortions in the level of a symmetrical, two-tailed t test that occur when asymptotic critical values are used. The bootstrap is implemented by resampling (Y,X) pairs, and the wild bootstrap is implemented as described above. The Monte Carlo experiments also investigate the level 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 level than it did with several other versions of the HCCME.

In the experiments reported here, $n = 25$. X is fixed in repeated samples and consists of an intercept and either 1 or 2 explanatory variables. In experiments in which X has an intercept and one explanatory variable, $\beta = (1,0)'$. In experiments in which X has an intercept and two explanatory variables, $\beta = (1,0,1)'$. The hypothesis tested in all experiments is $H_0: \beta_2 = 0$. 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

$$Y = X\beta + X\delta + V, \qquad (3.3)$$

where V and the components of δ are independently distributed as N(0,1).

There were 1000 Monte Carlo replications in each experiment. In the experiments with the HCCME, each replication consisted of the following steps:

1. Generate an estimation data set of size n by random sampling from model (3.1) or (3.3). Estimate β by OLS and compute the t statistic, T_n, for testing H_0. Also estimate β by OLS subject to the constraint that H_0 is true.

2. Generate a bootstrap sample of size n. This is done by either resampling (Y,X) pairs randomly with replacement or by using $Y_i^* = X_i b_n + U_i^*$ with b_n the constrained OLS estimate of β and U_i^* generated by the wild bootstrap. Using the bootstrap or wild bootstrap sample, re-estimate β by unconstrained OLS and compute the t statistic for testing $H_0^*: \beta_2 = b_{n2}$ if (Y,X) pairs are resampled or $H_0^*: \beta_2 = 0$ if the wild bootstrap is used. Call the value of the bootstrap or wild bootstrap t statistic T_n^*. Estimate the 0.05-level critical value of the t test from the empirical distribution of T_n^* that is obtained by repeating this step 100 times. Let $z_{0.05}^*$ denote the estimated critical value. Increasing the number of repetitions of this step beyond 100 has little effect on the results of the experiments.

3. Reject H_0 at the nominal 0.05 level based on the bootstrap critical value if $|T_n| > z_{0.05}*$. Reject H_0 at the nominal 0.05 level based on the asymptotic critical value if $|T_n| > 1.96$, which is the asymptotic 0.05-level critical value for the symmetrical, two-tailed t test.

The experiments with the jackknife version of the HCCME did not include bootstrapping or constrained OLS estimation but were identical in other respects.

Table 3 shows the empirical levels of the nominal 0.05-level t tests of H_0. The differences between the empirical and nominal levels 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 levels, but the empirical levels are still 2-3 times the nominal levels. With critical values obtained from the wild bootstrap, the differences between the empirical and nominal levels are very small. In these experiments, the wild bootstrap essentially removes the distortions of level that occur with asymptotic critical values.

c. Non-invariance of the Wald Test

The Wald statistic is not invariant to the specification of the null hypothesis being tested, H_0. The statistic is a different function of the data for different but algebraically equivalent specifications of H_0, and its numerical value can vary greatly according to the specification that is used. As a result the finite-sample level of the Wald test based on the asymptotic critical value depends on the specification of H_0 and can differ greatly from the nominal level. Gregory and Veall (1985), Lafontaine and White (1986), Breusch and Schmidt (1988), and Dagenais and Dufour (1991) discuss this problem.

A related problem concerns the t statistic for testing a hypothesis about a slope coefficient in a linear regression model with a Box-Cox (1964) transformed dependent variable. The t statistic is a Wald statistic and is not invariant to changes in the measurement units, or scale, of the dependent variable (Spitzer 1984). Thus, the numerical value of the t statistic and the finite-sample levels 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 levels of the t test with asymptotic critical values can be far from the nominal levels.

The distortions of level that occur when asymptotic critical values are used indicate that first-order asymptotic theory does not provide a good approximation to the finite-sample distribution of the Wald statistic. The bootstrap provides a better approximation to the finite-sample distribution and, therefore, better finite-sample critical values. This section reports the results of a Monte Carlo investigation of the ability of the bootstrap to provide improved finite-sample critical values for the Wald statistic and, thereby, to reduce the sensitivity of the finite-sample level of the Wald test to reparameterizations and rescalings.

Reparameterizations of the Null Hypothesis

This section reports the results of a Monte Carlo investigation of the finite-sample levels of Wald tests of algebraically equivalent specifications of H_0 when bootstrap critical values are used. See Horowitz and Savin (1992) for a more extensive investigation.

The model that generates the data is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U; \quad U \sim N(0,\sigma^2), \qquad (3.4)$$

where X_1 and X_2 are fixed in repeated samples. This is a slightly modified version of the model investigated by Gregory and Veall (1985). Two algebraically equivalent null hypotheses are tested. These are H_0^A: $\beta_1 - 1/\beta_2 = 0$, and H_0^B: $\beta_1\beta_2 - 1 = 0$. X_1 and X_2 are generated by independent random sampling from the N(0,1) distribution, and $\beta_0 = \sigma^2 = 1$ in all of the experiments. The values of β_1 and β_2 vary according to the experiment. Formulas for the Wald statistics for testing H_0^A and H_0^B are given by Gregory and Veall (1985).

The sample size in all of the Monte Carlo experiments is n = 20. Each experiment consisted of repeating the following sequence of steps 500 times:

1. Generate an estimation data set of size n by random sampling from model (3.4). Estimate the β's and σ^2 by maximum likelihood, and compute the Wald statistic, W_n, for testing the specified form of H_0. Also estimate the β's and σ^2 by maximum likelihood subject to the constraint that H_0 is true.

2. Generate a bootstrap sample of size n by random sampling from (3.4) but using the constrained maximum likelihood estimates of the parameters instead of the true values. Using this sample, re-estimate β and σ^2 by unconstrained maximum likelihood and compute the Wald statistic. Call its value W_n*. Estimate the 0.05-level critical value of the Wald test from the empirical distribution of W_n* that is obtained by repeating this step 100 times. Let $z_{0.05}$* denote the estimated critical value. Increasing the number of repetitions of this step beyond 100 has little effect on the results of the experiments.

3. Reject H_0 at the nominal 0.05 level based on the bootstrap critical value if $W_n > z_{0.05}*$. Reject H_0 at the nominal 0.05 level based on the asymptotic critical value if $W_n > 3.84$, which is the 0.95 quantile of the chi-square distribution with one degree of freedom.

The results are shown in Table 4. As in Gregory and Veall (1985), the empirical levels of the Wald tests based on the asymptotic critical value are larger than the nominal level, especially with H_0^A and $\beta_1 =$ 10 or 5. The reason for this is clear from the table: the empirical critical values of the various forms of the Wald statistic tend to be larger than the asymptotic critical value of 3.84. With H_0^A and $\beta_1 = 10$, the empirical critical value exceeds the asymptotic critical value by more than a factor of 25. In contrast, the mean bootstrap critical values are close to the empirical ones, and the differences between the nominal and empirical levels of the Wald tests with bootstrap critical values are small.

Rescaling the Dependent Variable in a Box-Cox Regression Model

This section reports the results of a Monte Carlo investigation of the finite-sample level of a symmetrical, two-tailed 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 given by equation (2.5) with $U \sim N(0,\sigma^2)$, $\beta_0 = 2$, $\beta_1 = 0$ and $\sigma^2 = 0.0625$. X was sampled from N(4,4) and was fixed in repeated samples. The hypothesis being tested is H_0: $\beta_1 = 0$. The value of λ is either 0 or 1, depending on the experiment, and the scale of Y is either 0.2, 1, or 5. The sample sizes were n = 50 and 100.

Each experiment consisted of repeating the following sequence of steps 1000 times.

1. Generate an estimation data set of size n by random sampling from model (2.5) and multiplying Y by 0.2, 1, or 5. Estimate the β's, λ, and σ^2 by maximum likelihood, and compute the t statistic, T_n, for testing H_0. Also estimate the β_0, λ, and σ^2 by maximum likelihood subject to the constraint that $\beta_1 = 0$.

2. Generate a bootstrap sample of size n by random sampling from (2.5) but using the constrained maximum likelihood estimates of the parameters instead of the true values. Using this sample, re-estimate the β's, λ, and σ^2 by unconstrained maximum likelihood and compute the t statistic. Call its value T_n*. Estimate the 0.05-level critical value of the t test from the empirical distribution of T_n* that is obtained by repeating this step 100 times. Let $z_{0.05}*$ denote the estimated critical value. Increasing the number of repetitions of this step beyond 100 has little effect on the results of the experiments.

3. Reject H_0 at the nominal 0.05 level based on the bootstrap critical value if $|T_n| > z_{0.05}^*$. Reject H_0 at the nominal 0.05 level based on the asymptotic critical value if $|T_n| > 1.96$.

The results are shown in Table 5. The empirical critical value of the t test tends to be much smaller than the asymptotic value of 1.96, especially in the experiments with a scale factor of 5. As a result, the empirical level of the t test is usually much smaller than its nominal level. The mean bootstrap critical values, however, are very close to the empirical critical values, and the levels based on bootstrap critical values are very close to the nominal level.

d. A t Test in a Trend Model with AR(1) Errors

This section reports the results of a Monte Carlo investigation of the finite-sample level of a symmetrical, two-tailed t test of a hypothesis about the parameter β in the following time-series model:

$$Y_t = \alpha_0 + \alpha_1 t + U_t ; \quad t = 1,2,\ldots$$

$$U_t = \beta U_{t-1} + V_t ; \quad V_t \sim \text{i.i.d.}$$

This model has been investigated in detail by Nankervis and Savin (1993). The results presented here are taken from their paper.

The test is based on the t statistic that is obtained from OLS estimation of the reduced-form model

$$Y_t = \gamma + \delta t + \beta Y_{t-1} + V_t ; \quad t = 1,2,\ldots \qquad (3.5)$$

where $\gamma = [\alpha_0(1 - \beta) + \alpha_1\beta]$ and $\delta = \alpha_1(1 - \beta)$. Two versions of the model were investigated. In one version, called the stationary model, δ is known to be zero and Y_0 is random. Thus, only γ, β and the variance of V are estimated. In the second version, called the trend model, δ is estimated by OLS along with the other parameters, and Y_0 is a constant equal to the first observed value of Y.

In all of the experiments, data were generated using $\gamma = \delta = 0$. In experiments with the trend model, $Y_0 = 0$. The value of β varies according to the experiment. The hypothesis being tested is H_0: $\beta = \beta_0$, where β_0 is the value of β used to generate the data. Three different distributions of V were used: N(0,1), the lognormal with log V ~ N(0,1), and a mixture of normals in which V is sampled from N(0,1) with probability 0.8 and from N(0,16) with probability 0.2. The sample sizes were 10 or 20 in experiments with the stationary model and 100 with the trend model, where larger samples were needed to obtain satisfactory numerical accuracy with the bootstrap.

For purposes of bootstrap sampling, it was assumed that the distribution of V is unknown. Bootstrap realizations of V were generated by sampling the residuals of the stationary or trend model that was estimated subject to the constraint that H_0 holds (that is, β was constrained to equal β_0). In experiments with the stationary model, the initial value of Y in the bootstrap sample was

$$Y_0^* = [\hat{\gamma}/(1 - \beta_0)] + \sum_{j=0}^{m} \beta_0^j V_{-j}^*,$$

where $\hat{\gamma}$ is the constrained estimate of γ, and the V_{-j}^* are sampled randomly from the empirical distribution of constrained least squares residuals. The values of m are 1, 50, 100, 150, 200 for β_0 = 0.0, 0.5, 0.9, 0.95, and 0.99, respectively.

Each experiment consisted of carrying out the following steps 15000 times:

1. Generate a sample of size n from (3.5) with the chosen parameter values and distribution of V. Estimate the parameters by OLS and compute the t statistic for testing $H_0: \beta = \beta_0$. Call this value of the t statistic T_n. Also estimate the parameters by OLS subject to the constraint $\beta = \beta_0$.

2. Generate a bootstrap sample of size n from the constrained estimate of (3.5). Bootstrap values of V are sampled from the empirical distribution of the residuals of the constrained estimated model. Using this sample, re-estimate γ, β and, in the trend model, δ. Compute the t statistic for testing H_0. Call its value T_n^*. Estimate the 0.05-level critical value of the t test from the empirical distribution of T_n^* that is obtained by repeating this step 1000 times. Let $z_{0.05}^*$ denote the estimated critical value.

3. Reject H_0 at the nominal 0.05 level based on the bootstrap critical value if $|T_n| > z_{0.05}^*$. Reject H_0 at the nominal 0.05 level based on the asymptotic critical value if $|T_n| > 1.96$.

The results are shown in Table 6. The empirical level of the t test using the asymptotic critical value is much larger than the nominal level when $\beta_0 > 0.5$. The error in the level increases as β_0 approaches 1. This is not surprising since the t statistic is not asymptotically normal when $\beta_0 = 1$. In contrast, the empirical level of the t test using bootstrap critical values is close to the nominal level in all of the experiments.

e. The Bootstrap as an Indicator of the Accuracy of Asymptotic Critical Values

In Section 2f it was noted that the bootstrap cannot be expected to perform well when an Edgeworth expansion provides a poor approximation to the distribution of interest. Phillips and Park (1988) have argued that even when an Edgeworth expansion does not improve the numerical quality of asymptotic approximations, it may provide information on whether first-order approximations are accurate. Specifically, first-order approximations are likely to be inaccurate if higher-order terms through $O(n^{-1})$ in the expansion of the distribution of a statistic are large compared to the first-order term.

Since the bootstrap amounts to a one-term Edgeworth expansion, it can provide an indication of the accuracy of first-order approximations to the distributions of test statistics without the tedious algebra associated with analytic Edgeworth expansions. In particular, large differences between bootstrap and asymptotic critical values may indicate that first-order approximations are inaccurate. This idea will now be illustrated with some Monte Carlo experiments.

The experiments consist of testing H_0: $\beta = 0$ in the following linear model with an endogenous right-hand side variable, X, and two instruments, Z_1 and Z_2:

$$Y = \beta X + U, \qquad (3.6)$$

$$X = \gamma Z_1 + \gamma Z_2 + V \qquad (3.7)$$

$$\begin{bmatrix} U \\ V \end{bmatrix} = N(0,\Sigma); \qquad \Sigma = \begin{bmatrix} 1 & r_{uv} \\ r_{uv} & 1 \end{bmatrix}. \qquad (3.8)$$

X and the Z's are scalars, and the true value of β is 0. X is endogenous if $r_{uv} \neq 0$. Z_1 and Z_2 are generated independently one another and of U by gaussian AR(1) processes with means of 0, variances of 1, and first-order serial correlation coefficients of r_z. Experiments were carried out with $\gamma = 0.25$ and 1.0, $r_z = 0$ and 0.75, and $r_{uv} = 0.75$. With $\gamma = 0.25$, the correlation between the instruments Z and the endogenous regressor X is relatively low; the asymptotic value of R^2 from the regression of X on Z_1 and Z_2 is 0.11. With $\gamma = 1$, the value of R^2 is 0.67. The sample sizes used in the experiments are n = 50 and n = 100.

The value of β is estimated by GMM using the moment conditions $E[Z_1(Y - \beta X)] = 0$ and $E[Z_2(Y - \beta X)] = 0$. GMM estimation is carried out in two stages. The first stage is two-stage least squares using the instruments Z_1 and Z_2, and the second stage uses an estimate of the asymptotically optimal GMM weight matrix. This yields an asymptotically efficient estimator of β. H_0 is tested with a symmetrical, two-tailed t test based on the second-stage estimation results.

In experiments with $r_z = 0$, bootstrap samples are obtained by resampling the quadruplets (Y,X,Z_1,Z_2) randomly with replacement. In experiments with $r_z = 0.75$, (Y,X,Z_1,Z_2) is resampled in non-overlapping blocks. Experiments were carried out using block lengths of n/5 and n/10. The block bootstrap procedure is summarized in Section 2e of this chapter and discussed in detail in Hall and Horowitz (1994). Recentering and correction of block bootstrap t statistics were carried out using the methods described in Sections 2d and 2e.

Each experiment consisted of carrying out the following steps 1000 times:

1. Generate an estimation sample of size n from (3.6)-(3.8) with $\beta = 0$. Estimate β by two-stage GMM. Call the estimate b_n and compute the t statistic, T_n, for testing H_0: $\beta = 0$.

2. Generate a bootstrap sample by sampling the quadruplet (Y,X,Z_1,Z_2) from the estimation data randomly with replacement or in non-overlapping blocks. Estimate β from the bootstrap sample by using GMM after recentering, and compute the bootstrap t statistic, T_n^*, for testing H_0^*: $\beta = b_n$. Estimate the 0.05-level critical value of the t test from the empirical distribution of T_n^* that is obtained by repeating this step 100 times. Let $z_{0.05}^*$ denote the estimated critical value.

3. Reject H_0 at the nominal 0.05 level based on the bootstrap critical value if $|T_n| > z_{0.05}^*$. Reject H_0 at the nominal 0.05 level based on the asymptotic critical value if $|T_n| > 1.96$.

The results of the experiments are shown in Table 7. In most cases the bootstrap reduces but does not remove the distortions of the level of the t test that occur with asymptotic critical values.

The bootstrap does, however, provide a warning that first-order asymptotic approximations are inaccurate. In Figure 1, the 25th, 50th, and 75th percentiles of the empirical distributions of the nominal 0.05-level bootstrap critical values in the experiments are plotted against the empirical levels of the tests based on the asymptotic critical value. The figure also contains a horizontal line indicating the asymptotic critical value. The figure shows that the difference between the bootstrap and asymptotic critical values is an increasing function of the distortion of the level of the asymptotic test. Thus, the bootstrap is informative about the accuracy of first-order asymptotic approximations despite its inability to fully correct the distortions of levels caused by these approximations. In particular, the 25th percentile of the bootstrap critical value of the t test exceeds the asymptotic critical value whenever the level of the asymptotic test exceeds roughly 0.08. Further research might usefully investigate ways to decide in applications whether an observed difference between asymptotic and bootstrap critical values is evidence that first-order approximations are inaccurate or is simply an artifact of random sampling errors.

4. CONCLUSIONS

In applied econometrics, the bootstrap is often used as a substitute for analytical asymptotic formulae when the statistics of interest have complicated asymptotic distributions. This chapter has focussed on another important use of the bootstrap: it often provides a better approximation to the finite-sample distribution of an asymptotically pivotal statistic than does first-order asymptotic theory.

First-order asymptotic theory often gives a poor approximation to the finite-sample distributions of test statistics. As a result, the true and nominal levels of hypothesis tests can be greatly different. The examples presented in Section 3 show that the use of bootstrap-based critical values instead of asymptotic ones can provide dramatic reductions in the differences between the true and nominal levels of tests based on asymptotically pivotal statistics. In many cases of practical importance, the bootstrap essentially eliminates finite-sample errors in levels. Even when the bootstrap is not numerically accurate, it can provide a warning that first-order asymptotic approximations are inaccurate.

Although the discussion here has concentrated on reducing distortions in the levels of hypothesis tests, similar conclusions can be drawn concerning confidence intervals. The use of bootstrap-based critical values instead of asymptotic ones often greatly reduces the difference between true and nominal coverage probabilities.

Of course, the bootstrap should not be used blindly. It does not always eliminate distortions of levels and sometimes makes them worse. Proper attention must be given to matters such as recentering, correction of test statistics in the block bootstrap for dependent data, and choosing the distribution from which bootstrap samples are drawn. In other words, the bootstrap, like any other statistical method, is not foolproof and works best in the hands of a careful, informed user.

FOOTNOTES

1. Härdle and Hart (1992) proved consistency of a modified version of the bootstrap estimator used by Härdle, *et al.* (1991). "Consistency" in this setting means that the bootstrap estimator of the distribution function of the test statistic converges weakly to the correct asymptotic distribution function. Proving that the bootstrap is consistent for the maximum score estimator is a difficult problem that has not yet been solved.

2. Strictly speaking, U cannot be normally distributed unless $\lambda = 0$ 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.

3. Horowitz (1994) also considered the 0.01-level critical value, but these results are not reported here.

REFERENCES

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

Andrews, D.W.K. 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59. 817-858.

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 1991. Bootstrapping unstable first-order autoregressive processes, *Annals of Statistics*, 19, 1098-1101.

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.

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

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.

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. and W.K. Newey 1992. Bootstrapping for GMM. Seminar notes, Department of Economics, Massachusetts Institute of Technology.

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.

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.

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.

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.

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

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

Hall, P. 1986. 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. 1992. *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P. and J.L. Horowitz 1995. Bootstrap critical values for tests based on generalized-method-of-moments estimators, *Econometrica*, forthcoming.

Hall, P., J.L. Horowitz, and B.-Y. Jing 1995. On blocking rules for the bootstrap with dependent data, *Biometrika*, 82, 561-574.

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. 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. 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.

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

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

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.

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.

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

Li, H. and G.S. Maddala 1995. Bootstrapping time series models, *Econometric Reviews*, forthcoming.

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

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*, 21m 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.

Nankervis, J.C. and N.E. Savin 1995. The level and power of the bootstrap t-test in the trend model with AR(1) errors, *Journal of Business and Economic Statistics*, forthcoming.

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*, 20.

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 1994. Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, 61, 631-653.

Newey, W.K. and K.D. West 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

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.

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

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.

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.

TABLE 1

EMPIRICAL LEVELS OF NOMINAL 0.05-LEVEL INFORMATION-MATRIX TESTS OF PROBIT AND TOBIT MODELS

		Level Using			Level Using		
	Distr.	Asymptotic Critical Values			Bootstrap-Based Crit. Values		
N	of X	White	Chesh.-Lan.	Orme	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

TABLE 2

POWER OF THE INFORMATION-MATRIX TEST WITH BOOTSTRAP CRITICAL VALUES

Form of	Power with Boot. Crit. Values		
Misspec.	White	Chesh.-Lan.	Orme
	Probit Models		
Interaction	0.652	0.667	0.311
Heterosked.	0.881	0.875	0.556
	Tobit Models		
Interaction	0.458	0.459	0.444
Heterosked.	0.506	0.401	0.028

TABLE 3

EMPIRICAL LEVELS OF t TESTS USING HETEROSKEDASTICITY-CONSISTENT COVARIANCE MATRIX ESTIMATORS

n = 25

Empirical Level at Nominal 0.05 Level

Form of Test	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

TABLE 4

EMPIRICAL LEVELS OF WALD TESTS OF H_0^A AND H_0^B

n = 20; Nominal Level = 0.05

	Null	Level with		Empirical	Mean
β_1, β_2	Hyp.	Crit. Val. from		Critical	Bootstrap
		Asymp.	Boot.	Value	Crit. Val.
10, 0.1	H_0^A	0.378	0.066	101.69	87.70
	H_0^B	0.082	0.074	4.12	3.69
5, 0.2	H_0^A	0.254	0.074	29.17	25.58
	H_0^B	0.092	0.074	4.01	3.68
2, 0.5	H_0^A	0.092	0.048	7.32	6.62
	H_0^B	0.074	0.074	4.31	3.75
1, 1	H_0^A	0.050	0.042	4.21	4.12
	H_0^B	0.104	0.080	5.04	4.71

TABLE 5

EMPIRICAL LEVELS OF t TESTS FOR BOX-COX REGRESSION MODEL

Nominal Level = 0.05

			Level Using			
			Crit. Val. from			Mean
		Scale			Empirical	Bootstrap
n	λ	Fac.	Asymp.	Boot.	Crit. Val.	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

TABLE 6

EMPIRICAL LEVELS OF t TESTS FOR AN AR(1) MODEL

Nominal Level = 0.05

	Empirical Level with			Empirical Level with		
	Asymptotic Critical Value			Bootstrap Critical Value		
	Distribution of V			Distribution of V		
β	Normal	Lognorm.	Mixture	Normal	Lognorm.	Mixture
Stationary model, n = 10 or 20						
0.0	0.032	0.017	0.024	0.050	0.035	0.050
0.50	0.054	0.023	0.038	0.052	0.035	0.052
0.90	0.129	0.097	0.103	0.055	0.040	0.059
0.95	0.159	0.138	0.145	0.057	0.042	0.063
0.99	0.182	0.193	0.198	0.056	0.048	0.073
Trend Model, n = 100						
0.0	0.051	0.031	0.049	0.051	0.042	0.053
0.50	0.066	0.034	0.054	0.053	0.041	0.052
0.90	0.164	0.129	0.148	0.054	0.050	0.054
0.95	0.255	0.225	0.240	0.054	0.050	0.054
0.99	0.479	0.461	0.465	0.056	0.055	0.059

TABLE 7

EMPIRICAL LEVELS OF t TESTS FOR LINEAR MODEL

WITH AN ENDOGENOUS REGRESSOR

Nominal level = 0.05

				Empirical Level	
n	γ	r_z	Blks	Asymp. Crit. Value	Boot. Crit. Value
50	0.25	0.0	50	0.128	0.124
	0.25	0.75	5	0.098	0.085
			10		0.079
100	0.25	0.0	100	0.082	0.085
	0.25	0.75	10	0.080	0.072
			20		0.076
50	1.0	0.0	50	0.077	0.058
	1.0	0.75	5	0.084	0.061
			10		0.049
100	1	0.0	100	0.063	0.056
	1	0.75	10	0.065	0.063
			20		0.060

LIST OF TABLES

LIST OF FIGURES