11-1-1996

On the Number of Bootstrap Repetitions for Bootstrap Standard Errors, Confidence Internals, and Tests

Donald W.K. Andrews

Moshe Buchinsky

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1141R

ON THE NUMBER OF BOOTSTRAP REPETITIONS FOR BOOTSTRAP STANDARD ERRORS, CONFIDENCE INTERVALS, AND TESTS

Donald W. K. Andrews and Moshe Buchinsky

November 1996
Revised August 1997

On the Number of Bootstrap Repetitions for Bootstrap Standard Errors, Confidence Intervals, and Tests

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Moshe Buchinsky
Department of Economics
Brown University
and
National Bureau of Economic Research

July 1996
Revised: July 1997

Abstract

This paper considers the problem of choosing the number of bootstrap repetitions B for bootstrap standard errors, confidence intervals, and tests. For each of these problems, the paper provides a three-step method for choosing B to achieve a desired level of accuracy. Accuracy is measured by the percentage deviation of the bootstrap standard error estimate, confidence interval endpoint(s), test's critical value, or test's p-value based on B bootstrap simulations from the corresponding ideal bootstrap quantities for which $B = \infty$. Monte Carlo simulations show that the proposed methods work quite well.

The results apply quite generally to parametric, semiparametric, and nonparametric models with independent and dependent data. The results apply to the standard nonparametric iid bootstrap, moving block bootstraps for time series data, parametric and semiparametric bootstraps, and bootstraps for regression models based on bootstrapping residuals.

Keywords: Bootstrap, bootstrap repetitions, coefficient of excess kurtosis, confidence interval, density estimation, hypothesis test, p-value, quantile, simulation, standard error estimate.

JEL Classification: C12, C13, C14, C15.

1. Introduction

Bootstrap methods have gained a great deal of popularity in empirical research. Although the methods are easy to apply, determining the number of bootstrap repetitions, B, to employ is a common problem in the existing literature. Typically, this number is determined in a somewhat ad hoc manner. This is problematic, because one can obtain a "different answer" from the same data merely by using different simulation draws if B is chosen to be too small. On the other hand, it is expensive to compute the bootstrap statistics of interest, if B is chosen to be extremely large. Thus, it is desirable to be able to determine a value of B that obtains a suitable level of accuracy for a given problem at hand. This paper addresses this issue in the context of the three main branches of statistical inference, viz., point estimation, interval estimation, and hypothesis testing.

We provide methods for determining B to attain specified levels of accuracy for bootstrap standard error estimates, confidence intervals, and hypothesis tests.[2] A three-step method for choosing B is proposed for each case. Three steps are required because one needs to determine the relevant features of the problem in the initial two steps before one can determine a suitable choice of B in the third step.

The measure of accuracy differs somewhat across the cases considered. For standard error estimates, we measure accuracy in terms of the percentage deviation of the bootstrap standard error estimate for a given value of B, from the ideal bootstrap estimate, for which $B = \infty$. For confidence intervals, we measure accuracy in terms of the percentage deviation of the bootstrap endpoint(s) of the confidence interval for a given value of B, from the ideal bootstrap endpoint(s). For symmetric two-sided confidence intervals, this measure of accuracy is equivalent to a measure based on the percentage deviation of the length of the confidence interval for a given value of B, from the ideal bootstrap length.

For tests with a specified significance level α, we measure accuracy in terms of the percentage deviation of the bootstrap critical value of the test for a given value of B, from the ideal bootstrap critical value. For tests in which one wants to report a p-value, we measure accuracy in terms of the percentage deviation of the bootstrap p-value for a given value of B, from the ideal bootstrap p-value. (We note that reporting a bootstrap p-value exploits the potential higher-order improvements of the bootstrap that are available for tests; see Section 4 below.) For each type of statistical inference, the measure of accuracy is directly related to the issue of whether one could obtain a "different answer" from the same data merely by using different simulation draws.

The accuracy obtained by a given choice of B is stochastic, because the bootstrap simulations are random. To determine a suitable value of B, we specify a bound on the relevant percentage deviation, denoted *pdb*, and we require that the actual percentage deviation is less than this bound with a specified probability, $1-\tau$, close to one. The three-step method takes *pdb* and τ as given and specifies a data-dependent method of determining a value of B, denoted B^*, such that the desired level of accuracy is obtained. For example, one might take $(pdb, \tau) = (10, .05)$. Then, the three-step method yields a value B^* such that the relevant percentage deviation is less than 10% with approximate probability .95.

The three-step methods are applicable in parametric, semiparametric, and nonparametric models with independent and identically distributed (iid) data, independent and non-identically distributed (inid) data, and time series data. The methods are applicable when the bootstrap employed is the standard nonparametric iid bootstrap, a moving block bootstrap for time series, a parametric or semiparameric bootstrap, or a bootstrap for regression models that is based on bootstrapping residuals. The methods are applicable to statistics that have normal and non-normal asymptotic distributions. Essentially, the results are applicable whenever the bootstrap samples are simulated to be iid across different bootstrap samples. We do not require that the simulations are iid within each bootstrap sample—in fact, they are not for most time series applications.

The results for confidence intervals apply to symmetric, equal-tailed, and one-sided percentile t confidence intervals, as defined in Hall (1992). Efron's (1987) AB_c confidence intervals are not considered. They will be considered elsewhere. The results for tests apply to a wide variety of tests of parametric restrictions and model specification based on t statistics, Wald statistics, Lagrange multiplier statistics, likelihood ratio statistics, etc.

For bootstrap standard error estimates, the three-step method depends on an estimate of the *coefficient of excess kurtosis*, γ_2, of the bootstrap distribution of the parameter estimator. We consider the usual estimator of γ_2 as well as a bias-corrected estimator of it. We compare these two methods via simulation. Because the computational cost of carrying out the bias correction is small and the gains are significant in some cases, we recommend use of the bias-corrected estimator of γ_2.

For confidence intervals and critical values of tests, the three-step method depends on estimates of the "density" evaluated at specific quantiles of the bootstrap distribution of the statistic that is used to construct the confidence interval or test. For this purpose, we use an estimator of Siddiqui (1960) with an optimal data-dependent smoothing parameter, which is a variant of that proposed by Hall and Sheather (1988).

The three-step methods are justified by asymptotic results. The small sample accuracy of the asymptotic results is evaluated via simulation experiments. We assess the performance of the three-step methods for standard error estimates and symmetric percentile t confidence intervals. In short, the simulations show that the methods work very well in the cases considered.

The closest results in the literature to the standard error results given here are those of Efron and Tibshirani (1986, Sec. 9). Efron and Tibshirani provide a simple formula that relates the coefficient of variation of the bootstrap standard error estimator, as an estimate of the true standard error, to the coefficient of variation of the ideal bootstrap standard error estimator, as an estimate of the true standard error. Their formula depends on some unknowns that are not estimable. Hence, Efron and Tibshirani only use their formula to suggest a range of plausible values of B. An advantage of our approach over that of Efron and Tibshirani is that the unknowns in our approach can be estimated quite easily. This allows us to specify an explicit method of choosing B to obtain a desired degree of accuracy of the bootstrap standard error estimator as an estimate of the ideal bootstrap standard error estimator.

The closest results in the literature to the confidence interval results given here are those of Hall (1986). His paper has two parts. The part that deals with coverage probabilities considers *unconditional* coverage probabilities, i.e., coverage probabilities with respect to the randomness in the data and the bootstrap simulations. In constrast, we consider *conditional* coverage probabilities, i.e., coverage probabilities with respect to the randomness in the data conditional on the bootstrap simulations. We do so because the bootstrap simulation randomness is *ancillary* and, hence, should be conditioned on when making inference according to the principle of ancillarity or conditionality; see Kiefer (1982). We do not want to be able to obtain "different answers" from the same data due to the use of different simulation draws.

The second part of Hall's (1986) paper (see Section 3) considers the asymptotic distribution of the difference between a bootstrap percentile t confidence interval endpoint based on B bootstrap repetitions and the ideal bootstrap endpoint. He considers the case where the bootstrap employed is the nonparametric iid bootstrap and the t statistic is a normalized sample mean of iid random variables.

While our three-step methods rely on similar results, we use a proof that differs from that of Hall in that it does not rely on smoothing the bootstrap distribution of the statistic. Furthermore, our results apply to a much wider array of statistics, types of bootstraps, and assumptions regarding the iid, inid, or dependent nature of the original sample than do Hall's, although it may be possible to generalize Hall's results along these lines. In addition, our results allow B to be data-dependent, as is necessary for the three-step methods.

In any event, the focus of Hall's results is quite different from that of this paper. Hall uses his results to demonstrate the near continuity of the discrete nonparametric iid bootstrap distribution of the t statistic. In contrast, we address the question of choosing a desired number, B^*, of bootstrap repetitions.

The closest results in the literature to the test results given here are those of Davidson and MacKinnon (1997). Davidson and MacKinnon consider the effect of the number of bootstrap repetitions on the unconditional power of a bootstrap test, i.e., the power with respect to the randomness in the data and the bootstrap simulations. They propose a pretesting method of choosing B that aims to achieve good unconditional power for a given significance level α. In contrast, the method that we consider aims to achieve a bootstrap test that has good conditional significance level given the simulation randomness. We do so for the same reason as given above for confidence intervals, viz., the bootstrap simulation randomness is *ancillary* and, hence, should be conditioned on when making inference according to the principle of ancillarity or conditionality.

The remainder of this paper is organized as follows. Sections 2, 3, and 4 provide the results for standard error estimates, confidence intervals, and tests respectively. Section 2.1 introduces notation and definitions for the standard error results. The notation follows that of Efron and Tibshirani (1993). Section 2.2 presents a formula for the accuracy of the bootstrap standard error estimator for finite B as an estimator of the ideal bootstrap standard error estimator. Section 2.3 introduces the three-step method for determining B for bootstrap standard error estimates. Section 2.4 presents Monte Carlo simulation results for the three-step method of determining B. Section 2.5 introduces a bias-corrected three-step method of determining B for bootstrap standard error estimates. Section 2.6 assesses its performance via Monte Carlo simulation.

Section 3.1 introduces notation and definitions for symmetric percentile t confidence intervals. The notation follows that of Hall (1992). Section 3.2 provides the three-step method for these confidence intervals. Section 3.3 provides the asymptotic justification of the three-step method for these confidence intervals. Section 3.4 presents Monte Carlo simulation results for the three-step method of determining B for symmetric confidence intervals. Section 3.5 extends the results of Sections 3.1–3.3 to equal-tailed and one-sided percentile t confidence intervals.

Section 4.1 provides the three-step method and its justification for tests with a specified significance level. Section 4.2 provides the three-step method and its justification for tests when a p-value is to be reported.

An Appendix of Proofs provides proofs of the results given in Sections 2–4.

2. Standard Error Results

In this section, we present the results for bootstrap standard error estimates.

2.1. Notation and Definitions

The observed data are a sample of size n: $\mathbf{X} = (X_1, ..., X_n)'$. Let $\widehat{\theta} = \widehat{\theta}(\mathbf{X})$ be an estimator of a scalar parameter θ_0 based on the sample $\mathbf{X}$. We are interested in estimation of the standard error, se, of $\widehat{\theta}$. By definition,

$$se = \left(E(\widehat{\theta}(\mathbf{X}) - E\widehat{\theta}(\mathbf{X}))^2\right)^{1/2}, \tag{2.1}$$

where E denotes expectation with respect to the randomness in $\mathbf{X}$. Of course, se depends on n, but we take n to be fixed in this section except where stated otherwise.

Let $\mathbf{X}^* = (X_1^*, ..., X_n^*)'$ be a bootstrap sample of size n based on the original sample $\mathbf{X}$. When the original sample $\mathbf{X}$ is comprised of iid or inid random variables, the bootstrap sample $\mathbf{X}^*$ often is an iid sample of size n drawn from some distribution $\widehat{F}$. For example, for the nonparametric bootstrap, $\widehat{F}$ is the empirical distribution function based on $\mathbf{X}$. That is, $\widehat{F}(x) = \frac{1}{n}\Sigma_{i=1}^n 1(X_i \leq x)$, where $1(X_i \leq x)$ denotes the indicator function of $X_i \leq x$. For parametric and semiparametric bootstraps, $\widehat{F}$ typically depends on estimators of θ_0 and other parameters. When the original sample $\mathbf{X}$ is comprised of dependent data, the bootstrap sample often is taken to be a moving block bootstrap or some variation of this; see Carlstein (1986), Kunsch (1989), Hall and Horowitz (1996), and Li and Maddala (1996). When the model is a regression model with independent or dependent data, the bootstrap sample is sometimes generated by bootstrapping the residuals; see Freedman (1981), Li and Maddala (1996), and the references therein. All of these bootstrap methods are covered by our results.

The "ideal" bootstrap standard error estimator of se is

$$\widehat{se}_\infty = \left(E^*(\widehat{\theta}(\mathbf{X}^*) - E^*\widehat{\theta}(\mathbf{X}^*))^2\right)^{1/2}, \tag{2.2}$$

where E^* denotes expectation with respect to the randomness in the bootstrap sample $\mathbf{X}^*$ conditional on the observed data $\mathbf{X}$.

Analytic calculation of the ideal bootstrap standard error is usually intractable. Instead one usually approximates it using bootstrap simulations. Consider B iid bootstrap samples $\{\mathbf{X}_b^* : b = 1, ..., B\}$ each with the same distribution as $\mathbf{X}^*$. The quantity B is referred to as the number of bootstrap repetitions. The corresponding B bootstrap estimates of θ_0 are denoted by $\widehat{\theta}_b^* = \widehat{\theta}(\mathbf{X}_b^*)$ for $b = 1, ..., B$. The bootstrap standard error estimator for B bootstrap repetitions is

$$\widehat{se}_B = \left(\frac{1}{B-1}\Sigma_{b=1}^B \left(\widehat{\theta}_b^* - \frac{1}{B}\Sigma_{c=1}^B \widehat{\theta}_c^*\right)^2\right)^{1/2}. \tag{2.3}$$

Note that

$$\lim_{B\to\infty} \widehat{se}_B = \widehat{se}_\infty \tag{2.4}$$

in probability and almost surely by the law of large numbers provided $E^*(\widehat{\theta}(\mathbf{X}^*))^2 < \infty$. The latter holds automatically for the nonparametric bootstrap due to its finite support.

Here and below (except as stated otherwise), all probability statements and the probability and expectation operators P^* and E^*, respectively, refer to the randomness in the iid bootstrap samples $\{\mathbf{X}_b^* : b = 1, ..., B\}$ conditional on the observed data $\mathbf{X}$. We note that our results are applicable in any bootstrap context in which the simulated bootstrap samples $\{\mathbf{X}_b^* : b = 1, ..., B\}$ are iid over the index b.

Let μ and γ_2 denote the *mean* and the *coefficient of excess kurtosis* of the bootstrap estimator $\widehat{\theta}_b^*$. By definition,

$$\begin{aligned} \mu &= E^*\widehat{\theta}(\mathbf{X}^*) \text{ and} \\ \gamma_2 &= \frac{E^*(\widehat{\theta}(\mathbf{X}^*) - \mu)^4}{(E^*(\widehat{\theta}(\mathbf{X}^*) - \mu)^2)^2} - 3 = \frac{E^*(\widehat{\theta}(\mathbf{X}^*) - \mu)^4}{\widehat{se}_\infty^4} - 3. \end{aligned} \tag{2.5}$$

(From above, the standard error of $\widehat{\theta}_b^*$ is $\widehat{se}_\infty$.) Note that $\gamma_2 = 0$ if $\widehat{\theta}_b^*$ has a normal distribution, $\gamma_2 > 0$ if $\widehat{\theta}_b^*$ has kurtosis greater than that of a normal distribution, and $\gamma_2 < 0$ otherwise. For example, for a t distribution with df degrees of freedom, the coefficient of excess kurtosis is $6/(df-4)$. Thus, for $df = 10$, $\gamma_2 = 1$; for $df = 7$, $\gamma_2 = 2$; and for $df = 5$, $\gamma_2 = 6$. The range of possible values of γ_2 across all distributions is $[-2, \infty)$. (The normal and t distributions are mentioned here for illustrative purposes only. For the nonparametric bootstrap, it is not possible for $\widehat{\theta}(\mathbf{X}^*)$ to have a normal or t distribution, because the distribution of $\widehat{\theta}(\mathbf{X}^*)$ is discrete. Nevertheless, $\widehat{\theta}(\mathbf{X}^*)$ may have a discrete distribution that is closely approximated by a normal or t distribution.)

Often $\{d_n(\widehat{\theta}_b^* - \theta_0) : n \geq 1\}$ converges to a normal distribution as $n \to \infty$ and is uniformly integrable to the fourth power, where $\{d_n : n \geq 1\}$ is a divergent sequence of positive constants, such as $n^{1/2}$ or $n^{2/5}$. In such cases, $\gamma_2 \to_p 0$ as $n \to \infty$. In many applications, however, γ_2 is greater than zero in finite samples.

Estimates of μ and γ_2 are given by

$$\widehat{\mu}_B = \frac{1}{B}\Sigma_{b=1}^{B}\widehat{\theta}_b^* \text{ and}$$

$$\widehat{\gamma}_{2B} = \frac{\frac{1}{B-1}\Sigma_{b=1}^{B}(\widehat{\theta}_b^* - \widehat{\mu}_B)^4}{\widehat{se}_B^4} - 3. \tag{2.6}$$

These estimators are consistent by the law of large numbers and Slutsky's Theorem:

$$\lim_{B\to\infty} \widehat{\mu}_B = \mu \text{ and } \lim_{B\to\infty} \widehat{\gamma}_{2B} = \gamma_2 \tag{2.7}$$

in probability and almost surely, provided $\widehat{se}_\infty \neq 0$.

Let $\chi^2_{1-\tau}$ denote the $(1-\tau)$-th quantile of a chi-squared distribution with one degree of freedom for $\tau \in (0,1)$. That is, $P(Y \leq \chi^2_{1-\tau}) = 1-\tau$, where Y has χ^2 distribution with one degree of freedom.

2.2. A Formula for the Accuracy of $\widehat{se}_B$ as an Approximation for $\widehat{se}_\infty$

In this section, we give a simple formula which provides a probabilistic statement of how close $\widehat{se}_B$ is to $\widehat{se}_\infty$ as a function of the number of bootstrap repetitions B. The results are justified by asymptotics as $B \to \infty$ and are established by a simple application of the delta method.

The percentage deviation of $\widehat{se}_B$ as an estimate of $\widehat{se}_\infty$ is

$$100\frac{|\widehat{se}_B - \widehat{se}_\infty|}{\widehat{se}_\infty}. \tag{2.8}$$

Let $1-\tau$ denote a probability close to one, such as .95. Let *pdb* be a bound on the percentage deviation of $\widehat{se}_B$ from $\widehat{se}_\infty$. We want to determine $B = B(pdb, \tau)$ such that

$$P^*\left(100\frac{|\widehat{se}_B - \widehat{se}_\infty|}{\widehat{se}_\infty} \leq pdb\right) = 1-\tau. \tag{2.9}$$

Alternatively, for given B and τ, we want to determine $pdb = pdb(B,\tau)$ such that (2.9) holds.

The relationship between B, *pdb*, and τ that is determined by (2.9) satisfies the following approximate formula:

$$pdb \doteq 50(\chi^2_{1-\tau}(2+\gamma_2)/B)^{1/2} \text{ or equivalently}$$

$$B \doteq 2{,}500\chi^2_{1-\tau}(2+\gamma_2)/pdb^2. \tag{2.10}$$

This formula is accurate in the sense that

$$\lim_{B\to\infty} P^*\left(100\frac{|\widehat{se}_B - \widehat{se}_\infty|}{\widehat{se}_\infty} \leq 50\left(\chi^2_{1-\tau}(2+\gamma_2)/B\right)^{1/2}\right) = 1-\tau. \tag{2.11}$$

The proof of this result and others below are given in the Appendix of Proofs.

Formula (2.10) is not operational because it depends on the unknown parameter γ_2. One can substitute the consistent estimator $\widehat{\gamma}_{2B}$ of (2.6) for γ_2 to obtain

$$\begin{aligned} pdb &\doteq 50\left(\chi^2_{1-\tau}(2+\widehat{\gamma}_{2B})/B\right)^{1/2} \text{ or equivalently} \\ B &\doteq 2{,}500\chi^2_{1-\tau}(2+\widehat{\gamma}_{2B})/pdb^2. \end{aligned} \tag{2.12}$$

Equations (2.7) and (2.11) combine to give

$$\lim_{B\to\infty} P^*\left(100\frac{|\widehat{se}_B-\widehat{se}_\infty|}{\widehat{se}_\infty} \le 50\left(\chi^2_{1-\tau}(2+\widehat{\gamma}_{2B})/B\right)^{1/2}\right)=1-\tau, \tag{2.13}$$

which justifies (2.12).

We now show how the formula of (2.12) can be utilized. Suppose B has been specified, perhaps by the author of some research paper of interest. We are interested in whether this choice of B is sufficiently large to yield $\widehat{se}_B$ close to $\widehat{se}_\infty$. Take $1-\tau$ close to one, say .95. Then, $\chi^2_{1-\tau}=3.84$ and

$$pdb \doteq 98\left((2+\widehat{\gamma}_{2B})/B\right)^{1/2}. \tag{2.14}$$

Table 1 provides the values of *pdb* that correspond to an array of values of $\widehat{\gamma}_{2B}$ and B when $1-\tau=.95$. For example, if $\widehat{\gamma}_{2B}=0$ (which corresponds to the kurtosis of the normal distribution) and $B=50$, then $pdb \doteq 20$. That is, with probability approximately .95, $\widehat{se}_B$ is within $\pm 20\%$ of $\widehat{se}_\infty$. Or, with probability approximately .95, $\widehat{se}_\infty$ is within $\pm 20\%$ of $\widehat{se}_B$. (The latter interpretation is valid because (2.13) holds with $\widehat{se}_\infty$ in the denominator replaced by $\widehat{se}_B$.)

Table 1 shows that to obtain very accurate estimates of $\widehat{se}_\infty$, say $pdb=5$, one needs quite large values of B, e.g., $B=750$ when $\widehat{\gamma}_{2B}=0$ and $B=2{,}000$ when $\widehat{\gamma}_{2B}=3$. Much smaller values of B are required to obtain moderate accuracy, say $pdb=20$, e.g., $B=50$ when $\widehat{\gamma}_{2B}=0$ and $B=100$ when $\widehat{\gamma}_{2B}=2$.

2.3. A Three-step Method for Determining the Number of Bootstrap Repetitions

We now specify a three-step method for determining B to achieve a desired accuracy of $\widehat{se}_B$ for estimating $\widehat{se}_\infty$. The desired accuracy is specified by a (pdb, τ) combination, such as (10, .05). The method involves the following steps:

Step 1. Suppose $\gamma_2=0$ and use (2.10) to specify a preliminary value of B, denoted B_0. By definition,

$$B_0=\text{int}(5{,}000\chi^2_{1-\tau}/pdb^2), \tag{2.15}$$

where $\text{int}(a)$ denotes the smallest integer greater than or equal to a.

Step 2. Simulate B_0 bootstrap estimates $\{\widehat{\theta}^*_b : b=1,...,B_0\}$ and compute $\widehat{\gamma}_{2B_0}$ as defined in (2.6) with B replaced by B_0.

Step 3. Take the desired number of bootstrap repetitions, B^*, to equal $B^* = \max\{B_1, B_0\}$, where

$$B_1 = \text{int}(2{,}500\chi^2_{1-\tau}(2+\widehat{\gamma}_{2B_0})/pdb^2). \quad (2.16)$$

If $\widehat{\gamma}_{2B_0} \leq 0$, then $B^* = B_0$ and one computes $\widehat{se}_{B^*}$ using the B_0 bootstrap estimates $\{\widehat{\theta}^*_b : b = 1, ..., B_0\}$ calculated in Step 2. If $\widehat{\gamma}_{2B_0} > 0$, one has to compute $B_1 - B_0$ additional bootstrap estimates $\{\widehat{\theta}^*_b : b = B_0+1, ..., B_1\}$ before computing $\widehat{se}_{B^*}$.

Using the three-step method above, as $pdb \to 0$ or $\tau \to 0$, we have $B_0 \to \infty$, $\widehat{\gamma}_{2B_0} \to \gamma_2$ in probability and almost surely, $B_1 \to \infty$ provided $\gamma_2 > -2$, and $B_0 \leq B^* \to \infty$. The justification of the above method is that as $pdb \to 0$, we have $B_1 \to \infty$ and

$$P^*\left(100\frac{|\widehat{se}_{B_1} - \widehat{se}_\infty|}{\widehat{se}_\infty} \leq pdb\right) \to 1 - \tau \quad (2.17)$$

provided $\gamma_2 > -2$. We stress that B_1 depends on *pdb* in (2.17) via (2.16). Equation (2.17) implies that the three-step method attains precisely the specified accuracy asymptotically using "small *pdb*" asymptotics when $\gamma_2 \geq 0$. Of course, if $\gamma_2 < 0$, then $B^* = B_0 > B_1$ with probability that goes to one as $pdb \to 0$ and the accuracy of $\widehat{se}_{B^*}$ for approximating $\widehat{se}_\infty$ exceeds that of (pdb, τ). (This is a consequence of the fact that it would be silly to throw away the extra $B_0 - B_1$ bootstrap estimates that have already been calculated in Step 2.) Because one normally specifies a small value of *pdb*, the asymptotic result (2.17) should be indicative of the relevant non-zero *pdb* behavior of the three-step method. The simulation results of Sections 2.4 and 2.6 are designed to examine this. We note that the asymptotics used here are completely analogous to large sample size asymptotics with *pdb* driving B_1 to infinity as $pdb \to 0$ and B_1 playing the role of the sample size.

When $\tau = .05$, equations (2.15) and (2.16) become

$$B_0 = \text{int}(19{,}200/pdb^2) \text{ and } B_1 = \text{int}(9{,}600(2+\widehat{\gamma}_{2B_0})/pdb^2). \quad (2.18)$$

For illustrative purposes, Table 2 provides values B_1 that correspond to several values of $\widehat{\gamma}_{2B_0}$ and *pdb*, with $\tau = .05$. The values of *pdb* considered are 20 (moderately accurate), 10 (accurate), and 5 (very accurate). Table 2 indicates that the necessary B_1 values increase very quickly as the desired level of accuracy increases.

The three-step method discussed above is based on a scalar parameter θ_0. In most applications, however, one has a vector of unknown parameters. In this case, one can apply the three-step method to several parameters of interest, or all the parameters in the model, to obtain $B^*_1, ..., B^*_\omega$, say, when considering ω parameters, and then take B^* to equal the maximum of these values. Then, B^* is the number of bootstrap repetitions needed to obtain the desired accuracy for all of the ω bootstrap standard error estimates (where accuracy is defined as above parameter by parameter).

2.4. Monte Carlo Simulations

In this section we evaluate the performance of the three-step method introduced in Section 2.3. The proposed method is justified by the limit result of (2.17). We wish to see whether this limit result is indicative of finite sample behavior for a range of values of *pdb* and τ in a standard econometric model. More specifically, given several (pdb, τ) combinations, we want to see how close $P^*(100|\widehat{se}_{B_1} - \widehat{se}_\infty|/\widehat{se}_\infty \leq pdb)$ is to $1 - \tau$.

Note that we focus initially on B_1 rather than B^* because, for B_1, equation (2.17) implies that $P^*(100|\widehat{se}_{B_1} - \widehat{se}_\infty|/\widehat{se}_\infty \leq pdb)$ should be approximately equal to $1 - \tau$, whereas for B^* equation (2.17) only implies the less precise result that $P^*(100|\widehat{se}_{B^*} - \widehat{se}_\infty|/\widehat{se}_\infty \leq pdb)$ should be approximately greater than or equal to $1 - \tau$. Of course, our interest ultimately is in the performance of B^*.

The model we consider is the linear regression model

$$y_i = x_i'\beta + u_i \text{ for } i = 1, ..., n, \tag{2.19}$$

where $n = 25$, $X_i = (y_i, x_i')'$ are iid over $i = 1, ..., n$, $x_i = (1, x_{1i}, ..., x_{5i})' \in R^6$, $(x_{1i}, ..., x_{5i})$ are mutually independent normal random variables, x_i is independent of u_i, and $Eu_i = 0$. The simulation results are invariant with respect to the means and variances of $(x_{1i}, ..., x_{5i})$, the variance of u_i, and the value of the regression parameter β, so we need not be specific as to their values. For reasons discussed below, we consider three error distributions: standard normal (denoted $N(0, 1)$), t with five degrees of freedom (denoted t_5), and chi-squared with five degrees of freedom shifted to have mean zero (denoted χ_5^2).

We estimate β by least squares (LS). We focus attention on bootstrap standard error estimates for the LS estimator of the first slope coefficient. Thus, the parameter θ of Sections 2.1–2.3 is β_2, the second element of β.

The LS estimator of θ is a linear combination of the errors $\{u_i : i \leq n\}$. Thus, for normal errors, the coefficient of excess kurtosis of the LS estimator of θ is zero. The crucial parameter γ_2, however, is the coefficient of excess kurtosis of the (discrete) *bootstrap* distribution of the LS estimator of θ. In general, the parameter γ_2 depends on the sample and need not equal zero. Nevertheless, the value of γ_2 will tend to be close to zero for normal errors for most samples, because the bootstrap distribution mimics the true distribution of the LS estimator. Correspondingly, for fat-tailed error distributions, the value of γ_2 will tend to be large for most samples.

To obtain samples for which γ_2 is close to zero, we consider normal errors. To obtain samples with larger values of γ_2, we consider the fat-tailed error distributions t_5 and χ_5^2. The t_5 and χ_5^2 distributions have similar tail behavior and generate samples with similar values of γ_2. The t_5 and normal distributions are symmetric, whereas the χ_5^2 distribution is highly skewed. The results for the χ_5^2 error distribution are used to determine whether skewness of the error distribution has an impact in finite samples on the performance of our three-step procedure for determining B. The results of Section 2.2 establish that skewness has no effect asymptotically.

We simulate 20 different samples from each error distribution because the value of γ_2 varies with the sample. For each of the 20 samples drawn (for a given distribution of u_i), we compute the LS estimate $\widehat{\theta}$ and the ideal bootstrap standard error estimate $\widehat{se}_\infty$ defined in (2.2). We accomplish the latter by employing 250,000 bootstrap repetitions (each of sample size 25). We explicitly assume that 250,000 is close enough to infinity to accurately obtain $\widehat{se}_\infty$.

Next, we run 2,500 Monte Carlo repetitions for each of the 20 samples, for a total of 50,000 repetitions. In each Monte Carlo repetition, we compute $\widehat{se}_{B_1}$ and $\widehat{se}_{B^*}$ for the LS estimate of the first slope coefficient following the three-step procedure outlined in Section 2.3. These calculations are made for several combinations of *pdb* (viz., 20%, 10%, and 5%) and $1-\tau$ (viz., .90, .95, and .975). For each repetition and each (pdb, τ) combination, we determine whether or not the estimate $\widehat{se}_{B_1}$ satisfies

$$100\frac{|\widehat{se}_{B_1} - \widehat{se}_\infty|}{\widehat{se}_\infty} < pdb. \tag{2.20}$$

We call the fraction of times that this condition is satisfied, out of the 2,500 repetitions, the *empirical level* based on B_1 bootstrap repetitions. The empirical level based on B^* bootstrap repetitions is defined analogously. For each (*pdb*, τ) combination and each sample, we compute the empirical levels based on B_1 and B^* bootstrap repetitions. The three-step method of Section 2.3 is considered to perform well if the empirical level based on B_1 bootstrap repetitions is close to $1-\tau$, or if the empirical level based on B^* bootstrap repetitions is close to, or greater than, $1-\tau$.

The results from this set of experiments are reported in Table 3 for the $N(0,1)$ and t_5 error distributions. The numbers reported in Table 3 are averages over the 20 samples. (For example, Med is the average median over the 20 samples.) Results for the χ^2_5 error distribution are almost the same as those for the t_5. In consequence, we do not report the χ^2_5 results. They indicate that asymmetry of the errors is not an important factor for the performance of the three-step method.

Table 3(A) shows that the empirical levels are very close to their asymptotic counterparts for the experiment with the $N(0,1)$ error distribution,. This is true even though the bootstrap distribution of the LS estimate with only 25 observations in the sample can be far away from its asymptotic normal distribution. Note that the empirical levels for the more stringent bounds (i.e., smaller *pdb*'s) and higher probabilities (i.e., higher $1-\tau$'s) are closer to the asymptotic levels. The reason is that the asymptotic approximation improves as B_1, or B^*, increases. Smaller *pdb* values and/or larger $1-\tau$ values lead to larger B_1 and B^* values and, hence, better performance.

The average of the γ_2 values over the 20 samples used in Table 3(A) (computed using 250,000 bootstrap repetitions for each sample) is .37. The mean (over 2,500 simulation repetitions) of the estimator $\widehat{\gamma}_{2B_0}$ averaged over the 20 samples, as reported in Table 3(A), is markedly lower than .37 when B_0 is small (or equivalently, when *pdb* is large). This downward bias of $\widehat{\gamma}_{2B_0}$ leads to B_1 and B^* values that are smaller than desired. In turn, this leads to empirical levels based on B_1 and B^* bootstrap repetitions that are less than $1-\tau$ when B_0 is small. For larger values of B_0 (which occur with smaller *pdb* values), this bias vanishes and the empirical levels are closer to $1-\tau$. The problem of underestimating γ_2 stems from the fact that neither the numerator nor the denominator of the estimator $\widehat{\gamma}_{2B_0}$ in (2.6) is an unbiased estimator of its population counterpart, although both are consistent estimators.

Note that there is significant variation in the values of B_1 over the various (*pdb*, τ) combinations in Table 3(A). The mean values of B_1 range between 37 and 11,257. The corresponding values for B^* are very similar, because $\gamma_2 > 0$ for all 20 samples, $\widehat{\gamma}_{2B_0}$ is positive or close to zero for the vast majority of repetitions, and $B^* = B_1$ whenever $\widehat{\gamma}_{2B_0} > 0$. If one is satisfied with a modest percentage deviation (e.g., *pdb* = 10%), then the required number of bootstrap repetitions is not very large. On the other hand, if one sets a very stringent percentage deviation (e.g., *pdb* = 5%) and a very high probability (e.g., $1 - \tau = .975$), then the number of bootstrap repetitions needed to achieve this level of accuracy is large.

Table 3(B) presents the results based on t_5 errors. The average value of γ_2 over the 20 samples with t_5 errors is 1.26, which is noticeably larger than the value of .37 for normal errors. In Table 3(B), the empirical levels based on B_1 and B^* bootstrap repetitions are lower than $1 - \tau$ and lower than their values in Table 3(A). Nevertheless, the same basic pattern is observed as in Table 3(A). That is, the difference between the empirical levels and $1-\tau$ are largest when B_0 is small, which corresponds to *pdb* being large. When B_0 is small, $\widehat{\gamma}_{2B_0}$ is markedly downward biased and its bias is greater in magnitude than in Table 3(A). This causes B_1 and B^* to be smaller than desired by a greater magnitude than in Table 3(A).

Overall, the empirical level results of Table 3(B) are not as good as those of Table 3(A). Nevertheless, the three-step method still performs quite well with t_5 errors. The largest deviation of an empirical level based on B_1 repetitions from its asymptotic counterpart is .052 and for all other (*pdb*, τ) combinations the deviations are less than half as large. Furthermore, the deviations based on B^* are smaller than those based on B_1.

2.5. A Bootstrap Bias-corrected Estimator of γ_2

The simulation results of Section 2.4 suggest that the performance of the three-step method of Section 2.3 could be improved, especially when γ_2 is large, if a less-biased estimator of γ_2 than $\widehat{\gamma}_{2B_0}$ is employed. In this section, we specify such an estimator.

The iid sample of B_0 bootstrap estimates of θ_0 in Step 2 of the three-step method of Section 2.3 is $\mathbf{\Theta}^* = (\widehat{\theta}_1^*, ..., \widehat{\theta}_{B_0}^*)$. By definition, γ_2 is the coefficient of excess kurtosis of the distribution of $\widehat{\theta}_b^*$ for any $b = 1, ..., B_0$. For present purposes, we think of $(\widehat{\theta}_1^*, ..., \widehat{\theta}_{B_0}^*)$ as being an original sample and $\widehat{\gamma}_{2B_0}$ as being an estimator based on this sample that we want to bootstrap bias correct.

Let $\widehat{G}$ denote the empirical distribution of $(\widehat{\theta}_1^*, ..., \widehat{\theta}_{B_0}^*)$. A bootstrap sample $\mathbf{\Theta}^{**} = (\widehat{\theta}_1^{**}, ..., \widehat{\theta}_{B_0}^{**})$ is a random sample of size B_0 drawn from $\widehat{G}$. Let $\widehat{\gamma}_2(\mathbf{\Theta}^{**})$ denote the estimate $\widehat{\gamma}_{2B_0}$ of γ_2 computed using the bootstrap sample $\mathbf{\Theta}^{**}$, rather than the original sample Θ^*. That is,

$$(2.21) \qquad \widehat{\gamma}_2(\Theta^{**}) = \frac{\frac{1}{B_0-1}\sum_{b=1}^{B_0}\left(\widehat{\theta}_b^{**} - \frac{1}{B_0}\sum_{j=1}^{B_0}\widehat{\theta}_j^{**}\right)^4}{\left(\frac{1}{B_0-1}\sum_{b=1}^{B_0}\left(\widehat{\theta}_b^{**} - \frac{1}{B_0}\sum_{j=1}^{B_0}\widehat{\theta}_j^{**}\right)^2\right)^2} - 3.$$

Note that $\widehat{\gamma}_{2B_0} = \widehat{\gamma}_2(\Theta^*)$.

The "ideal" bootstrap estimate of the bias of $\widehat{\gamma}_{2B_0}$ for estimating γ_2 is

$$(2.22) \qquad E^{**}\widehat{\gamma}_2(\Theta^{**}) - \widehat{\gamma}_{2B_0},$$

where E^{**} denotes expectation with respect to the randomness in Θ^{**}, e.g., see Efron and Tibshirani (1993, eqn. (10.2), p. 125). The "ideal" bootstrap bias-corrected estimate $\widehat{\gamma}_{2B_0\infty}$ of γ_2 is

$$(2.23) \qquad \widehat{\gamma}_{2B_0\infty} = \widehat{\gamma}_{2B_0} - (E^{**}\widehat{\gamma}_2(\Theta^{**}) - \widehat{\gamma}_{2B_0}) = 2\widehat{\gamma}_{2B_0} - E^{**}\widehat{\gamma}_2(\Theta^{**}).$$

Analytic calculation of the ideal bootstrap bias-corrected estimate of γ_2 is intractable. Instead we approximate it using bootstrap simulations. Consider R independent bootstrap samples $\{\Theta_r^{**} : r = 1, ..., R\}$, where each bootstrap sample $\Theta_r^{**} = (\widehat{\theta}_{1r}^{**}, ..., \widehat{\theta}_{B_0r}^{**})$ is a random sample of size B_0 drawn from $\widehat{G}$. The corresponding R bootstrap estimates of γ_2 are $\widehat{\gamma}_2(\Theta_r^{**})$ for $r = 1, ..., R$. The bootstrap bias-corrected estimator $\widehat{\gamma}_{2B_0R}$ of γ_2 for R bootstrap repetitions is

$$(2.24) \qquad \widehat{\gamma}_{2B_0R} = 2\widehat{\gamma}_{2B_0} - \frac{1}{R}\sum_{r=1}^{R}\widehat{\gamma}_2(\Theta_r^{**}).$$

Now, the three-step method of Section 2.3 can be altered by (i) adding a step between Steps 2 and 3 in which $\widehat{\gamma}_{2B_0R}$ is calculated and (ii) replacing $\widehat{\gamma}_{2B_0}$ in Step 3 by $\widehat{\gamma}_{2B_0R}$. The added step is summarized as follows:

Step 2(b). Simulate R bootstrap samples $\{\Theta_r^{**} : r = 1, ..., R\}$, compute R bootstrap estimates $\{\widehat{\gamma}_2(\Theta_r^{**}) : r = 1, ..., R\}$ from these samples using (2.21), and compute $\widehat{\gamma}_{2B_0R}$ from these bootstrap estimates and $\widehat{\gamma}_{2B_0}$ using (2.24).

We refer to the new procedure as the *bias-corrected three-step method* for determining B for bootstrap standard error estimates.

The computational requirements of Step 2(b) are quite modest. Step 2(b) requires that one simulate R bootstrap samples and calculate the simple closed form expressions for $\widehat{\gamma}_2(\Theta_r^{**})$ for $r = 1, ..., R$. For example, when B_0 is 192 (which corresponds to $(pdb, \tau) = (10, .05)$) and $R = 400$, the computational time is only about four seconds using a Sun Sparc-20 computer. Note that the computational requirements of Step 2(b) are the same no matter how difficult the computation of $\widehat{\theta}$ is and no matter how large the original sample size n is. Thus, if a single bootstrap estimate of $\widehat{\theta}$ takes several minutes or several hours to compute, the time required to carry out Step 2(b) is a small fraction of the total computational time.

2.6. Monte Carlo Simulations for the Bias-corrected Three-step Method

Here we evaluate the performance of the bias-corrected three-step method of Section 2.5 via Monte Carlo simulation. Table 4 reports simulation results for the bias-corrected three-step method of determining B for the linear regression model of Section 2.4 with t_5 errors. That is, the results of Table 4 are analogous to those of Table 3(B) except that the three-step method is replaced by the bias-corrected three-step method. We only consider the t_5 errors because they yield the worst results of the three error distributions considered in Section 2.4. The number of repetitions, R, used in the bootstrap bias correction is taken to be 407. This number is chosen, somewhat arbitrarily, to be a value that yields a reasonable tradeoff between computational time for our simulation experiment and accuracy of the bootstrap bias-corrected estimator.

The results of Table 4 show a significant improvement in the performance of the three-step method when it is augmented by a bias-correction of $\widehat{\gamma}_{2B_0}$. Most of the empirical levels are very close to their theoretical counterparts. The largest deviation is .025 and all the rest are about half as large or less. The results in the "Mean" column for $\widehat{\gamma}_{2B_0R}$ indicate that the bias of the bias-corrected estimator $\widehat{\gamma}_{2B_0R}$ is much smaller than that of $\widehat{\gamma}_{2B_0}$ in Table 3(B). The "Mean" numbers of bootstrap repetitions B_1 are larger than in Table 3(B) due to the bias-correction, but the increase is not substantial. Analogous results hold for B^*.

We conclude that the bias-corrected three-step method yields a noticeable improvement over the three-step method in cases where γ_2 is large. The computational cost of the bias-correction is minimal in absolute terms. Also, it is minimal relative to the total computational cost for calculating the bootstrap standard error estimate $\widehat{se}_{B^*}$ whenever $\widehat{\theta}$ is difficult to compute. Thus, we recommend use of the bias-corrected three-step method.

3. Confidence Interval Results

In this section, we consider the problem of choosing the number of bootstrap repetitions B for percentile t confidence intervals. The first three subsections deal with symmetric two-sided confidence intervals and the last subsection extends the results to equal-tailed two-sided and one-sided confidence intervals.

3.1. Notation and Definitions

We begin by introducing some notation and definitions. As in the previous section, $\mathbf{X}$ denotes the observed data and $\widehat{\theta} = \widehat{\theta}(\mathbf{X})$ is an estimator of an unknown scalar parameter θ_0. We wish to construct a confidence interval for θ_0 of (approximate) confidence level $100(1-\alpha)\%$ for some $0 < \alpha < 1$. Here we consider symmetric confidence intervals about $\widehat{\theta}$. (Note that symmetric confidence intervals typically yield greater coverage accuracy as measured by higher order expansions than equal-tailed confidence intervals; see Hall (1992, Secs. 3.5 and 3.6).)

We assume that the normalized estimator $n^\kappa(\widehat{\theta} - \theta_0)$ has an asymptotic normal distribution as $n \to \infty$. (Adjustments for the non-normal case are specified below.) In many cases of interest, $\kappa = 1/2$. We allow for $\kappa \neq 1/2$, however, to cover non-parametric estimators, such as nonparametric estimators of a density or regression function at a point. Let $\widehat{\sigma} = \widehat{\sigma}(\mathbf{X})$ denote a consistent estimator of the asymptotic standard error of $n^\kappa(\widehat{\theta} - \theta_0)$. Let

$$T = n^\kappa(\widehat{\theta} - \theta_0)/\widehat{\sigma} \tag{3.1}$$

denote the t statistic for testing whether $\theta = \theta_0$. The t statistic has an asymptotic standard normal distribution when the true parameter is θ_0.

The "theoretical" symmetric percentile t confidence interval of confidence level $100(1-\alpha)\%$ is

$$J_{SY} = [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}k_\alpha, \widehat{\theta} + n^{-\kappa}\widehat{\sigma}k_\alpha], \tag{3.2}$$

where k_α is the solution to

$$P(|T| \leq k_\alpha) = 1 - \alpha. \tag{3.3}$$

By definition of k_α, J_{SY} has exact confidence level $100(1-\alpha)\%$. In practice, however, one typically does not know k_α. We consider using the bootstrap to estimate k_α.

Define a bootstrap sample $\mathbf{X}^* = (X_1^*, ..., X_n^*)'$ and a bootstrap estimator $\widehat{\theta}^* = \widehat{\theta}(\mathbf{X}^*)$ as in the previous section. Let $\widehat{\sigma}^* = \widehat{\sigma}(\mathbf{X}^*)$ denote the asymptotic standard error estimator based on the bootstrap sample $\mathbf{X}^*$. Let $T^* = n^\kappa(\widehat{\theta}^* - \widehat{\theta})/\widehat{\sigma}^*$ denote the bootstrap t statistic based on $\mathbf{X}^*$. Let $\widehat{k}_{\alpha,\infty}$ denote the *ideal bootstrap estimate* of k_α. Because the bootstrap statistic T^* has a discrete distribution (at least for the nonparametric bootstrap), there typically is no value $\widehat{k}_{\alpha,\infty}$ that satisfies the equation $P^*(|T^*| \leq \widehat{k}_{\alpha,\infty}) = 1 - \alpha$ exactly, where $P^*(\cdot)$ denotes probability with respect to the bootstrap sample $\mathbf{X}^*$ conditional on the original sample $\mathbf{X}$. Thus, to be precise, we define $\widehat{k}_{\alpha,\infty} = \inf\{k : P^*(|T^*| \leq k) \geq 1 - \alpha\}$.

The ideal bootstrap symmetric percentile t confidence interval of approximate confidence level $100(1-\alpha)\%$ is

$$\widehat{J}_{SY,\infty} = [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,\infty}, \widehat{\theta} + n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,\infty}]. \tag{3.4}$$

Analytic calculation of the ideal bootstrap estimate $\widehat{k}_{\alpha,\infty}$ of the critical point k_α is usually intractable. Nevertheless, one can approximate $\widehat{k}_{\alpha,\infty}$ using bootstrap simulations.

As above, consider B iid bootstrap samples $\{\mathbf{X}_b^* : b = 1, ..., B\}$, each with the same distribution as $\mathbf{X}^*$, and the corresponding bootstrap statistics $\widehat{\theta}_b^*$ $(= \widehat{\theta}(\mathbf{X}_b^*))$, $\widehat{\sigma}_b^*$ $(= \widehat{\sigma}(\mathbf{X}_b^*))$, and $T_b^* = n^\kappa(\widehat{\theta}_b^* - \widehat{\theta})/\widehat{\sigma}_b^*$ for $b = 1, ..., B$. Let $\{|T^*|_{B,b} : b = 1, ..., B\}$ denote the ordered sample of the absolute values of T_b^*.

Following Hall (1992, p. 307), we choose B not to be just any positive integer, but one that satisfies $\nu/(B+1) = 1 - \alpha$ for some positive integer ν. (This has advantages in terms of the unconditional coverage probability of the resultant confidence interval; see Hall (1992, p. 307).) Then, the bootstrap estimate of k_α based on B bootstrap repetitions is defined to be

$$\hat{k}_{\alpha,B} = |T^*|_{B,\nu}. \tag{3.5}$$

That is, $\hat{k}_{\alpha,B}$ is the ν-th order statistic of $\{|T_b^*| : b = 1, ..., B\}$. Furthermore, the bootstrap symmetric percentile t confidence interval of approximate confidence level $100(1-\alpha)\%$ based on B bootstrap repetitions is

$$\hat{J}_{SY,B} = [\hat{\theta} - n^{-\kappa}\hat{\sigma}\hat{k}_{\alpha,B}, \hat{\theta} + n^{-\kappa}\hat{\sigma}\hat{k}_{\alpha,B}]. \tag{3.6}$$

Note that B can be chosen as in the previous paragraph only if α is rational. We assume therefore that

$$\alpha = \alpha_1/\alpha_2 \tag{3.7}$$

for some positive integers α_1 and α_2 (with no common integer divisors). Then,

$$B = \alpha_2 a - 1 \text{ and } \nu = (\alpha_2 - \alpha_1)a \tag{3.8}$$

for some positive integer a. For example, if $\alpha = .05$, then $\alpha_1 = 1$, $\alpha_2 = 20$, $B = 20a-1$, and $\nu = 19a$ for some integer $a > 0$. That is, $B = 19, 39, 59$, etc. If $\alpha = .10$, then $\alpha_1 = 1$, $\alpha_2 = 10$, $B = 10a - 1$, and $\nu = 9a$ for some integer $a > 0$.

3.2. A Three-step Method for Determining the Number of Bootstrap Repetitions

In this section, we introduce a three-step method for determining B for the bootstrap confidence interval $\hat{J}_{SY,B}$ defined above. Our main interest is determining B such that $\hat{J}_{SY,B}$ is close to the ideal bootstrap confidence interval $\hat{J}_{SY,\infty}$. A secondary interest is in the unconditional coverage probability of $\hat{J}_{SY,B}$ (where "unconditional" refers to the randomness in both the data *and* the simulations).

Our primary interest is the former, because the simulated random variables are ancillary with respect to the parameter θ_0. Hence, the principle of ancillarity or conditionality (e.g., see Kiefer (1982) and references therein) implies that we should seek a confidence interval that has confidence level that is (approximately) $100(1-\alpha)\%$ conditional on the simulation draws. To obtain such an interval, we need to choose B to be sufficiently large that $\hat{J}_{SY,B}$ is close to $\hat{J}_{SY,\infty}$. Otherwise, two researchers using the same data and the same statistical method could reach different conclusions due only to the use of different simulation draws.

We measure the closeness of $\hat{J}_{SY,B}$ to $\hat{J}_{SY,\infty}$ by comparing the endpoints of the two intervals. The percentage deviation of the endpoints of $\hat{J}_{SY,B}$ to the endpoints of $\hat{J}_{SY,\infty}$ is

$$100\frac{|n^{-\kappa}\hat{\sigma}\hat{k}_{\alpha,B} - n^{-\kappa}\hat{\sigma}\hat{k}_{\alpha,\infty}|}{n^{-\kappa}\hat{\sigma}\hat{k}_{\alpha,\infty}} = 100\frac{|\hat{k}_{\alpha,B} - \hat{k}_{\alpha,\infty}|}{\hat{k}_{\alpha,\infty}}. \tag{3.9}$$

We could also measure the closeness of $\hat{J}_{SY,B}$ to $\hat{J}_{SY,\infty}$ by comparing the lengths of the two intervals. This is a particularly appropriate measure in the present case because each interval is symmetric about the same value $\hat{\theta}$. Furthermore, the length of a two-sided confidence interval is of inherent interest because it directly reflects the precision of the interval estimate.

Denote the length of a confidence interval J by $L(J)$. We have

$$L(\widehat{J}_{SY,B}) = 2n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B} \text{ and } L(\widehat{J}_{SY,\infty}) = 2n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,\infty}. \tag{3.10}$$

The percentage deviation of $L(\widehat{J}_{SY,B})$ to $L(\widehat{J}_{SY,\infty})$ is

$$100\frac{|L(\widehat{J}_{SY,B}) - L(\widehat{J}_{SY,\infty})|}{L(\widehat{J}_{SY,\infty})} = 100\frac{|\widehat{k}_{\alpha,B} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}}. \tag{3.11}$$

Thus, for the symmetric confidence interval, our measure of the closeness of $\widehat{J}_{SY,B}$ to $\widehat{J}_{SY,\infty}$, based on the percentage deviation of the endpoints, is equivalent to a measure based on the percentage deviation of the length of the confidence interval. The former measure is applicable more generally, however, because in the case of one-sided confidence intervals the lengths of all confidence intervals are infinite.

As in the previous section, let $1-\tau$ denote a probability close to one, such as .95. Let *pdb* be a bound on the percentage deviation of the endpoints of $\widehat{J}_{SY,B}$ to the endpoints of $\widehat{J}_{SY,\infty}$. We want to determine $B = B(pdb, \tau)$ such that

$$P^*\left(100\frac{|\widehat{k}_{\alpha,B} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}} \leq pdb\right) = 1-\tau. \tag{3.12}$$

That is, we want to specify a method of determining B to obtain a desired level of accuracy *pdb* with probability (approximately) equal to $1-\tau$.

We introduce a three-step method of doing so. The method relies on an estimator of the reciprocal of a density function at a point, which appears in the asymptotic distribution of the sample quantile $\widehat{k}_{\alpha,B}$. For this, we use Siddiqui's (1960) estimator (analyzed by Bloch and Gastwirth (1968) and Hall and Sheather (1988)) with a plug-in estimator of the bandwidth parameter that is chosen to maximize the higher order asymptotic coverage probability of the resultant confidence interval, as calculated by Hall and Sheather (1988). To reduce the noise of the plug-in estimator, we take advantage of the fact that we know the asymptotic value of the density and use it to generate our estimators of the unknown coefficients in the plug-in formula.

The three-step method is defined as follows:

Step 1. Compute a preliminary number of bootstrap repetitions B_0 via

$$\begin{aligned} B_0 &= \alpha_2 a_0 - 1, \text{ where} \\ a_0 &= \text{int}\left(\frac{2,500\alpha(1-\alpha)\chi^2_{1-\tau}}{z^2_{1-\alpha/2}\phi^2(z_{1-\alpha/2})pdb^2\alpha_2}\right) \text{ and } \alpha = \alpha_1/\alpha_2. \end{aligned} \tag{3.13}$$

Step 2. Simulate B_0 bootstrap t statistics $\{T^*_b : b = 1, ..., B_0\}$; order the absolute values of the bootstrap t statistics, which are denoted $\{|T^*|_{B_0,b} : b = 1, ..., B_0\}$; and calculate $\nu_0 = (\alpha_2 - \alpha_1)a_0$, $\widehat{m} = \text{int}(c_\alpha B_0^{2/3})$, $\widehat{k}_{\alpha,B_0} = |T^*|_{B_0,\nu_0}$, $|T^*|_{B_0,\nu_0-\widehat{m}}$, and $|T^*|_{B_0,\nu_0+\widehat{m}}$, where

$$c_\alpha = \left(\frac{6z^2_{1-\alpha/2}\phi^2(z_{1-\alpha/2})}{2z^2_{1-\alpha/2}+1}\right)^{1/3}. \tag{3.14}$$

Step 3. Take the desired number of bootstrap repetitions, B^*, to equal $B^* = \max\{B_0, B_1\}$, where

$$B_1 = \alpha_2 a_1 - 1 \text{ and}$$
$$(3.15) \quad a_1 = \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{\widehat{k}^2_{\alpha,B_0} pdb^2 \alpha_2}\left(\frac{B_0}{2\widehat{m}}\right)^2 \left(|T^*|_{B_0,\nu_0+\widehat{m}} - |T^*|_{B_0,\nu_0-\widehat{m}}\right)^2\right).$$

Note that $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-th quantile of a standard normal distribution, $\phi(\cdot)$ denotes the standard normal density function, and $\chi^2_{1-\tau}$ denotes the $(1-\tau)$-th quantile of a chi-square distribution with one degree of freedom.

Having determined B^*, one simulates $B^* - B_0$ (≥ 0) additional bootstrap t statistics $\{T^*_b : b = B_0 + 1, ..., B^*\}$ and orders the absolute values of the B^* bootstrap t statistics, which are denoted $\{|T^*|_{B^*,b} : b = 1, ..., B^*\}$. The desired cutoff value, ν^*, and the desired critical point, $\widehat{k}_{\alpha,B^*}$, are then given by

$$(3.16) \quad \begin{aligned} \nu^* &= \max\{\nu_0, \nu_1\}, \ \nu_1 = (\alpha_2 - \alpha_1)a_1, \text{ and} \\ \widehat{k}_{\alpha,B^*} &= |T^*|_{B^*,\nu^*}. \end{aligned}$$

The resulting bootstrap confidence interval, based on B^* bootstrap repetitions, is equal to

$$(3.17) \quad \widehat{J}_{SY,B^*} = [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B^*}, \widehat{\theta} + n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B^*}].$$

In Table 5, we provide the values of B_0, a_0, ν_0, c_α, and $\widehat{m}$ that correspond to common values of α, τ, and *pdb*. Table 5 indicates that B_0 increases significantly as τ decreases and even more so as *pdb* decreases. For example, the combination $(\alpha, pdb, \tau) = (.05, 15, .10)$ requires $B_0 = 119$. In contrast, $(\alpha, pdb, \tau) = (.05, 5, .01)$ requires $B_0 = 2399$. In addition, B_0 increases as α decreases. The pattern of variation of ν_0 and $\widehat{m}$ is the same as that of B_0 except that $\widehat{m}$ decreases as α decreases. This occurs because the height of the "density" that is being estimated decreases as α decreases and it decreases quickly enough to offset the increase in B_0.

To assess the magnitude of the B_1 values generated by our three-step method, we carried out the following procedure: (i) we assumed that the bootstrap distribution of T^* was some specified distribution, viz., $N(0,1)$, t_{10}, t_5, or χ^2_5; (ii) for each (α, pdb, τ) combination in Table 5, we took B_0 draws from the specified distribution and calculated B_1 according to Steps 2 and 3 of the three-step procedure; (iii) we repeated the simulation of part (ii) 5000 times; and (iv) we computed the median, mean, minimum, maximum, and standard deviation of the 5000 values of B_1. For brevity, we do not provide tables of the results, but just summarize them briefly.

The use of the $N(0,1)$ distribution illustrates typical B_1 values when the bootstrap distribution of T^* equals (or is close to) its asymptotic distribution, which is $N(0,1)$. In this case, we found that the ratio of the median B_1 value to B_0 was in the range 1.0–1.1 for most (α, pdb, τ) combinations with $\alpha = .05$ or $.10$. Thus, in this case, the initial choice of B_0 is an accurate starting value to determine B_1.

When we used the t_{10}, t_5, and χ^2_5 distributions, which differ increasingly from the $N(0,1)$ distribution, we found that the ratio of the median B_1 value to B_0 increased significantly. For example, for the t_5 distribution, which has very thick tails, the ratio was in the range 2.3–3.5 for $\alpha = .05$ and 1.8–2.1 for $\alpha = .10$. Thus, when the bootstrap distribution is far from its asymptotic normal distribution noticeably more bootstrap repetitions B_1 are needed than B_0. This illustrates the importance of using a three-step method, which takes account of the actual bootstrap distribution of T^* in Steps 2 and 3 and does not rely on its asymptotic distribution.

To illustrate the magnitude of B_1 for what may be a typical scenario, we consider the t_{10} distribution and specify (α, pdb, τ) to equal $(.05, 10, .05)$. In this case, the median B_1 value is 699. This is larger than the number of bootstrap repetitions often used in the econometrics literature.

As in the previous section, the three-step method introduced here is based on a scalar parameter θ_0. When one is interested in separate confidence intervals for several parameters, say ω parameters, one can apply the three-step method for each of the parameters to obtain $B^*_{(1)}$, $B^*_{(2)}$, ..., $B^*_{(\omega)}$ and take B^* to equal the maximum of these values.

If the asymptotic distribution of T is not normal, then a_0 and c_α in Steps 1 and 2 above have to be adjusted. Suppose the asymptotic distribution of $|T|$ is F, the $(1-\alpha)$-th quantile of F is $q_{1-\alpha}$, and F has a density $f(\cdot)$ with respect to Lebesgue measure at $q_{1-\alpha}$, then $a_0 = \text{int}((10,000\alpha(1-\alpha)\chi^2_{1-\tau})\ /(q^2_{1-\alpha} f^2(q_{1-\alpha}) pdb^2 \alpha_2))$ and c_α is defined as in (5.31) of the Appendix of Proofs. If F depends on unknown parameters, then consistent estimates of these parameters can be used to provide estimates of $q_{1-\alpha}$ and $f(q_{1-\alpha})$ for use in the definitions of a_0 and c_α.

3.3. Asymptotic Justification of the Three-step Method

We now discuss the justification of the three-step method introduced above. The three-step method relies on the fact that $\widehat{k}_{\alpha,B}$ is a sample quantile based on an iid sample of random variables each with distribution given by the bootstrap distribution of $|T^*|$. If the bootstrap distribution of $|T^*|$ was absolutely continuous at $\widehat{k}_{\alpha,\infty}$, then $B^{1/2}(\widehat{k}_{\alpha,B} - \widehat{k}_{\alpha,\infty})$ would be asymptotically normally distributed as $B \to \infty$ for fixed n with asymptotic variance given by $\alpha(1-\alpha)/f^2(\widehat{k}_{\alpha,\infty})$, where $f(\cdot)$ denotes the density of $|T^*|$. (Here and below, we condition on the data and the asymptotics are based on the randomness of the simulations alone.)

But, the bootstrap distribution of $|T^*|$ is a discrete distribution (at least for the nonparametric bootstrap, which is based on the empirical distribution). In consequence, the asymptotic distribution of $B^{1/2}(\widehat{k}_{\alpha,B} - \widehat{k}_{\alpha,\infty})$ as $B \to \infty$ for fixed n is a *pointmass at zero* for all α values except for those in a set of Lebesgue measure zero. (The latter set is the set of values that the distribution function of $|T^*|$ takes on at its points of support.)

Although $|T^*|$ has a discrete distribution in the case of the nonparametric iid bootstrap, its distribution is very nearly continuous even for small values of n. The largest probability π_n of any of its atoms is very small: $\pi_n = n!/n^n \sim (2\pi n)^{1/2} e^{-n}$ provided the original sample $\mathbf{X}$ consists of distinct vectors and distinct bootstrap samples $\mathbf{X}^*$ give rise to distinct values of $|T^*|$ (as is typically the case); see Hall (1992, Appendix I). This suggests that we should consider asymptotics as $n \to \infty$, as well as $B \to \infty$, in order to account for the essentially continuous nature of the distribution of $|T^*|$. If we do so, then $B^{1/2}(\widehat{k}_{\alpha,B} - \widehat{k}_{\alpha,\infty})$ has a nondegenerate asymptotic distribution with asymptotic variance that depends on the value of a density at a point, just as in the case where the distribution of $|T^*|$ is continuous. This is what we do. It is in accord with Hall's (1992, p. 285) view that "for many practical purposes the bootstrap distribution of a statistic may be regarded as continuous."

We note that the (potential) discreteness of T^* significantly increases the complexity of the asymptotic justification of the three-step method given below and its proof.

We now introduce a strengthening of the assumption of asymptotic normality of the t statistic T that is needed for the asymptotic justification of the three-step method. We assume: For some $\xi > 0$ and all sequences of constants $\{x_n : n \geq 1\}$ for which $x_n \to z_{1-\alpha/2}$, we have

$$\begin{aligned} & P(|T| \leq x_n) = P(|Z| \leq x_n) + O(n^{-\xi}) \text{ as } n \to \infty \text{ and} \\ (3.18) \qquad & P^*(|T^*| \leq x_n) = P(|Z| \leq x_n) + O(n^{-\xi}) \text{ as } n \to \infty, \end{aligned}$$

where $Z \sim N(0,1)$. (The assumption on $|T^*|$ is assumed to hold with probability one with respect to the randomness in the data, i.e., with respect to $P(\cdot)$.)

Assumption (3.18) holds whenever the t statistic and the bootstrap t statistic have one-term Edgeworth expansions. This occurs in any context in which the bootstrap delivers higher order improvements in the coverage probability of confidence intervals based on T. The literature on the bootstrap is full of results that establish (3.18) for different t statistics. For example, see Hall (1992, Sec. 3.3 and Ch. 5), Hall and Horowitz (1996), and references therein. When $\kappa = 1/2$ and $\widehat{\sigma}$ is an $n^{1/2}$-consistent estimator of the asymptotic standard error of $\widehat{\theta}$, then (3.18) typically holds with $\xi = 1$. (The $n^{-1/2}$ terms in the Edgeworth expansions of T and T^* typically are even functions of x_n and hence cancel out in the Edgeworth expansions of $|T|$ and $|T^*|$, leaving the order of the first terms of the latter equal to n^{-1}.) One example where (3.18) holds with $\kappa = 1/2$ and $\xi < 1$ is when $\widehat{\theta}$ is a sample quantile and $\widehat{\sigma}$ is an estimator of its asymptotic standard error (which is not $n^{1/2}$-consistent because it involves the nonparametric estimation of a density at a point); see Hall and Sheather (1988) and Hall and Martin (1991). When $\kappa < 1/2$, as occurs with nonparametric estimators $\widehat{\theta}$, then (3.18) typically holds with $\xi < 1$; see Hall (1992, Ch. 4) and references therein.

The discussion above considers letting $B \to \infty$. This is not really appropriate because we want B to be determined endogenously by the three-step method. Rather, we consider asymptotics in which the accuracy measure *pdb* $\to 0$ and this, in turn, forces $B \to \infty$. Thus, the asymptotic justification of the three-step method of choosing B^* is in terms of the limit as *both pdb* $\to 0$ and $n \to \infty$.

We assume that *pdb* $\to 0$ sufficiently slowly that

$$(3.19) \qquad pdb \times n^{\xi} \to \infty \text{ as } n \to \infty,$$

where ξ is as in (3.18).

The asymptotic justification of the three-step method is that

$$P^*\left(100\frac{|\widehat{k}_{\alpha,B_1}-\widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}}\leq pdb\right)\to 1-\tau \quad \text{as } pdb\to 0 \text{ and } n\to\infty, \text{ where}$$

$$\widehat{k}_{\alpha,B_1}=|T^*|_{B_1,\nu_1} \text{ and } \widehat{J}_{SY,B_1}=[\widehat{\theta}-n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B_1},\widehat{\theta}+n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B_1}]. \tag{3.20}$$

As above, the probability $P^*(\cdot)$ denotes probability with respect to the simulation randomness conditional on the infinite sequence of data vectors. Under the assumptions above, this conditional result holds with probability one with respect to the randomness in the data. The proof of (3.20) is given in the Appendix of Proofs.

Equation (3.20) implies that the three-step method attains precisely the desired level of accuracy using "small *pdb* and large *n*" asymptotics when $B^*=B_1>B_0$. When $B^*=B_0>B_1$, then the accuracy of the three-step method exceeds the desired level of accuracy. (This is a consequence of the fact that it would be silly to throw away the extra B_0-B_1 bootstrap estimates in Step 3 that have already been calculated in Step 2.)

3.4. Monte Carlo Simulations for Symmetric Confidence Intervals

In this section, we evaluate the performance of the three-step method introduced in Section 3.2. As in Section 2.4, the purpose of the Monte Carlo experiments reported here is to evaluate whether or not the limit result of (3.20) is indicative of finite sample behavior for a range of values of α, *pdb*, and τ. That is, we want to see how close $P^*(100\mid\widehat{k}_{\alpha,B_1}-\widehat{k}_{\alpha,\infty}\mid/\widehat{k}_{\alpha,\infty}\leq pdb)$ is to $1-\tau$.

As in Section 2.4, we focus our attention on B_1 rather than on B^*, because equation (3.20) implies that for B_1, $P^*(100\mid\widehat{k}_{\alpha,B_1}-\widehat{k}_{\alpha,\infty}\mid/\widehat{k}_{\alpha,\infty}\leq pdb)$ should be approximately equal to $1-\tau$, whereas for B^* equation (3.20) only implies that $P^*(100\mid\widehat{k}_{\alpha,B^*}-\widehat{k}_{\alpha,\infty}\mid/\widehat{k}_{\alpha,\infty}\leq pdb)$ should be approximately greater than, or equal to, $1-\tau$. Nevertheless, as above, the ultimate interest is in the performance of the three-step method based on B^*.

We consider the same linear regression model as in (2.19) with the same three error distributions : $N(0,1)$, t_5, and χ^2_5. Again, we estimate β by LS and focus our attention on the first slope coefficient. Thus, the parameter θ of Sections 3.1–3.3 is β_2 (the second element of β). The standard error estimator $\widehat{\sigma}$ is defined using the standard formula. That is, $\widehat{\sigma}^2$ is the (2,2) term of the matrix $\widehat{\sigma}^2_u(\sum_{i=1,\ldots,25}x_ix_i'/25)^{-1}$, where $\widehat{\sigma}^2_u=e'e/(n-6)$ and e is the vector of LS residuals.

We simulate 250 different samples from each of the three error distributions. For each of the 250 samples, we compute the LS estimate $\widehat{\theta}$ and the standard error estimate $\widehat{\sigma}$. Then, we simulate $\widehat{k}_{\alpha,\infty}$ using 250,000 bootstrap repetitions (each of size 25). As in Section 2.4, we explicitly assume that 250,000 is close enough to infinity to accurately obtain $\widehat{k}_{\alpha,\infty}$. Given $\widehat{\theta}$, $\widehat{\sigma}$, and $\widehat{k}_{\alpha,\infty}$, we calculate the ideal symmetric confidence interval $\widehat{J}_{SY,\infty}$ defined in (3.4) for each of the 250 samples for each error distribution.

Next, we run 2,000 Monte Carlo repetitions for each of the 250 samples for a total of 500,000 repetitions. In each Monte Carlo repetition, we compute $\widehat{J}_{SY,B_1}$, $\widehat{J}_{SY,B^*}$, $\widehat{k}_{\alpha,B_1}$, and $\widehat{k}_{\alpha,B^*}$ using the three-step method of Section 3.2. We make this calculation for several combinations of α (viz., .10 and .05), *pdb* (viz., 15%, 10%, and 5%), and $1-\tau$ (viz., .10 and .05). For each repetition and each (α, pdb, τ) combination, we check whether $\widehat{k}_{\alpha,B_1}$ satisfies

$$(3.21) \qquad 100\frac{|\widehat{k}_{\alpha,B_1} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}} \le pdb,$$

or equivalently, whether $L(\widehat{J}_{SY,B_1})$ satisfies

$$(3.22) \qquad 100\frac{|L(\widehat{J}_{SY,B_1}) - L(\widehat{J}_{SY,\infty})|}{L(\widehat{J}_{SY,\infty})} \le pdb.$$

We call the fraction of times this condition is satisfied, out of the 2,000 repetitions, the empirical level based on B_1. The empirical level based on B^* bootstrap repetitions is computed analogously. In addition, we compute the fraction of times that θ falls within the constructed confidence interval $\widehat{J}_{SY,B_1}$. We call this fraction the *empirical unconditional coverage probability.* The empirical unconditional coverage probability based on B^* bootstrap repetitions is defined analogously.

The three-step method of Section 3.2 is considered to perform well if the empirical levels based on B_1 bootstrap repetitions are close to $1-\tau$. Based on B^* bootstrap repetitions, the method is considered to perform well if the empirical levels are close to, or greater than, $1-\tau$.

The results from this set of experiments are reported in Table 6 for the $N(0,1)$ and t_5 error distributions. The numbers reported in this table are averages over the 250 samples. The results for the χ^2_5 error distribution are very similar to those given in Table 6(B) for the t_5 error distribution in terms of both the empirical levels obtained and the number of bootstrap repetitions B_1 needed. These results show that the high skewness of the χ^2_5 error distribution does not have any effect on the performance of the three-step method. For brevity, we do not report these results.

Table 6(A) shows that the empirical levels are somewhat higher than the corresponding $1-\tau$ values for the experiments with the $N(0,1)$ error distribution. Nevertheless, with low *pdb* (5), the empirical levels are quite close to their asymptotic counterparts.

Table 6(A) indicates that the performance of the three-step method is determined by the number of bootstrap repetitions B_1, or B^*, employed. The (α, pdb, τ) combinations that yield the best results are those that induce a relatively large number of bootstrap repetitions. Thus, the smaller the bound *pdb*, the closer are the empirical levels to their asymptotic counterparts, and the more so, the higher the $1-\tau$ value. For example, for the (.10, 5, .10) combination, the median B_1 value is 1348, while for the combination (.10, 15, .10), it is only 230. As a result, the empirical level for the former case is .907, which is quite close to .900, while for the latter it is .942.

Also, the empirical levels are closer to their asymptotic counterparts for the confidence intervals with lower confidence level $1-\alpha$. This occurs because it is more difficult to estimate the .95 quantile of $|T^*|$ needed for a 95% confidence interval than to estimate the .90 quantile of $|T^*|$ needed for a 90% confidence interval.

Table 6(B) reports the results from the Monte Carlo simulations with the t_5 error distribution. The general picture revealed by Table 6(B) is very similar to that of Table 6(A). The empirical levels are comparable to those reported in Table 6(A). They are somewhat higher than their asymptotic counterparts. The most pronounced difference between the two sets of experiments is that for all (α, pdb, τ) combinations, the number of bootstrap repetitions B_1 is somewhat larger for the experiment with the t_5 error distribution, but not by much. This indicates that even with a relatively small sample size (25 observations) the bootstrap distribution of T^* with a fat-tailed t_5 error distribution is not much different than with a $N(0,1)$ error distribution. Certainly, the bootstrap distribution of T^* based on t_5 errors is far from being a t_5 distribution itself.

We conclude that the three-step method does pretty well in attaining the desired accuracy of the bootstrap endpoints and confidence interval length in relation to their ideal bootstrap counterpoints. The three-step method is slightly conservative, because the accuracy obtained is slightly greater than the nominal accuracy.

Lastly, we consider the empirical unconditional coverage probabilities. In all cases, they are the same whether based on B_1 or B^* bootstrap repetitions. In Table 6(A), they equal .908 or .909 for all cases where $\alpha = .900$ and they equal .957 for all cases where $\alpha = .950$. In Table 6(B), they are in the range .900–.902 for all cases where $\alpha = .900$ and they are in the range .951–.953 for all cases where $\alpha = .950$. Thus, the empirical unconditional coverage probabilities are extremely close to their asymptotic counterparts. This is consistent with Hall's (1986) result that one need not employ a large number of bootstrap repetitions in order to obtain good unconditional coverage probabilities. Nevertheless, our results show that in order to construct confidence intervals whose endpoints, length, and conditional coverage probability are close to that of the ideal bootstrap confidence interval, one does need to employ a relatively large number of bootstrap repetitions.

3.5. Equal-tailed and One-sided Confidence Intervals

We now develop three-step methods for choosing B for the case of equal-tailed and one-sided bootstrap percentile t confidence intervals. We take $\mathbf{X}$, $\widehat{\theta}$, θ_0, $\widehat{\sigma}$, T, $\mathbf{X}^*$, $\widehat{\theta}^*$, $\widehat{\sigma}^*$, T^*, and $\{(\mathbf{X}_b^*, \widehat{\theta}_b^*, T_b^*) : b = 1, ..., B\}$ as in Section 3.1. We assume that the normalized estimator $n^\kappa(\widehat{\theta} - \theta_0)$ has an asymptotic normal distribution as $n \to \infty$ for some $\kappa > 0$. (Adjustments for the non-normal case are provided below.)

The "theoretical" equal-tailed and one-sided percentile t confidence intervals with exact confidence levels $100(1-2\alpha)\%$ and $100(1-\alpha)\%$, respectively, are

$$(3.23) \qquad J_2 = [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}q_{1-\alpha}, \widehat{\theta} - n^{-\kappa}\widehat{\sigma}q_\alpha] \text{ and } J_1 = [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}q_{1-\alpha}, \infty),$$

where q_α is the solution to $P(T \leq q_\alpha) = \alpha$ for any $\alpha \in (0,1)$.

We use the bootstrap to estimate the quantiles q_α and $q_{1-\alpha}$. Let $\widehat{q}_{\alpha,\infty}$ denote the *ideal bootstrap estimate* of q_α, i.e., the α-th quantile of the distribution of T^*.

Again, to be precise, we define $\widehat{q}_{\alpha,\infty} = \inf\{q : P^*(T^* \leq q) \geq \alpha\}$. The ideal bootstrap equal-tailed and one-sided percentile t confidence intervals of approximate confidence levels $100(1-2\alpha)\%$ and $100(1-\alpha)\%$, respectively, are

$$\begin{aligned} \widehat{J}_{2,\infty} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,\infty}, \widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{\alpha,\infty}] \text{ and} \\ \widehat{J}_{1,\infty} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,\infty}, \infty). \end{aligned} \tag{3.24}$$

We approximate $\widehat{q}_{1-\alpha,\infty}$ and $\widehat{q}_{\alpha,\infty}$ using bootstrap simulations. Let $\{T^*_{B,b} : b = 1, ..., B\}$ denote the ordered sample of $\{T^*_b : b = 1, ..., B\}$. We assume that α is rational. That is, $\alpha = \alpha_1/\alpha_2$ for some positive integers α_1 and α_2 (with no common integer divisors). We choose B to be a positive integer that satisfies $\nu/(B+1) = 1-\alpha$ and $\eta/(B+1) = \alpha$ for some positive integers ν and η. That is, $B = \alpha_2 a - 1$, $\nu = (\alpha_2 - \alpha_1)a$, and $\eta = \alpha_1 a$, for some positive integer a. For example, if $\alpha = .05$, then $\alpha_1 = 1$, $\alpha_2 = 20$, $B = 20a - 1$, $\nu = 19a$, and $\eta = a$ for some integer $a > 0$.

The bootstrap estimates of $q_{1-\alpha}$ and q_α based on B bootstrap repetitions are defined to be

$$\widehat{q}_{1-\alpha,B} = T^*_{B,\nu} \text{ and } \widehat{q}_{\alpha,B} = T^*_{B,\eta}. \tag{3.25}$$

That is, $\widehat{q}_{1-\alpha,B}$ and $\widehat{q}_{\alpha,B}$ are the ν-th and η-th order statistics of $\{T^*_b : b = 1, ..., B\}$. Then, the bootstrap equal-tailed and one-sided percentile t confidence intervals of approximate confidence levels $100(1-2\alpha)\%$ and $100(1-\alpha)\%$, respectively, based on B bootstrap repetitions are defined to be

$$\begin{aligned} \widehat{J}_{2,B} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B}, \widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{\alpha,B}] \text{ and} \\ \widehat{J}_{1,B} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B}, \infty). \end{aligned} \tag{3.26}$$

We now introduce three-step methods for determining B for the bootstrap confidence intervals $\widehat{J}_{2,B}$ and $\widehat{J}_{1,B}$. We measure the closeness of $\widehat{J}_{h,B}$ to $\widehat{J}_{h,\infty}$ by comparing the endpoints of these intervals for $h = 1, 2$. The percentage deviation of the lower endpoint of $\widehat{J}_{h,B}$ to the lower endpoint of $\widehat{J}_{h,\infty}$ is

$$100\frac{|n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,\infty}|}{n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,\infty}} = 100\frac{|\widehat{q}_{1-\alpha,B} - \widehat{q}_{1-\alpha,\infty}|}{\widehat{q}_{1-\alpha,\infty}} \tag{3.27}$$

for $h = 1, 2$. The percentage deviation of the upper endpoint of $\widehat{J}_{2,B}$ to the upper endpoint of $\widehat{J}_{2,\infty}$ is defined analogously.

As above, let $1 - \tau$ denote a probability close to one, such as .95. Let *pdb* be a bound on the percentage deviation of an endpoint of $\widehat{J}_{h,B}$ to the corresponding endpoint of $\widehat{J}_{h,\infty}$ for $h = 1$ or 2. For the lower endpoint, we want to determine $B = B(pdb, \tau)$ such that

$$P^*\left(100\frac{|\widehat{q}_{1-\alpha,B} - \widehat{q}_{1-\alpha,\infty}|}{\widehat{q}_{1-\alpha,\infty}} \leq pdb\right) = 1 - \tau. \tag{3.28}$$

For the upper endpoint, we want to determine an analogous value of B with $1 - \alpha$ replaced by α.

The three-step method of determining B for $\widehat{J}_{2,B}$ is designed to obtain a given desired level of accuracy pdb for both endpoints, each with probability (approximately) equal to $1-\tau$. The three-step method for the equal-tailed confidence interval $\widehat{J}_{2,B}$ is defined as follows:

Step 1. Compute a preliminary number of bootstrap repetitions B_0 via

$$\begin{aligned} B_0 &= \alpha_2 a_0 - 1, \text{ where} \\ a_0 &= \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{z^2_{1-\alpha}\phi^2(z_{1-\alpha})pdb^2\alpha_2}\right) \text{ and } \alpha = \alpha_1/\alpha_2. \end{aligned} \tag{3.29}$$

Step 2. Simulate B_0 bootstrap t statistics $\{T^*_b : b = 1, ..., B_0\}$; compute the ordered bootstrap t statistics, which are denoted $\{T^*_{B_0,b} : b = 1, ..., B_0\}$; and calculate $\nu_0 = (\alpha_2 - \alpha_1)a_0$, $\eta_0 = \alpha_1 a_0$, $\widehat{m} = \text{int}(c_\alpha B_0^{2/3})$, $\widehat{q}_{1-\alpha,B_0} = T^*_{B_0,\nu_0}$, $T^*_{B_0,\nu_0-\widehat{m}}$, $T^*_{B_0,\nu_0+\widehat{m}}$, $\widehat{q}_{\alpha,B_0} = T^*_{B_0,\eta_0}$, $T^*_{B_0,\eta_0-\widehat{m}}$, $T^*_{B_0,\eta_0+\widehat{m}}$, where

$$c_\alpha = \left(\frac{1.5z^2_{1-\alpha/2}\phi^2(z_{1-\alpha})}{2z^2_{1-\alpha}+1}\right)^{1/3}. \tag{3.30}$$

Step 3. Take the desired number of bootstrap repetitions, B^*, to equal $B^* = \max\{B_0, B_{1\ell}, B_{1u}\}$, where

$$\begin{aligned} B_{1\ell} &= \alpha_2 a_{1\ell} - 1,\ B_{1u} = \alpha_2 a_{1u} - 1, \\ a_{1\ell} &= \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{\widehat{q}^2_{1-\alpha,B_0}pdb^2\alpha_2}\left(\frac{B_0}{2\widehat{m}}\right)^2\left(T^*_{B_0,\nu_0+\widehat{m}} - T^*_{B_0,\nu_0-\widehat{m}}\right)^2\right), \text{ and} \\ a_{1u} &= \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{\widehat{q}^2_{\alpha,B_0}pdb^2\alpha_2}\left(\frac{B_0}{2\widehat{m}}\right)^2\left(T^*_{B_0,\eta_0+\widehat{m}} - T^*_{B_0,\eta_0-\widehat{m}}\right)^2\right). \end{aligned} \tag{3.31}$$

Note that the term in (3.30) that depends on $z_{1-\alpha/2}$ rather than $z_{1-\alpha}$ follows from the formula of Hall and Sheather (1988); see (5.31). It is not a typographical error.

Having determined B^*, one simulates $B^* - B_0$ (≥ 0) additional bootstrap t statistics $\{T^*_b : b = B_0+1, ..., B^*\}$ and orders the values of the B^* bootstrap t statistics, which are denoted $\{T^*_{B^*,b} : b = 1, ..., B^*\}$. For the equal-tailed confidence interval $\widehat{J}_{2,B}$, the desired cutoff values, ν^* and η^*, and the desired critical points, $\widehat{q}_{1-\alpha,B^*}$ and $\widehat{q}_{\alpha,B^*}$, are then given by

$$\begin{aligned} \nu^* &= \max\{\nu_0, \nu_{1\ell}, \nu_{1u}\},\ \nu_{1\ell} = (\alpha_2-\alpha_1)a_{1\ell},\ \nu_{1u} = (\alpha_2-\alpha_1)a_{1u}, \\ \eta^* &= \max\{\eta_0, \eta_{1\ell}, \eta_{1u}\},\ \eta_{1\ell} = \alpha_1 a_{1\ell},\ \eta_{1u} = \alpha_1 a_{1u}, \\ \widehat{q}_{1-\alpha,B^*} &= T^*_{B^*,\nu^*}, \text{ and } \widehat{q}_{\alpha,B^*} = T^*_{B^*,\eta^*}. \end{aligned} \tag{3.32}$$

The three-step method of determining B for the one-sided confidence interval $\widehat{J}_{1,B}$ is the same as that for $\widehat{J}_{2,B}$ except that one does not need to calculate η_0, $\widehat{q}_{\alpha,B_0}$,

$T^*_{B_0,\eta_0-\widehat{m}}$, or $T^*_{B_0,\eta_0+\widehat{m}}$ in Step 2 and one defines $B^* = \max\{B_0, B_{1\ell}\}$ in Step 3. (For the one-sided confidence interval $(-\infty, \widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{\alpha,B}]$ that has a finite upper bound, rather than lower bound, the three-step method is the same as that for $\widehat{J}_{2,B}$ except that one does not need to calculate ν_0, $\widehat{q}_{1-\alpha,B_0}$, $T^*_{B_0,\nu_0-\widehat{m}}$, or $T^*_{B_0,\nu_0+\widehat{m}}$ in Step 2 and one defines $B^* = \max\{B_0, B_{1u}\}$ in Step 3.) The desired cutoff value, ν^*, and the desired critical point, $\widehat{q}_{1-\alpha,B^*}$, for the one-sided confidence interval $J_{1,B}$ are given by (3.32) with ν_{1u} deleted in the definition of ν^*.

The equal-tailed and one-sided bootstrap confidence intervals based on B^* bootstrap repetitions, then, are equal to

$$\begin{aligned} \widehat{J}_{2,B^*} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B^*}, \widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{\alpha,B^*}] \text{ and} \\ \widehat{J}_{1,B^*} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B^*}, \infty). \end{aligned} \tag{3.33}$$

Table 7 provides the values of B_0 for equal-tailed and one-sided confidence intervals that correspond to different (α, pdb, τ) combinations, along with corresponding values of ν_0, η_0, c_α, and $\widehat{m}$. The pattern of Table 7 as (α, pdb, τ) varies is exactly the same as that of Table 5 for symmetric confidence intervals, as is expected from the formulae.

Tables 5 and 7 indicate that the B_0 values for equal-tailed and one-sided confidence intervals are noticeably larger than those for symmetric two-sided confidence intervals with the same confidence level. The ratio of the B_0 value for equal-tailed or one-sided confidence intervals to that for symmetric confidence intervals only depends on the confidence level and not on *pdb* or τ (except for rounding effects from the int$(\cdot)$ function.) For equal-tailed confidence intervals, this ratio is 2.0 and 2.1 for confidence levels .95 and .90 respectively. For one-sided confidence intervals, this ratio is 1.8 and 2.3 for confidence levels .95 and .90 respectively. (Note that to make these comparisons correctly one has to take account of the fact that the confidence levels of symmetric and one-sided confidence intervals are both $1-\alpha$, whereas the confidence level of equal-tailed confidence intervals is $1-2\alpha$.) The reason fewer repetitions are needed for symmetric confidence intervals is that the asymptotic density of $|T^*|$ is twice as large as that of T^* at any positive value.

As in Section 3.2, we computed the magnitude of B_1 values generated by the three-step method for equal-tailed and one-sided confidence intervals by specifying certain distributions for T^*. In this case, we considered the distributions $N(0,1)$, t_5, and χ^2_5. The results (in terms of the ratios of the median B_1 values to the B_0 values) and their implications are quite similar to those for symmetric confidence intervals, so we do not repeat them here.

The asymptotic justifications of the three-step methods introduced above are analogous to the asymptotic justification given for the symmetric percentile t confidence intervals. Details are given in Section 5.3 of the Appendix.

As in previous sections, when one is interested in separate confidence intervals for several parameters, say ω parameters, one can apply the three-step method for each of the parameters to obtain $B^*_{(1)}$, $B^*_{(2)}$, ..., $B^*_{(\omega)}$ and take B^* to equal the maximum of these values.

If the asymptotic distribution of T is not normal, then Steps 1 and 2 of the three-step method have to be adjusted. Suppose the asymptotic distribution of T is F, the α-th and the $(1-\alpha)$-th quantiles of F are $\overline{q}_\alpha$ and $\overline{q}_{1-\alpha}$, and F has a density $f(\cdot)$ with respect to Lebesgue measure at $\overline{q}_\alpha$ and $\overline{q}_{1-\alpha}$. To allow for the case where F is not symmetric about zero, we have to define two pairs of values $(B_{0\ell}, B_{0u})$ and $(a_{0\ell}, a_{0u})$ in place of B_0 and a_0 in Step 1 and four pairs of values $(\nu_{0\ell}, \nu_{0u})$, $(\eta_{0\ell}, \eta_{0u})$, $(\widehat{m}_\ell, \widehat{m}_u)$, and $(c_{\alpha\ell}, c_{\alpha u})$ in place of ν_0, η_0, $\widehat{m}$, and c_α, respectively, in Step 2. Then, in Step 3, $a_{0\ell}$ is defined with B_0, ν_0, η_0, and $\widehat{m}$ replaced by $B_{0\ell}$, $\nu_{0\ell}$, $\eta_{0\ell}$, and $\widehat{m}_\ell$, respectively, and a_{0u} is defined with B_0 ν_0, η_0, and $\widehat{m}$ replaced by B_{0u} ν_{0u}, η_{0u}, and $\widehat{m}_u$, respectively.

It suffices to define $(a_{0\ell}, a_{0u})$ and $(c_{\alpha\ell}, c_{\alpha u})$. The other new terms above follow from these via the same definitions as in Steps 1 and 2 above. We define $a_{0\ell} = \text{int}((10,000\alpha(1-\alpha)\chi^2_{1-\tau}) \,/(\overline{q}^2_{1-\alpha} f^2(\overline{q}_{1-\alpha}) pdb^2\alpha_2))$ and a_{0u} the same way except with $\overline{q}_\alpha$ in place of $\overline{q}_{1-\alpha}$ twice in the formula. We define $c_{\alpha\ell}$ as in (5.31) with $q_{1-\alpha}$ replaced by $\overline{q}_{1-\alpha}$ and $c_{\alpha u}$ as in (5.31) with $q_{1-\alpha}$ replaced by $\overline{q}_\alpha$.

If F depends on unknown parameters, then consistent estimates of these parameters can be used to provide estimates of $\overline{q}_{1-\alpha}$, $\overline{q}_\alpha$, $f(\overline{q}_{1-\alpha})$, and $f(\overline{q}_\alpha)$ for use in the definitions of $a_{0\ell}$, a_{0u}, $c_{\alpha\ell}$, and $c_{\alpha u}$.

4. Test Results

In this section, we consider the problem of choosing the number of bootstrap repetitions B for tests. First, we provide a method of doing so when a specific significance level α is of interest. Next, we provide a method of doing so for p-values. We recommend the use of the p-value results in most circumstances, because they convey more information.

4.1. Tests of Significance Level α

We begin by introducing some notation and definitions. As above, $\mathbf{X}$ denotes the observed data. We wish to construct a test of some null hypothesis of (approximate) significance level α for some $0 < \alpha < 1$. Our results apply to a wide variety of tests, such as tests of parametric restrictions and of model specification in parametric, semiparametric, and nonparametric models. Let T denote a test statistic based on $\mathbf{X}$. For example, T could be a t-statistic, a Wald statistic, a Lagrange multiplier statistic, a likelihood ratio statistic, etc.

We assume that the test statistic T has an asymptotic distribution, G, under the null hypothesis. For example, G could be a normal distribution, the distribution of the absolute value of a normal random variable, a chi-squared distribution with d degrees of freedom for some positive integer d, etc. We asssume that G has a unique $(1-\alpha)$-th quantile, denoted $q_{G,1-\alpha}$, and that G has a density with respect to Lebesgue measure at $q_{G,1-\alpha}$, denoted $g(q_{G,1-\alpha})$.

The "theoretical" test of significance level α rejects the null hypothesis if $T > k_\alpha$, where k_α is the solution to $P(T > k_\alpha) = \alpha$. By definition of k_α, the "theoretical" test has exact signifance level α. In practice, however, one typically does not know k_α. We consider using the bootstrap to approximate k_α.

Let $\mathbf{X}^* = (X_1^*, ..., X_n^*)'$ denote a bootstrap sample based on $\mathbf{X}$. Depending upon the circumstances, the bootstrap employed could be a variant of a nonparametric iid bootstrap, a moving block bootstrap for time series data, a parametric or semiparametric bootstrap for iid or time series data, or a bootstrap for regression models based on bootstrapping residuals . Let T^* denote a bootstrap version of the test statistic T based on $\mathbf{X}^*$. We assume that T^* is defined such that its asymptotic distribution conditional on the data is G with probability one (with respect to the randomness in the data). It is important for purposes of power of the test to define T^* such that the latter holds whether or not the null hypothesis is true; see Hall and Wilson (1991), Hall and Horowitz (1996), and Li and Maddala (1996).

If T is an asymptotically pivotal statistic (i.e., the asymptotic distribution G of T is the same for all distributions in the null), then the significance level of the bootstrap test typically exhibits higher order improvements (in terms of the closeness of its exact and nominal significance levels) over the standard test that results from using the delta method to estimate k_α; e.g., see Beran (1988) and Hall (1992).

Let $\widehat{k}_{\alpha,\infty}$ denote the *ideal bootstrap estimate* of k_α, i.e., the $(1-\alpha)$-th quantile of the distribution of T^*. To be precise, we define $\widehat{k}_{\alpha,\infty} = \inf\{k : P^*(T^* \leq k) \geq 1 - \alpha\}$, where $P^*(\cdot)$ denotes probability with respect to $\mathbf{X}^*$ conditional on the data $\mathbf{X}$. The ideal bootstrap test of approximate significance level α rejects the null hypothesis if $T > \widehat{k}_{\alpha,\infty}$. Analytic calculation of $\widehat{k}_{\alpha,\infty}$ is usually intractable. In consequence, one typically approximates $\widehat{k}_{\alpha,\infty}$ using bootstrap simulations.

As above, consider B iid bootstrap samples $\{\mathbf{X}_b^* : b = 1, ..., B\}$, each with the same distribution as $\mathbf{X}^*$. The corresponding B bootstrap test statistics are $\{T_b^* : b = 1, ..., B\}$. Let $\{T_{B,b}^* : b = 1, ..., B\}$ denote the ordered sample of $\{T_b^* : b = 1, ..., B\}$.

We assume that $\alpha = \alpha_1/\alpha_2$ for some positive integers α_1 and α_2 (with no common integer divisors). We choose B to be a positive integer that satisfies $\nu/(B+1) = 1-\alpha$ for some positive integer ν. That is, $B = \alpha_2 a - 1$ and $\nu = (\alpha_2 - \alpha_1)a$ for some positive integer a.

The bootstrap estimate of k_α based on B bootstrap repetitions is defined to be

$$\widehat{k}_{\alpha,B} = T_{B,\nu}^*. \tag{4.1}$$

That is, $\widehat{k}_{\alpha,B}$ is the ν-th order statistic of $\{T_b^* : b = 1, ..., B\}$. Then, the bootstrap test of approximate significance level α based on B bootstrap repetitions rejects the null hypothesis if

$$T > \widehat{k}_{\alpha,B}. \tag{4.2}$$

We now introduce the three-step method for determining B for the bootstrap test. The percentage deviation of the simulated critical value $\widehat{k}_{\alpha,B}$ from the ideal bootstrap critical value $\widehat{k}_{\alpha,\infty}$ is

$$100\frac{|\widehat{k}_{\alpha,B} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}}. \tag{4.3}$$

As above, let $1-\tau$ denote a probability close to one, such as .95. Let pdb be a bound on the percentage deviation of the critical value $\widehat{k}_{\alpha,B}$ from the critical value $\widehat{k}_{\alpha,\infty}$. We want to determine $B = B(pdb,\tau)$ such that

$$P^*\left(100\frac{|\widehat{k}_{\alpha,B}-\widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}} \leq pdb\right) = 1-\tau. \tag{4.4}$$

The three-step method of determining B is designed to do so. It is defined as follows:

Step 1. Compute a preliminary number of bootstrap repetitions B_0 via

$$\begin{aligned} B_0 &= \alpha_2 a_0 - 1, \text{ where} \\ a_0 &= \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{q^2_{G,1-\alpha}g^2(q_{G,1-\alpha})pdb^2\alpha_2}\right) \text{ and } \alpha = \alpha_1/\alpha_2. \end{aligned} \tag{4.5}$$

Step 2. Simulate B_0 bootstrap test statistics $\{T^*_b : b = 1,...,B_0\}$; compute the ordered bootstrap test statistics, which are denoted $\{T^*_{B_0,b} : b = 1,...,B_0\}$; and calculate $\nu_0 = (\alpha_2-\alpha_1)a_0$, $\widehat{m} = \text{int}(c_\alpha B_0^{2/3})$, $\widehat{k}_{\alpha,B_0} = T^*_{B_0,\nu_0}$, $T^*_{B_0,\nu_0-\widehat{m}}$, and $T^*_{B_0,\nu_0+\widehat{m}}$, where

$$c_\alpha = \left(\frac{1.5\chi^2_{1-\alpha}g^4(q_{G,1-\alpha})}{3g'(q_{G,1-\alpha})^2 - g(q_{G,1-\alpha})g''(q_{G,1-\alpha})}\right)^{1/3}. \tag{4.6}$$

Step 3. Take the desired number of bootstrap repetitions, B^*, to equal $B^* = \max\{B_0, B_1\}$, where

$$\begin{aligned} B_1 &= \alpha_2 a_1 - 1 \text{ and} \\ a_1 &= \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{\widehat{k}^2_{\alpha,B_0}pdb^2\alpha_2}\left(\frac{B_0}{2\widehat{m}}\right)^2\left(T^*_{B_0,\nu_0+\widehat{m}} - T^*_{B_0,\nu_0-\widehat{m}}\right)^2\right). \end{aligned} \tag{4.7}$$

Steps 1 and 2 require the calculation of $q_{G,1-\alpha}$, $g(q_{G,1-\alpha})$, $g'(q_{G,1-\alpha})$, and $g''(q_{G,1-\alpha})$, where $g'(\cdot)$ and $g''(\cdot)$ denote the first and second derivatives of $g(\cdot)$. If G is a standard normal distribution, then these quantities equal $z_{1-\alpha}$, $\phi(z_{1-\alpha})$, $-z_{1-\alpha}\phi(z_{1-\alpha})$, and $(z^2_{1-\alpha}-1)\phi(z_{1-\alpha})$, respectively, where $z_{1-\alpha}$ denotes the $(1-\alpha)$-th quantile of a standard normal distribution. If G is the distribution of the absolute value of a standard normal random variable and $\alpha < .5$, then these quantities equal $z_{1-\alpha/2}$, $2\phi(z_{1-\alpha/2})$, $-2z_{1-\alpha/2}\phi(z_{1-\alpha/2})$, and $2(z^2_{1-\alpha/2}-1)\phi(z_{1-\alpha/2})$ respectively. If G is a chi-squared distribution with d degrees of freedom, then these quantities equal q, $g(q) = (2^{d/2}\Gamma(d/2))^{-1}q^\psi\exp(-q/2)$, $g'(q) = (\psi/q - 1/2)g(q)$, and $g''(q) = (\psi(\psi-1)/q^2 - \psi/q + 1/4)g(q)$, respectively, where $\psi = d/2-1$ and q denotes the $(1-\alpha)$-th quantile of a chi-squared distribution with d degrees of freedom.

If G depends on unknown parameters, then consistent estimates of these parameters can be used to provide estimates of $q_{G,1-\alpha}$ and $f(q_{G,1-\alpha})$ for use in Steps 1 and 2.

Having determined B^*, one simulates $B^* - B_0$ (≥ 0) additional bootstrap test statistics $\{T_b^* : b = B_0 + 1, ..., B^*\}$ and orders the values of the B^* bootstrap test statistics, which are denoted $\{T_{B^*,b}^* : b = 1, ..., B^*\}$. The desired cutoff value, ν^*, and the desired critical value, $\widehat{k}_{\alpha,B^*}$, are given by

$$(4.8) \qquad \nu^* = \max\{\nu_0, \nu_1\}, \ \nu_1 = (\alpha_2 - \alpha_1)a_1, \ \text{and } \widehat{k}_{\alpha,B^*} = T_{B^*,\nu^*}^*.$$

The bootstrap test based on B^* bootstrap repetitions, then, rejects the null hypothesis if

$$(4.9) \qquad T > \widehat{k}_{\alpha,B^*}.$$

To assess the computational burden of the three-step procedure for tests with specified significance levels, Table 8 provides values for B_0, as well as a_0, ν_0, and $\widehat{m}$, for a variety of (α, τ) combinations when *pdb* $= 10$ and the asymptotic null distribution of the test statistic is absolute $N(0,1)$, $N(0,1)$, χ_5^2, and χ_{15}^2. (Results for the χ_{10}^2 distribution are intermediate between those of the χ_5^2 and χ_{15}^2 distributions, but are somewhat closer to the χ_{15}^2 results.)

Table 8 shows that for tests with absolute $N(0,1)$ asymptotic null distribution the same number of initial bootstrap repetitions B_0 are needed as for symmetric confidence intervals. For tests with $N(0,1)$ asymptotic null distribution, noticeably larger B_0 values are required—the ratio of B_0 values for $N(0,1)$ to absolute $N(0,1)$ tests is in the range 1.8–2.3 for $\alpha = .05$ or $\alpha = .10$. For tests with χ_5^2 asymptotic null distribution, similar B_0 values are required as for absolute $N(0,1)$ tests—the ratio of B_0 values for χ_5^2 to absolute $N(0,1)$ tests is in the range 1.0–1.2 for $\alpha = .05$ or $\alpha = .10$. For tests with χ_{15}^2 asymptotic null distribution, noticeably smaller B_0 values are required than for absolute $N(0,1)$ tests—the ratio of B_0 values for χ_{15}^2 to absolute $N(0,1)$ tests is in the range .38–.50 for $\alpha = .05$ or $\alpha = .10$. Thus, there is considerable variation in suitable values of B_0 for test statistics with different asymptotic null distributions.

In all cases, B_0 increases quickly as α or τ decreases. It is also true that B_0 increases very quickly as *pdb* decreases, but Table 8 only reports results for *pdb* $= 10$. For most combinations reported, the number of bootstrap repetitions required is greater than that commonly used in empirical econometric applications.

The asymptotic justification of the three-step method introduced above is analogous to the asymptotic justification given for symmetric percentile t confidence intervals in Section 3.3. See Section 5.4 of the Appendix of Proofs for details.

4.2. p-values

We now consider choosing the number of bootstrap repetitions B for a testing problem in which one wants to report a p-value. By definition, the p-value is the infimum of the significance levels for which the test rejects the null hypothesis given the observed value of the test statistic T. We view the reporting of a p-value to be an efficient method of communicating the result of hypothesis tests for all significance levels $\alpha \in (0,1)$. The use of a bootstrap p-value exploits the higher-order improvements of the bootstrap, because given the p-value and a significance level α of interest (which may vary across individuals), one can determine whether the test rejects the null hypothesis and the significance level of the test is accurate to the level obtained by the bootstrap test.

We use the same notation, definitions, and assumptions as in Section 4.1 except that assumptions (5.36) and (5.37) are not needed and n is treated as fixed. We consider probabilities with respect to the bootstrap simulation randomness conditional on the sample $\mathbf{X}$.

The ideal bootstrap p-value is defined to be

$$\widehat{p}_\infty = P^*(T^* > T), \tag{4.10}$$

where $P^*(\cdot)$ denotes probability with respect to $\mathbf{X}^*$ (and hence T^*) conditional on $\mathbf{X}$ (and hence T). We assume that $\widehat{p}_\infty$ does not equal zero or one. (This holds with probability one with respect to the randomness in the sample $\mathbf{X}$ except in pathological cases.) We estimate $\widehat{p}_\infty$ using bootstrap simulations. Given B bootstrap repetitions, the bootstrap p-value is

$$\widehat{p}_B = \frac{1}{B}\sum_{b=1}^{B} 1(T_b^* > T). \tag{4.11}$$

We now introduce a three-step method for determining B for the bootstrap p-value. The percentage deviation of the simulated p-value $\widehat{p}_B$ from the ideal bootstrap p-value $\widehat{p}_\infty$is

$$100\frac{|\widehat{p}_B - \widehat{p}_\infty|}{\widehat{p}_\infty}. \tag{4.12}$$

Let $1-\tau$ denote a probability close to one, such as .95. Let pdb be a bound on the percentage deviation of $\widehat{p}_B$ from $\widehat{p}_\infty$. We want to determine $B = B(pdb, \tau)$ such that

$$P^*\left(100\frac{|\widehat{p}_B - \widehat{p}_\infty|}{\widehat{p}_\infty} \leq pdb\right) = 1-\tau. \tag{4.13}$$

The three-step method of determining B is designed to do this. It is defined as follows:

Step 1. Compute a preliminary number of bootstrap repetitions B_0 via

$$B_0 = \text{int}\left(\frac{10,000\chi^2_{1-\tau}G(T)}{(1-G(T))pdb^2}\right). \tag{4.14}$$

Step 2. Simulate B_0 bootstrap test statistics $\{T_b^* : b = 1, ..., B_0\}$ and compute

$$\widehat{p}_{B_0} = \frac{1}{B_0}\sum_{b=1}^{B_0} 1(T_b^* > T). \tag{4.15}$$

Step 3. Take the desired number of bootstrap repetitions, B^*, to equal $B^* = \max\{B_0, B_1\}$, where

$$B_1 = \text{int}\left(\frac{10,000\chi^2_{1-\tau}(1-\widehat{p}_{B_0})}{\widehat{p}_{B_0}pdb^2}\right). \tag{4.16}$$

In Step 1, the term $1-G(T)$ is an initial estimate of the p-value that is obtained by using the asymptotic null distribution G of the test statistic T. For example, if G is a standard normal distribution, then $G(T) = \Phi(T)$, where $\Phi(\cdot)$ is the standard normal distribution function. If G is the distribution of the absolute value of a standard normal random variable, then $G(T) = 2\Phi(T) - 1$. If G is a chi-squared distribution, then $G(T)$ is the corresponding chi-squared distribution function evaluated at T.

If G depends on unknown parameters, then consistent estimates of these parameters can be used to provide estimates of $G(T)$ for use in Step 1.

We note that one should adjust one's choice of *pdb* in light of the initial p-value estimates $1-G(T)$ and/or $\widehat{p}_{B_0}$, because the level of accuracy needed depends on the p-value. In particular, when the p-value is small or large, e.g., .001 or .70, then one does not need to take *pdb* to be as small as when the p-value is of intermediate magnitude, e.g., .075. This is quite important in terms of minimizing the computational burden.

If one or more significance levels α are of particular interest, then one should choose B_0 and B_1 (by rounding up) such that $\nu/(B+1) = 1-\alpha$ for some positive integer ν for each α of interest. For example, if significance levels .05 and .10 are of particular interest, then B_0 and B_1 should be rounded up to the nearest value $B = 20a - 1$ for some $a = 1, 2, \ldots.$ (i.e., $B = 19, 39, 59$, etc.). If .01, .05, and .10 are of particular interest, then B_0 and B_1 should be rounded up to the nearest value $B = 100a - 1$ for some $a = 1, 2, \ldots.$ The reason for doing this is that it has advantages in terms of the unconditional significance level for the values of α of particular interest; see Hall (1992, p. 307).

Having determined B^*, one simulates $B^* - B_0$ (≥ 0) additional bootstrap test statistics $\{T_b^* : b = B_0 + 1, \ldots, B^*\}$ and computes the bootstrap p-value

$$\widehat{p}_{B^*} = \frac{1}{B^*} \sum_{b=1}^{B^*} 1(T_b^* > T). \tag{4.17}$$

Table 9 provides representative values of B_0 for the three-step method for p-values. Three different values of τ are considered, viz., .01, .05 and .10. A range of values of the initial p-value estimate $1 - G(T)$ and the accuracy bound *pdb* are considered. For clarity, the Table only provides $(1 - G(T), pdb)$ combinations that are of some interest. For example, it is not of interest to consider the combination (.001,5), because this combination yields excessive accuracy and, hence, requires an excessively large value of B_0.

Table 9 indicates that the required magnitude of B_0 depends on the initial p-value estimate $1-G(T)$. If it is quite small or large, then one does not need a small value of *pdb* and the required magnitude of B_0 is not large. On the other hand, if $1-G(T)$ is in an intermediate range, such as (.01, .15), then one may want to employ a relatively small value of *pdb* and the required magnitude of B_0 may be quite large.

The asymptotic justification of the three-step method of choosing B^* is in terms of the limit as $pdb \to 0$ with n fixed:

$$P^*\left(100\frac{|\widehat{p}_{B_1} - \widehat{p}_\infty|}{\widehat{p}_\infty} \leq pdb\right) \to 1 - \tau \quad \text{as } pdb \to 0, \tag{4.18}$$

where $\widehat{p}_{B_1} = \frac{1}{B_1}\sum_{b=1}^{B_1}(T_b^* > T)$. This conditional result holds provided $\widehat{p}_\infty$ does not equal zero or one. The proof is given in the Appendix of Proofs.

Equation (4.18) implies that the three-step method attains precisely the desired level of accuracy using "small *pdb*" asymptotics when $B^* = B_1 \geq B_0$. When $B^* = B_0 > B_1$, then the accuracy of the three-step method exceeds the desired level of accuracy.

5. Appendix of Proofs

5.1. Proofs of the Standard Error Results

First, we prove (2.11). We rewrite $\widehat{se}_B$ of (2.3) as

$$\begin{aligned}\widehat{se}_B &= \left(\frac{1}{B}\Sigma_{b=1}^B\left(\widehat{\theta}_b^* - \mu\right)^2 - \left(\frac{1}{B}\Sigma_{b=1}^B\widehat{\theta}_b^* - \mu\right)^2\right)^{1/2}\\ &= m(A_B), \text{ where}\end{aligned}$$

$$A_B = \begin{pmatrix}\frac{1}{B}\Sigma_{b=1}^B\left(\widehat{\theta}_b^* - \mu\right)^2\\ \frac{1}{B}\Sigma_{b=1}^B\widehat{\theta}_b^* - \mu\end{pmatrix} \text{ and } m(a) = \left(a_1 - a_2^2\right)^{1/2} \text{ for } a = (a_1,\, a_2)'. \tag{5.1}$$

For convenience, we have replaced $B-1$ in the denominator of $\widehat{se}_B$ by B. By the central limit theorem,

$$B^{1/2}(A_B - A) \xrightarrow{d} N(\mathbf{0},\, \Omega) \text{ as } B \to \infty, \text{ where}$$

$$A = \begin{pmatrix}\widehat{se}_\infty^2\\ 0\end{pmatrix} \text{ and } \Omega = \begin{pmatrix}E^*\left((\widehat{\theta}_b^* - \mu)^2 - \widehat{se}_\infty^2\right)^2 & E^*(\widehat{\theta}_b^* - \mu)^3\\ E^*(\widehat{\theta}_b^* - \mu)^3 & \widehat{se}_\infty^2\end{pmatrix}. \tag{5.2}$$

We have $\frac{\partial}{\partial a}m(a) = \frac{1}{2}\left(a_1 - a_2^2\right)^{-1/2}(1,\, -2a_2)'$ and $\frac{\partial}{\partial a}m(A) = (1/(2\widehat{se}_\infty),\, 0)'$. The delta method now gives

$$B^{1/2}(\widehat{se}_B - \widehat{se}_\infty) = B^{1/2}(m(A_B) - m(A)) \xrightarrow{d} N(0, V), \text{ where}$$

$$V = \frac{1}{4\widehat{se}_\infty^2}E^*\left((\widehat{\theta}_b^* - \mu)^2 - \widehat{se}_\infty^2\right)^2 = \frac{\widehat{se}_\infty^2}{4}(2 + \gamma_2). \tag{5.3}$$

In turn, this gives

$$\left(100\frac{|\widehat{se}_B - \widehat{se}_\infty|}{\widehat{se}_\infty}\right)^2 / (2{,}500(2+\gamma_2)/B) \xrightarrow{d} \chi^2, \tag{5.4}$$

where χ^2 denotes a chi-squared random variable with one degree of freedom, which establishes (2.11).

Next, we prove (2.17). Let $B_{11} = 2{,}500\chi_{1-\tau}^2(2+\gamma_2)/pdb^2$. Note that B_{11} is nonrandom. Equations (5.2)–(5.4) hold with B replaced by B_{11} throughout and with the limit as $B \to \infty$ replaced by the limit as $pdb \to 0$ (because the latter forces $B_{11} \to \infty$). Now, by the central limit theorem of Doeblin–Anscombe for a sum of independent random variables with a random number of terms in the sum (e.g., see Chow and Teicher (1978, Thm. 9.4.1, p. 317)), provided $B_1/B_{11} \to_p 1$ as $pdb \to 0$, the result of (5.2) holds with B replaced by B_1 and with the limit as $B \to \infty$ replaced by the limit as $pdb \to 0$. In turn, this implies that (5.3) and (5.4) hold with the same changes. The latter can be rewritten using (2.16) as

$$\left(100\frac{|\widehat{se}_{B_1} - \widehat{se}_\infty|}{\widehat{se}_\infty}\right)^2(\chi_{1-\tau}^2/pdb^2)(2+\widehat{\gamma}_{2B_0})/(2+\gamma_2) \xrightarrow{d} \chi^2 \text{ as } pdb \to 0. \tag{5.5}$$

(The effect of the int(·) function in (2.16) is asymptotically negligable and, hence, is ignored in obtaining the previous equation from (5.4).) By (2.7) and the fact that $B_0 \to \infty$ as $pdb \to 0$, this yields

$$(5.6) \qquad \left(100\frac{|\widehat{se}_{B_1} - \widehat{se}_\infty|}{\widehat{se}_\infty}\right)^2 \chi^2_{1-\tau}/pdb^2 \xrightarrow{d} \chi^2 \text{ as } pdb \to 0,$$

which establishes (2.17).

It remains to show that $B_1/B_{11} \to_p 1$ as $pdb \to 0$. This follows from (2.7) because $B_1/B_{11} = (2 + \widehat{\gamma}_{2B_0})/(2 + \gamma_2)$.

5.2. Proofs of the Symmetric Confidence Interval Results

We now prove (3.20). First we show that (3.20) holds with B_1 replaced by the non-random quantity B_{11}. By definition,

$$\begin{aligned} & B_{11} = \alpha_2 a_{11} - 1, \ \nu_{11} = (\alpha_2 - \alpha_1)a_{11}, \text{ and} \\ (5.7) \qquad & a_{11} = \text{int}\left(\frac{10,000\alpha(1-\alpha)\chi^2_{1-\tau}}{z^2_{1-\alpha/2}pdb^2\alpha_2}\left(\frac{1}{2\phi(z_{1-\alpha/2})}\right)^2\right). \end{aligned}$$

Note that $B_{11} \to \infty$ as $pdb \to 0$ and B_{11} does not depend on n.

We establish the asymptotic distribution of $B_{11}^{1/2}(\widehat{k}_{\alpha,B_{11}} - \widehat{k}_{\alpha,\infty})$ as $pdb \to 0$ and $n \to \infty$, where $\widehat{k}_{\alpha,B_{11}} = |T^*|_{B_{11},\nu_{11}}$, using an argument developed for proving the asymptotic distribution of the sample median based on an iid sample of random variables that are absolutely continuous at their population median; see Lehmann (1983, Thm. 5.3.2, p. 354). (In contrast, recall that $\widehat{k}_{\alpha,B_{11}}$ is the sample $(1-\alpha)$-*th* quantile of iid observations each with the bootstrap distribution of $|T^*|$, which depends on n and may be discrete.)

We have: For any $x \in R$,

$$(5.8) \qquad P^*(B_{11}^{1/2}(\widehat{k}_{\alpha,B_{11}} - \widehat{k}_{\alpha,\infty}) \le x) = P^*(|T^*|_{B_{11},\nu_{11}} \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}).$$

Let S_B be the number of $|T_b^*|$'s for $b = 1, ..., B$ that exceed $\widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}$. Presently, we consider $S_{B_{11}}$. Below, we consider S_{B_1}. (In both cases, the cutoff point $\widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}$ depends on B_{11}.) We have

$$(5.9) \quad |T^*|_{B_{11},\nu_{11}} \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2} \text{ if and only if } S_{B_{11}} \le B_{11} - \nu_{11} = B_{11}\alpha - (1-\alpha).$$

The random variable $S_{B_{11}}$ has a binomial distribution with parameters $(B_{11}, p_{B_{11},n})$, where

$$(5.10) \qquad p_{B_{11},n} = 1 - P^*(|T^*| \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}).$$

The probability in (5.8) equals

$$\begin{aligned} & P^*\left(S_{B_{11}} \le B_{11}\alpha - (1-\alpha)\right) \\ (5.11) \qquad & = P^*\left(\frac{S_{B_{11}} - B_{11}p_{B_{11},n}}{(B_{11}p_{B_{11},n}(1-p_{B_{11},n}))^{1/2}} \le \frac{B_{11}\alpha - (1-\alpha) - B_{11}p_{B_{11},n}}{(B_{11}p_{B_{11},n}(1-p_{B_{11},n}))^{1/2}}\right). \end{aligned}$$

Note that the random variable in the right-hand side probability has mean zero and variance one and satisfies the conditions of the Lindeberg central limit theorem (applied with $pdb \to 0$ and $n \to \infty$).

Using the assumption of (3.18), we obtain

$$
\begin{aligned}
\widehat{k}_{\alpha,\infty} &= \inf\{k : P^*(|T^*| \le k) \ge 1-\alpha\} = \inf\{k : P(|Z| \le k) \ge 1-\alpha\} + o(1) \\
&= z_{1-\alpha/2} + o(1) \text{ as } n \to \infty \text{ and} \\
p_{B_{11},n} &= 1 - P^*(|T^*| \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}) \to \alpha \text{ as } pdb \to 0 \text{ and } n \to \infty. \qquad (5.12)
\end{aligned}
$$

The upper bound in the right-hand side probability of (5.11) can be written as

$$
\begin{aligned}
w_{B_{11},n} &= \frac{B_{11}^{1/2}(\alpha - p_{B_{11},n}) - (1-\alpha)/B_{11}^{1/2}}{(p_{B_{11},n}(1-p_{B_{11},n}))^{1/2}} \\
&= \left((\alpha(1-\alpha))^{-1/2} + o(1)\right) B_{11}^{1/2}(\alpha - p_{B_{11},n}) + o(1) \qquad (5.13)
\end{aligned}
$$

as $pdb \to 0$ and $n \to \infty$. In addition, we have

$$
\begin{aligned}
B_{11}^{1/2}(\alpha - p_{B_{11},n}) &= B_{11}^{1/2}(P^*(|T^*| \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}) - (1-\alpha)) \\
&= B_{11}^{1/2}(P(|T| \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}) - P(|T| \le \widehat{k}_{\alpha,\infty})) + o(1) \\
&= B_{11}^{1/2}(P(|Z| \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}) - P(|Z| \le \widehat{k}_{\alpha,\infty})) + o(1) \\
&= B_{11}^{1/2} 2\phi(\zeta_{B_{11},n}) x/B_{11}^{1/2} + o(1) \\
&\to 2\phi(z_{1-\alpha/2})x \text{ as } pdb \to 0 \text{ and } n \to \infty. \qquad (5.14)
\end{aligned}
$$

The first equality of 5.14 holds by the definition of $p_{B_{11},n}$. The second and third equalities hold by (3.18) and (3.19) (using the fact that the latter and the definition of B_{11} imply that $B_{11}^{1/2} = O(1/pdb) = n^{\xi}O(1/(pdb \times n^{\xi})) = o(n^{\xi})$). The fourth equality holds for some $\zeta_{B_{11},n}$ that lies between $\widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}$ and $\widehat{k}_{\alpha,\infty}$ by a mean value expansion using the fact that the density at $y > 0$ of the absolute value of a standard normal random variable is $2\phi(y)$. The convergence result of 5.14 holds because $\zeta_{B_{11},n} \to z_{1-\alpha/2}$ as $pdb \to 0$ and $n \to \infty$.

Equations (5.13) and (5.14) give

$$
\lim_{pdb \to 0, n \to \infty} w_{B_{11},n} = 2\phi(z_{1-\alpha/2})/(\alpha(1-\alpha))^{1/2}. \qquad (5.15)
$$

Equations (5.8), (5.11), and (5.15) plus the Lindeberg central limit theorem applied to (5.11) yield

$$
\begin{aligned}
&P^*(B_{11}^{1/2}(\widehat{k}_{\alpha,B_{11}} - \widehat{k}_{\alpha,\infty}) \le x) \to \Phi(x 2\phi(z_{1-\alpha/2})/(\alpha(1-\alpha))^{1/2}) \text{ and} \\
&B_{11}^{1/2}(\widehat{k}_{\alpha,B_{11}} - \widehat{k}_{\alpha,\infty}) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)}{(2\phi(z_{1-\alpha/2}))^2}\right) \qquad (5.16)
\end{aligned}
$$

as $pdb \to 0$ and $n \to \infty$.

This result, (3.11), and (5.7) imply that

$$
\begin{aligned}
& P^*\left(100\frac{|\widehat{k}_{\alpha,B_{11}} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}} \leq pdb\right) \\
&= P^*\left(100\frac{|\widehat{k}_{\alpha,B_{11}} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}} \leq \frac{100(\alpha(1-\alpha))^{1/2}\chi_{1-\tau}}{z_{1-\alpha/2}B_{11}^{1/2}}\left(\frac{1}{2\phi(z_{1-\alpha/2})}\right)(1+o(1))\right) \\
&\to 1-\tau \ \text{ as } pdb \to 0 \text{ and } n \to \infty, \text{ where}
\end{aligned}
$$

$$
\widehat{J}_{SY,B_{11}} = [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B_{11}}, \widehat{\theta} + n^{-\kappa}\widehat{\sigma}\widehat{k}_{\alpha,B_{11}}]. \tag{5.17}
$$

Thus, (3.20) holds with B_1 replaced with B_{11}.

Next, we show that

$$
(B_1 - B_{11})/B_{11} \xrightarrow{p} 0 \text{ as } pdb \to 0 \text{ and } n \to \infty \tag{5.18}
$$

(with respect to the simulation randomness conditional on the data). This follows from

$$
\widehat{k}_{\alpha,B_0} \xrightarrow{p} z_{1-\alpha/2} \text{ and } \left(\frac{B_0}{2\widehat{m}}\right)\left(|T^*|_{B_0,\nu_0+\widehat{m}} - |T^*|_{B_0,\nu_0-\widehat{m}}\right) \xrightarrow{p} \frac{1}{2\phi(z_{1-\alpha/2})} \tag{5.19}
$$

as $pdb \to 0$ and $n \to \infty$. The former holds by (5.12) and (5.16) with B_{11} replaced by B_0. The latter is established as follows.

Define the inverse of a distribution function F to be $F^{-1}(t) = \inf\{x : F(x) \geq t\}$. Let $\widehat{F}_{|T^*|}(\cdot)$ denote the distribution function of the bootstrap distribution of $|T^*|$. Let $\{U_b : b = 1, ..., B_0\}$ denote iid uniform [0,1] random variables. Let $\{U_{B_0,b} : b = 1, ..., B_0\}$ denote the ordered sample of $\{U_b : b = 1, ..., B_0\}$. Then, $\widehat{F}_{|T^*|}^{-1}(U_b)$ has the same distribution as $|T_b^*|$ and $\widehat{F}_{|T^*|}^{-1}(U_{B_0,b})$ has the same distribution as $|T^*|_{B_0,b}$. It suffices to show that

$$
\left(\frac{B_0}{2\widehat{m}}\right)\left(\widehat{F}_{|T^*|}^{-1}(U_{B_0,\nu_0+\widehat{m}}) - \widehat{F}_{|T^*|}^{-1}(U_{B_0,\nu_0-\widehat{m}})\right) \xrightarrow{p} \frac{1}{2\phi(z_{1-\alpha/2})} \tag{5.20}
$$

as $pdb \to 0$ and $n \to \infty$. (Note that $B_0 \to \infty$ as $pdb \to 0$.) Let

$$
\Psi(x) = P(|Z| \leq x) = \Phi(x) - \Phi(-x), \text{ where } Z \sim N(0,1). \tag{5.21}
$$

The left-hand side of (5.20) equals

$$
\left(\frac{\widehat{F}_{|T^*|}^{-1}(U_+) - \widehat{F}_{|T^*|}^{-1}(U_-)}{U_+ - U_-}\right)\left(\frac{B_0}{2\widehat{m}}\right)(U_+ - U_-) = \left(\frac{\widehat{F}_{|T^*|}^{-1}(U_+) - \widehat{F}_{|T^*|}^{-1}(U_-)}{U_+ - U_-}\right)(1 + o_p(1)) \tag{5.22}
$$

where U_+ and U_- abbreviate $U_{B_0,\nu_0+\widehat{m}}$ and $U_{B_0,\nu_0-\widehat{m}}$ respectively. Equation (5.22) holds by the argument of Bloch and Gastwirth (1968, Pf. of Thm. 1) (which relies on the fact that the spacings of the order statistics of uniform random variables have beta distributions). The first term in parentheses on the right-hand side of (5.22) equals

$$\frac{\Psi^{-1}(U_+)-\Psi^{-1}(U_-)}{U_+-U_-}+\frac{B_0^{1/3}(\widehat{F}_{|T^*|}^{-1}(U_+)-\Psi^{-1}(U_+))}{B_0^{1/3}(U_+-U_-)}-\frac{B_0^{1/3}(\widehat{F}_{|T^*|}^{-1}(U_-)-\Psi^{-1}(U_-))}{B_0^{1/3}(U_+-U_-)}. \tag{5.23}$$

The first summand of (5.23) satisfies

$$U_+ \xrightarrow{p} 1-\alpha,\ U_- \xrightarrow{p} 1-\alpha, \text{ and}$$

$$\frac{\Psi^{-1}(U_+)-\Psi^{-1}(U_-)}{U_+-U_-} \xrightarrow{p} \frac{\partial}{\partial x}\Psi^{-1}(1-\alpha)=\frac{1}{\psi(\Psi^{-1}(1-\alpha))}=\frac{1}{2\phi(z_{1-\alpha/2})} \tag{5.24}$$

as *pdb* $\to 0$ and $n\to\infty$, where $\psi(\cdot)=\Psi'(\cdot)$. The first two results of (5.24) hold by standard results for the sample quantiles of iid uniform random variables. The third result follows from the first two results using the definition of differentiability of $\Psi^{-1}(\cdot)$ and an almost sure representation argument.

Next, we show that the second and third summands of (5.23) are $o_p(1)$. By the argument of Bloch and Gastwirth referred to above, $B_0^{1/3}(U_+-U_-) \xrightarrow{p} 2c_\alpha > 0$. Thus, it suffices to show that

$$B_0^{1/3}(\widehat{F}_{|T^*|}^{-1}(U_+)-\Psi^{-1}(U_+)) \xrightarrow{p} 0 \text{ as } pdb\to 0 \text{ and } n\to\infty \tag{5.25}$$

and likewise with "U_+" replaced by "U_-". The proofs of these two results are the same, so we just prove the former.

It suffices to prove (5.25) with $B_0^{1/3}$ replaced by $n^{2\xi/3}$ because $B_0^{1/3}=o(n^{2\xi/3})$ by the assumption of (3.19). For any distribution function F, $x_1 < F^{-1}(t)\le x_2$ if and only if $F(x_1)<t\le F(x_2)$; see Shorack and Wellner (1986, p. 5). Thus, for any $\varepsilon>0$,

$$\begin{aligned}
&n^{2\xi/3}|\widehat{F}_{|T^*|}^{-1}(U_+)-\Psi^{-1}(U_+)|\le\varepsilon\\
&\quad\text{iff } \Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon < F_{|T^*|}^{-1}(U_+)\le \Psi^{-1}(U_+)+n^{-2\xi/3}\varepsilon\\
&\quad\text{iff } F_{|T^*|}(\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon) < U_+ \le F_{|T^*|}(\Psi^{-1}(U_+)+n^{-2\xi/3}\varepsilon).
\end{aligned} \tag{5.26}$$

We have

$$\begin{aligned}
&F_{|T^*|}(\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon)\\
&=(F_{|T^*|}(\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon)-\Psi(\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon))+\Psi(\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon)\\
&=O_p(n^{-\xi})+\left(U_++2\phi(\gamma_{B_0,n})(-n^{-2\xi/3}\varepsilon)\right)\\
&< U_+ \text{ with probability that goes to one as } pdb\to 0 \text{ and } n\to\infty,
\end{aligned} \tag{5.27}$$

where the second equality holds by (i) the assumption of (3.18), the fact that $\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon \to_p z_{1-\alpha/2}$, and the use of an almost sure representation argument and (ii) a mean value expansion, where $\gamma_{B_0,n}$ lies between $\Psi^{-1}(U_+)-n^{-2\xi/3}\varepsilon$

and $\Psi^{-1}(U_+)$ and, hence, $\gamma_{B_0,n} \to_p z_{1-\alpha/2}$. An analogous result (with the inequality reversed) holds for $F_{|T^*|}(\Psi^{-1}(U_+) + n^{-2\xi/3}\varepsilon)$. Hence, the right-hand side of (5.26) holds with probability that goes to one, which establishes (5.25), and the proof of the second result of (5.19) is complete.

Now we use equation (5.18) and the above proof that (3.20) holds with the random quantity B_1 replaced by the non-random quantity B_{11} to establish (3.20) as is.

First, we have: For any $x \in R$,

$$P^*(B_{11}^{1/2}(\widehat{k}_{\alpha,B_1} - \widehat{k}_{\alpha,\infty}) \le x) = P^*(|T^*|_{B_1,\nu_{11}} \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2}). \tag{5.28}$$

(Note that we take the normalization factor to be B_{11} not B_1.) Let S_{B_1} be as defined above. We have

$$|T^*|_{B_1,\nu_{11}} \le \widehat{k}_{\alpha,\infty} + x/B_{11}^{1/2} \text{ iff } S_{B_1} \le B_1 - \nu_1 = B_1\alpha - (1-\alpha). \tag{5.29}$$

The random variable S_{B_1} has a binomial distribution with parameters $(B_1, p_{B_{11},n})$, where $p_{B_{11},n}$ is the same as above. The probability in (5.28) equals

$$\begin{aligned} & P^*\left(S_{B_1} \le B_1\alpha - (1-\alpha)\right) \\ & = P^*\left(\frac{S_{B_1} - B_1 p_{B_{11},n}}{(B_1 p_{B_{11},n}(1-p_{B_{11},n}))^{1/2}} \le \frac{B_1\alpha - (1-\alpha) - B_1 p_{B_{11},n}}{(B_1 p_{B_{11},n}(1-p_{B_{11},n}))^{1/2}}\right). \end{aligned} \tag{5.30}$$

The random variable depending on S_{B_1} in the right-hand side probability is a normalized sum of independent random variables with a random number, B_1, of terms in the sum. By the central limit theorem of Doeblin-Anscombe (e.g., see Chow and Teicher (1978, Thm. 9.4.1, p. 317)), it has a standard normal asymptotic distribution, because (i) it has a standard normal asymptotic distribution when B_1 is replaced by the non-random quantity B_{11} and (ii) $B_1/B_{11} \to_p 1$ by (5.18).

Now, for present purposes, equation (5.12) holds without any changes. In addition, by the same argument as in (5.13)–(5.15), coupled with (5.18), the upper bound in the right-hand side of (5.30) converges in probability to $z_{1-\alpha/2}$ as $pdb \to 0$ and $n \to \infty$. These results and the result of the previous paragraph combine to verify (5.16) and (5.17) with B_{11} replaced by B_1. That is, (3.20) holds, as desired.

We finish by showing that our formula for the bandwidth parameter $\widehat{m}$ used with the Siddiqui estimator corresponds to that given by Hall and Sheather (1988). In our notation, Hall and Sheather's formula is

$$\widehat{m} = \text{int}(c_\alpha B_0^{2/3}) \text{ and } c_\alpha = \left(\frac{1.5 z_{1-\alpha/2}^2 f^4(q_{1-\alpha})}{3f'(q_{1-\alpha})^2 - f(q_{1-\alpha})f''(q_{1-\alpha})}\right)^{1/3}, \tag{5.31}$$

where $f(\cdot)$ denotes the density of the iid random variables upon which the sample quantile is based, $f'(\cdot)$ and $f''(\cdot)$ denote the first two derivatives of $f(\cdot)$, $q_{1-\alpha}$ denotes the population quantile, and $z_{1-\alpha/2}$ is as above. In our case, we use the asymptotic analogues of $f(\cdot)$ and $q_{1-\alpha}$, viz., $2\phi(\cdot)$ and $z_{1-\alpha/2}$, in the formula. Note that $\phi'(x) = -x\phi(x)$ and $\phi''(x) = (x^2-1)\phi(x)$. Hence, our constant c_α satisfies

$$c_\alpha = \left(\frac{6 z_{1-\alpha/2}^2 \phi^2(z_{1-\alpha/2})}{2z_{1-\alpha/2}^2 + 1}\right)^{1/3} = \left(\frac{6 z_{1-\alpha/2}^2 \phi^4(z_{1-\alpha/2})}{3\phi'(z_{1-\alpha/2})^2 - \phi(z_{1-\alpha/2})\phi''(z_{1-\alpha/2})}\right)^{1/3}, \tag{5.32}$$

which corresponds to Hall and Sheather's constant c_α.

5.3. Asymptotic Justification for the Three-step Methods for Equal-tailed and One-sided Confidence Intervals

The asymptotic justification of the three-step methods introduced in Section 3.5 for equal-tailed and one-sided confidence intervals is analogous to that given for the symmetric percentile t confidence intervals in Section 3.3. First, we introduce an analogous strengthening of the assumption of asymptotic normality of the t statistic T. For $\widehat{J}_{2,B^*}$, we assume: For some $\xi > 0$ and all sequences of constants $\{x_n : n \geq 1\}$ for which $x_n \to z_{1-\alpha/2}$ or $x_n \to -z_{1-\alpha/2}$, we have

$$\begin{aligned} P(T \leq x_n) &= P(Z \leq x_n) + O(n^{-\xi}) \text{ as } n \to \infty \text{ and} \\ P^*(T^* \leq x_n) &= P(Z \leq x_n) + O(n^{-\xi}) \text{ as } n \to \infty, \end{aligned} \tag{5.33}$$

where $Z \sim N(0,1)$. For $\widehat{J}_{1,B^*}$, we make the same assumption except that we do not need it to hold for sequences for which $x_n \to -z_{1-\alpha/2}$. (As above, the assumption on T^* is assumed to hold with probability one with respect to the randomness in the data.)

Assumption (5.33) holds whenever the t statistic and the bootstrap t statistic have one-term Edgeworth expansions, just as in the case of symmetric confidence intervals. When $\kappa = 1/2$ and $\widehat{\sigma}$ is an $n^{1/2}$-consistent estimator of the asymptotic standard error of $\widehat{\theta}$, then (5.33) typically holds with $\xi = 1/2$ rather than $\xi = 1$, however, because there is no cancelling out of the $n^{-1/2}$ terms in the Edgeworth expansions of T and T^* as there are in the Edgeworth expansions of $|T|$ and $|T^*|$.

The asymptotic justifications of the three-step method of choosing B^* are in terms of the limit as $pdb \to 0$ and $n \to \infty$. We assume that $pdb \to 0$ sufficiently slowly that

$$pdb \times n^{\xi} \to \infty \text{ as } n \to \infty, \tag{5.34}$$

where ξ is as in (5.33).

The asymptotic justifications of the three-step methods for $\widehat{J}_{2,B^*}$ and $\widehat{J}_{1,B^*}$ are that

$$\begin{aligned} P^*\left(100\frac{|\widehat{q}_{1-\alpha,B_{1\ell}} - \widehat{q}_{1-\alpha,\infty}|}{\widehat{q}_{1-\alpha,\infty}} \leq pdb\right) &\to 1-\tau \text{ as } pdb \to 0 \text{ and } n \to \infty \text{ and} \\ P^*\left(100\frac{|\widehat{q}_{\alpha,B_{1u}} - \widehat{q}_{\alpha,\infty}|}{\widehat{q}_{\alpha,\infty}} \leq pdb\right) &\to 1-\tau \text{ as } pdb \to 0 \text{ and } n \to \infty, \text{ where} \end{aligned}$$

$$\begin{aligned} \widehat{q}_{1-\alpha,B_{1h}} &= T^*_{B_{1h},\nu_{1h}},\ \widehat{q}_{\alpha,B_{1h}} = T^*_{B_{1h},\eta_{1h}}, \\ \widehat{J}_{2,B_{1h}} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B_{1h}}, \widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{\alpha,B_{1h}}] \text{ for } h = \ell, u, \text{ and} \\ \widehat{J}_{1,B_{1\ell}} &= [\widehat{\theta} - n^{-\kappa}\widehat{\sigma}\widehat{q}_{1-\alpha,B_{1\ell}}, \infty). \end{aligned} \tag{5.35}$$

As above, the probability $P^*(\cdot)$ denotes probability with respect to the simulation randomness conditional on the infinite sequence of data vectors. Under the assumptions, this conditional result holds with probability one with respect to the randomness in the data.

Equation (5.35) implies that the three-step method for $\widehat{J}_{2,B^*}$ attains precisely the desired level of accuracy using "small *pdb* and large n" asymptotics when $B^* = \max(B_{1\ell}, B_{1u}) > B_0$. When $B^* = B_0 > \max(B_{1\ell}, B_{1u})$, then the accuracy of the three-step method exceeds the desired level of accuracy. An analogous statement holds for the three-step method for $\widehat{J}_{1,B^*}$.

The proof of (5.35) is the same as that given in Section 5.2 for symmetric percentile t confidence intervals except that $|T|$ and $|T^*|$ are replaced throughout by T and T^*, $\widehat{k}_{\alpha,B}$ and $\widehat{k}_{\alpha,\infty}$ are replaced throughout by either $\widehat{q}_{1-\alpha,B}$ and $\widehat{q}_{1-\alpha,\infty}$ or $\widehat{q}_{\alpha,B}$ and $\widehat{q}_{\alpha,\infty}$, and the formulae for a_0, B_0, c_α, and $\widehat{m}$ are changed to reflect the fact that we are estimating a density that asymptotically equals either $\phi(z_{1-\alpha})$ or $\phi(z_\alpha)$ rather than $2\phi(z_{1-\alpha/2})$.

5.4. Asymptotic Justification for the Three-step Method for Tests with a Specified Significance Level

The asymptotic justification of the three-step method for tests of significance level α introduced in Section 4.1 is analogous to the asymptotic justification given for symmetric percentile t confidence intervals in Section 3.3. First, we introduce an analogous strengthening of the assumption of convergence in distributuion to G of the test statistic T under the null hypothesis. We assume: For some $\xi > 0$ and all sequences of constants $\{x_n : n \geq 1\}$ for which $x_n \to q_{G,1-\alpha}$, we have

$$\begin{aligned} &P(T \leq x_n) = G(x_n) + O(n^{-\xi}) \text{ as } n \to \infty \text{ and} \\ &P^*(T^* \leq x_n) = G(x_n) + O(n^{-\xi}) \text{ as } n \to \infty \end{aligned} \tag{5.36}$$

when the null hypothesis is true. (The assumption on T^* is assumed to hold with probability one with respect to the randomness in the data.) Assumption (5.36) holds whenever the test statistic and the bootstrap test statistic have one-term Edgeworth expansions, just as in the case of symmetric confidence intervals. This assumption is widely applicable. See Hall and Horowitz (1996) for an example in which (5.36) holds with the asymptotic distribution G being a chi-squared distribution.

The asymptotic justification of the three-step method of choosing B^* is in terms of the limit as $pdb \to 0$ and $n \to \infty$. We assume that $pdb \to 0$ sufficiently slowly that

$$pdb \times n^{\xi} \to \infty \text{ as } n \to \infty, \tag{5.37}$$

where ξ is as in (5.36).

The asymptotic justification of the three-step method is that

$$P^*\left(100\frac{|\widehat{k}_{\alpha,B_1} - \widehat{k}_{\alpha,\infty}|}{\widehat{k}_{\alpha,\infty}} \leq pdb\right) \to 1-\tau \quad \text{as } pdb \to 0 \text{ and } n \to \infty, \tag{5.38}$$

where $\widehat{k}_{\alpha,B_1} = T^*_{B_1,\nu_1}$. Under the assumptions, the conditional result above holds with probability one with respect to the randomness in the data.

Equation (5.38) implies that the three-step method attains precisely the desired level of accuracy using "small *pdb* and large n" asymptotics when $B^* = B_1 \geq B_0$. When $B^* = B_0 > B_1$, then the accuracy of the three-step method exceeds the desired level of accuracy.

The proof of (5.38) is the same as that given in Section 5.2 for symmetric percentile t confidence intervals except that $|T|$ and $|T^*|$ are replaced throughout by T and T^* and the formulae for a_0, B_0, c_α, and $\widehat{m}$ are changed to reflect the fact that we are estimating a density that asymptotically equals $g(q_{G,1-\alpha})$ rather than $2\phi(z_{1-\alpha/2})$.

5.5. Proof of the p-value Results

All of the probabilistic statements below refer to the bootstrap simulation randomness conditional on the sample **X**. First, by the central limit theorem for iid random variables,

$$\frac{B(\widehat{p}_B - \widehat{p}_\infty)^2}{\widehat{p}_\infty(1-\widehat{p}_\infty)} \xrightarrow{d} \chi^2 \text{as } B \to \infty, \tag{5.39}$$

because $\widehat{p}_\infty$ does not equal zero or one.

Next, let $B_{11} = 10{,}000\chi^2_{1-\tau}(1-\widehat{p}_\infty)/(\widehat{p}_\infty pdb^2)$. Note that B_{11} is non-random. Equation (5.39) holds with B replaced by B_{11} and with the limit as $B \to \infty$ replaced by the limit as $pdb \to 0$ (because the latter forces $B_{11} \to \infty$). Now, by the central limit theorem of Doeblin–Anscombe for a sum of independent random variables with a random number of terms in the sum (e.g., see Chow and Teicher (1978, Thm. 9.4.1, p. 317)), provided $B_1/B_{11} \to_p 1$ as $pdb \to 0$, the result of (5.39) holds with B replaced by B_1 and with the limit as $B \to \infty$ replaced by the limit as $pdb \to 0$. With these changes, (5.39) can be rewritten using (4.16) as

$$\left(100\frac{|\widehat{p}_{B_1} - \widehat{p}_\infty|}{\widehat{p}_\infty}\right)^2 \left(\frac{\chi^2_{1-\tau}}{pdp^2}\right) \left(\frac{\widehat{p}_\infty(1-\widehat{p}_{B_0})}{(1-\widehat{p}_\infty)\widehat{p}_{B_0}}\right) \xrightarrow{d} \chi^2 \text{ as } pdb \to 0. \tag{5.40}$$

(The effect of the int($\cdot$) function in (4.16) is asymptotically negligible and, hence, is ignored in obtaining the previous equation from (5.39).) By the law of large numbers for iid bounded random variables, $\widehat{p}_{B_0} \to_p \widehat{p}_\infty$ as $pdb \to 0$ because $B_0 \to \infty$ as $pdb \to 0$. This and (5.40) yield

$$\left(100\frac{|\widehat{p}_{B_1} - \widehat{p}_\infty|}{\widehat{p}_\infty}\right)^2 \chi^2_{1-\tau}/pdp^2 \xrightarrow{d} \chi^2 \text{ as } pdb \to 0, \tag{5.41}$$

which establishes (4.18).

It remains to show that $B_1/B_{11} \to_p 1$ as $pdb \to 0$. This follows from $\widehat{p}_{B_0} \to_p \widehat{p}_\infty$, because $B_1/B_{11} = (\widehat{p}_\infty(1-\widehat{p}_{B_0}))/((1-\widehat{p}_\infty)\widehat{p}_{B_0})$.

6. Footnotes

[1]The authors thank Ariel Pakes, three referees, and the co-editor for helpful comments; Glena Ames for typing the manuscript; and Rosemarie Lewis for proofreading the manuscript. The first author acknowledges the research support of the National Science Foundation via grant number SBR–9410975. The second author acknowledges the research support of the National Science Foundation via grant number SBR–9320386.

[2]We received the following comment on an earlier version of this paper, which only considered bootstrap standard errors: "This paper addresses the wrong bootstrap statistics. Standard errors are not interesting quantities for statistical inference. A standard error is useful only if it can be used to obtain a confidence interval or test statistic."

We disagree with this comment. It is not in accord with standard statistical theory or practice. Point estimation accompanied by a measure dispersion of the distribution of the estimator, usually its standard error, is a perfectly valid and widely used method of statistical inference. For example, see Dawid's (1983) entry on "Statistical Inference" in the *Encyclopedia of Statistical Science* and Efron and Tibshirani's (1993, Ch. 6) discussion of the use of the bootstrap for estimating standard errors.

Point estimation with standard errors, confidence intervals, and tests are the three most commonly used forms of statistical inference. Each is a special case of statistical decision theory. None is right or wrong. Each is just more or less suitable for a given application. In practice, point estimation is probably the most common form of statistical inference used in econometrics and many other fields.

7. References

Beran, R. (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687–697.

Bloch, D. A. and J. L. Gastwirth (1968): "On a Simple Estimate of the Reciprocal of the Density Function," *Annals of Mathematical Statistics,* 39, 1083–1085.

Carlstein, E. (1986): "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence," *Annals of Statistics*, 14, 1171–1179.

Chow, Y. S. and H. Teicher (1978): *Probability Theory: Independence Interchangeability Martingales*. New York: Springer-Verlag.

Davidson, R. and J. G. MacKinnon (1997): "Bootstrap Tests: How Many Bootstraps?" unpublished working paper, Department of Economics, Queen's University, Kingston, Ontario.

Dawid, A. P. (1983): "Inference, Statistical: I," in *Encyclopedia of Statistical Sciences*, Vol. 4, edited by N. L. Johnson and S. Kotz. New York: Wiley.

Efron, B. (1987): "Better Bootstrap Confidence Intervals," (with discussion) *Journal of the American Statistical Association,* 82, 171–200.

Efron, B. and R. Tibshirani (1986): "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, 1, 54–77.

——— (1993): *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Freedman, D. (1981): "Bootstrapping Regression Models," *Annals of Statistics*, 9, 1218–1228.

Hall, P. (1986): "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval," *Annals of Statistics*, 14, 1453–1462.

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

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moment Estimators," *Econometrica,* 64, 891–916.

Hall, P. and M. A. Martin (1991): "On the Error Incurred Using the Bootstrap Variance Estimator when Constructing Confidence Intervals for Quantiles," *Journal of Multivariate Analysis,* 38, 70–81.

Hall, P. and S. J. Sheather (1988): "On the Distribution of a Studentized Quantile," *Journal of the Royal Statistical Society,* Series B, 50, 381–391.

Hall, P. and S. R. Wilson (1991): "Two Guidelines for Bootstrap Hypothesis Testing," *Biometrics*, 47, 757–762.

Kiefer, J. (1982): "Conditional Inference," in *Encyclopedia of Statistical Sciences*, Vol. 2, 103–109.

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

Lehmann, E. L. (1983): *Theory of Point Estimation*. New York: Wiley.

Li, H. and G. S. Maddala (1996): "Bootstrapping Time Series Models," *Econometric Reviews*, 15, 115–158.

Shorack, G. R. and J. A. Wellner (1986): *Empirical Processes with Applications to Statistics*. New York: Wiley.

Siddiqui, M. M. (1960): "Distribution of Quantiles in Samples from a Bivariate Population," *Journal of Research of the National Bureau of Standards*, B, 64, 145–150.

Table 1. Values of *pdb* as a function of $\hat{\gamma}_{2B}$ and B when $\tau = .05$ for Standard Errors

	B									
$\hat{\gamma}_{2B}$	10	25	50	100	200	350	500	750	1,000	2,000
0	44	28	20	14	10	7.4	6.2	5.1	4.4	3.1
1	54	34	24	17	12	9.1	7.6	6.2	5.4	3.8
2	62	39	28	20	14	10.5	8.8	7.2	6.2	4.4
3	69	44	31	22	15	11.7	10.0	8.0	6.9	4.9

Table 2. Values of B_1 as a function of $\hat{\gamma}_{2B_0}$ and *pdb* when $\tau = .05$ for Standard Errors

	pdb		
$\hat{\gamma}_{2B_0}$	20	10	5
0	48	192	768
1	72	288	1,152
2	96	384	1,536
3	120	480	1,920

Table 3. Monte Carlo Simulation Results for Standard Errors

A. Error Distribution $N(0,1)$

pdb	$1-\tau$	Empirical Level		B_0	B_1				$\widehat{\gamma}_{2B_0}$			
		B^*	B_1		Mean	Med	Min	Max	Mean	Med	Min	Max
20	.900	.891	.870	34	37	33	12	222	0.15	−0.07	−1.34	11.07
10	.900	.897	.890	136	158	148	76	1,297	0.32	0.17	−0.89	17.13
5	.900	.896	.894	542	645	620	417	4,449	0.38	0.28	−0.46	14.42
20	.950	.945	.931	48	54	49	19	386	0.21	0.01	−1.24	14.06
.10	.950	.947	.944	192	225	213	120	1,805	0.34	0.21	−0.76	16.80
5	.950	.951	.950	768	917	884	630	6,901	0.39	0.30	−0.36	15.97
20	.975	.973	.964	63	71	65	27	545	0.25	0.07	−1.14	15.35
10	.975	.973	.972	251	297	281	165	2,440	0.36	0.24	−0.69	17.44
5	.975	.975	.974	1,004	1,204	1,161	854	11,257	0.40	0.31	−0.30	20.42

B. Error Distribution t_5

pdb	$1-\tau$	Empirical Level		B_0	B_1				$\widehat{\gamma}_{2B_0}$			
		B^*	B_1		Mean	Med	Min	Max	Mean	Med	Min	Max
20	.900	.859	.848	34	45	39	13	276	0.65	0.29	−1.28	14.24
10	.900	.878	.876	136	208	189	88	1,568	1.07	0.78	−0.71	21.14
5	.900	.889	.889	542	874	828	516	6,885	1.22	1.05	−0.10	23.40
20	.950	.925	.917	48	67	59	21	456	0.79	0.43	−1.14	16.97
10	.950	.934	.934	192	300	276	135	2,227	1.12	0.87	−0.60	21.20
5	.950	.944	.944	768	1,244	1,189	781	9,516	1.24	1.10	0.03	22.78
20	.975	.960	.955	63	91	80	31	657	0.88	0.54	−1.03	18.92
10	.975	.966	.966	251	397	368	196	2,727	1.16	0.93	−0.44	19.73
5	.975	.971	.971	1,004	1,631	1,567	1,044	11,264	1.25	1.12	0.08	20.44

Note: The reported numbers are the averages over the simulations performed for 20 samples, each of which consists of 25 observations. For each sample, we carry out 2,500 Monte Carlo repetitions. γ_2 is calculated for each of the 20 samples using 250,000 bootstrap repetitions. The average of the 20 γ_2 values is .37 in part A of the Table and 1.26 in part B of the Table.

Table 4. Monte Carlo Simulation Results for the Bias-corrected Three-step Method for Standard Errors with Error Distribution t_5

pdb	$1-\tau$	Empirical Level		B_0	B_1				$\widehat{\gamma}_{2B_0R}$			
		B^*	B_1		Mean	Med	Min	Max	Mean	Med	Min	Max
20	.900	.887	.875	34	54	49	13	349	1.03	0.75	−1.31	17.98
10	.900	.892	.890	136	217	209	83	1,710	1.19	0.90	−0.84	23.01
5	.900	.899	.897	542	913	849	505	7,327	1.20	0.96	−0.27	24.77
20	.950	.945	.939	48	79	77	19	511	1.08	0.79	−1.20	18.85
10	.950	.947	.945	192	324	290	139	2,463	1.19	0.92	−0.67	23.37
5	.950	.949	.949	768	1,334	1,203	766	10,186	1.24	0.98	−0.18	24.34
20	.975	.972	.970	63	110	99	31	696	1.12	0.80	−1.08	18.59
10	.975	.973	.971	251	439	395	192	3,305	1.21	0.98	−0.55	24.39
5	.975	.974	.974	1,004	1,645	1,595	1,032	14,267	1.25	1.03	−0.07	24.36

Note: The reported numbers are the averages over the simulations performed for 20 samples, each of which consists of 25 observations. For each sample we carry out 2,500 Monte Carlo repetitions. γ_2 is calculated for each of the 20 samples using 250,000 bootstrap repetitions. The average of the 20 γ_2 values is 1.26. The value of R is 407.

Table 5. Values of a_0, B_0, ν_0, c_α, and $\widehat{m}$ as a Function of α, τ, and *pdb* for Symmetric Confidence Intervals

α, c_α	.01, .0837			.05, .2086			.10, .2993		
τ	.01	.05	.10	.01	.05	.10	.01	.05	.10
pdb = 5:									
a_0	48	28	20	120	70	49	208	121	85
B_0	4799	2799	1999	2399	1399	979	2079	1209	849
ν_0	4752	2772	1980	2280	1330	931	1872	1089	765
$\widehat{m}$	24	17	14	38	27	21	49	34	27
pdb = 10:									
a_0	12	7	5	30	18	13	52	31	22
B_0	1199	699	499	599	359	259	519	309	219
ν_0	1188	693	495	570	342	247	468	279	198
$\widehat{m}$	10	7	6	15	11	9	20	14	11
pdb = 15:									
a_0	6	4	3	14	8	6	24	14	10
B_0	599	399	299	279	159	119	239	139	99
ν_0	594	396	297	266	152	114	216	126	90
$\widehat{m}$	6	5	4	9	7	6	12	9	7

Note: All quantities are defined in the three-step procedure of Section 3.2.

Table 6. Monte Carlo Simulation Results for Symmetric Confidence Intervals

A. Error Distribution $N(0,1)$

$1-\alpha$	pdb	$1-\tau$	Empirical Level		B_0	B_1			
			B^*	B_1		Mean	Med	Min	Max
.90	15	.90	.946	.943	99	258	216	28	1837
.90	10	.90	.924	.920	219	394	364	74	1482
.90	5	.90	.907	.905	849	1317	1280	481	3003
.90	15	.95	.970	.968	139	309	273	47	1532
.90	10	.95	.960	.957	309	524	493	124	1652
.90	5	.95	.952	.951	1209	1825	1785	756	3829
.95	15	.90	.952	.949	119	564	360	31	16709
.95	10	.90	.947	.946	259	754	654	104	4346
.95	5	.90	.915	.915	979	1920	1843	591	5104
.95	15	.95	.989	.989	159	1228	804	69	35579
.95	10	.95	.969	.968	359	884	801	159	3809
.95	5	.95	.955	.955	1399	2611	2531	947	6046

B. Error Distribution t_5

$1-\alpha$	pdb	$1-\tau$	Empirical Level		B_0	B_1			
			B^*	B_1		Mean	Med	Min	Max
.90	15	.90	.945	.942	99	275	230	29	1927
.90	10	.90	.924	.920	219	418	385	79	1560
.90	5	.90	.908	.907	849	1388	1348	505	3196
.90	15	.95	.969	.967	139	329	291	48	1686
.90	10	.95	.959	.957	309	555	521	130	1792
.90	5	.95	.953	.952	1209	1922	1878	792	4048
.95	15	.90	.950	.948	119	587	377	32	18426
.95	10	.90	.947	.946	259	800	696	107	4439
.95	5	.90	.917	.916	979	2055	1972	635	5320
.95	15	.95	.989	.989	159	1274	839	75	39633
.95	10	.95	.969	.968	359	941	854	163	3958
.95	5	.95	.957	.956	1399	2799	2714	1007	6530

Note: The reported numbers are the averages over the simulations performed for 250 samples, each of which consists of 25 observations. For each sample we carry out 2,000 Monte Carlo repetitions.

Table 7. Values of a_0, B_0, ν_0, η_0, c_α, and $\widehat{m}$ as a Function of α, τ, and *pdb* for Equal-tailed and One-sided Confidence Intervals

α, c_α	.01, .0838			.025, .1436			.05, .2122			.10, .3074		
τ	.01	.05	.10	.01	.05	.10	.01	.05	.10	.01	.05	.10
$pdb = 5$:												
a_0	69	40	29	124	72	51	219	127	90	473	274	194
B_0	6899	3999	2899	4959	2879	2039	4379	2539	1799	4729	2739	1939
ν_0	6831	3960	2871	4836	2808	1989	4161	2413	1710	4257	2466	1746
η_0	69	40	29	124	72	51	219	127	90	473	274	194
$\widehat{m}$	31	22	18	42	30	24	57	40	32	87	61	48
$pdb = 10$:												
a_0	18	10	8	31	18	13	55	32	23	119	69	49
B_0	1799	999	799	1239	719	519	1099	639	459	1189	689	489
ν_0	1782	990	792	1209	702	507	1045	608	437	1071	621	441
η_0	18	10	8	31	18	13	55	32	23	119	69	49
$\widehat{m}$	13	9	8	17	12	10	23	16	13	35	24	20
$pdb = 15$:												
a_0	8	5	4	14	8	6	25	15	10	53	31	22
B_0	799	499	399	559	319	239	499	299	199	529	309	219
ν_0	792	495	396	546	312	234	475	285	190	477	279	198
η_0	8	5	4	14	8	6	25	15	10	53	31	22
$\widehat{m}$	8	6	5	10	7	6	14	10	8	21	15	12

Note: All quantities are defined in the three-step procedure of Section 3.5.

Table 8. Values of a_0, B_0, ν_0, c_α, and $\widehat{m}$ as a Function of α and τ for $pdb = 10$ for Tests with Significance Level α

A. Test Statistics with Absolute $N(0,1)$ Asymptotic Null Distribution

α, c_α	.01, .0837			.05, .2086			.10, .2993		
τ	.01	.05	.10	.01	.05	.10	.01	.05	.10
$pdb = 10$:									
a_0	12	7	5	30	18	13	52	31	22
B_0	1199	699	499	599	359	259	519	309	219
ν_0	1188	693	495	570	342	247	468	279	198
$\widehat{m}$	10	7	6	15	11	9	20	14	11

B. Test Statistics with $N(0,1)$ Asymptotic Null Distribution

α, c_α	.01, .0838			.05, .2122			.10, .3074		
τ	.01	.05	.10	.01	.05	.10	.01	.05	.10
$pdb = 10$:									
a_0	18	10	8	55	32	23	119	69	49
B_0	1799	999	799	1099	639	459	1189	689	489
ν_0	1782	990	792	1045	608	437	1071	621	441
$\widehat{m}$	13	9	8	23	16	13	35	24	20

C. Test Statistics with χ^2_5 Asymptotic Null Distribution

α, c_α	.01, .0800			.05, .1963			.10, .2820		
τ	.01	.05	.10	.01	.05	.10	.01	.05	.10
$pdb = 10$:									
a_0	18	10	7	35	21	15	52	30	22
B_0	1799	999	699	699	419	299	519	299	219
ν_0	1782	990	693	665	399	285	468	270	198
$\widehat{m}$	12	8	7	16	11	9	19	13	11

D. Test Statistics with χ^2_{15} Asymptotic Null Distribution

α, c_α	.01, .0811			.05, .2022			.10, 2912		
τ	.01	.05	.10	.01	.05	.10	.01	.05	.10
$pdb = 10$:									
a_0	8	5	4	14	9	6	20	12	9
B_0	799	499	399	279	179	119	199	119	89
ν_0	792	495	396	266	171	114	180	108	81
$\widehat{m}$	7	6	5	9	7	5	10	8	6

Note: All quantities are defined in the three-step method of Section 4.1.

Table 9. Values of B_0 as a Function of τ, *pdb*, and $1 - G(T)$ for p-values

$1 - G(T)$	*pdb*										
	5	10	15	20	30	40	50	100	150	200	300
$\tau = .01$:											
.001								6624	2944	1656	736
.005							5278	1320	587	330	147
.010					7293	4103	2626	657	292		
.025			11492	6465	2873	1617	1035				
.050	50388	12597	5599	3150	1400	788					
.10	23869	5968	2653	1492	664						
.15	15028	3757	1670	940	418						
.20	10609	2653	1179	664	295						
.30			688	387	172	97	62				
.50				166	74	42	27				
.70					32	18	12				
.90					9	5	3				
$\tau = .05$:											
.001								3837	1705	960	427
.005							3057	765	340	192	85
.010					4224	2376	1521	381	169		
.025			6656	3744	1664	936	600				
.050	29184	7296	3243	1824	811	456					
.10	13825	3457	1537	865	385						
.15	8704	2176	968	544	242						
.20	6145	1537	683	385	171						
.30			399	224	100	56	36				
.50			171	96	43	24	16				
.70					19	11	7				
.90					5	3	2				
$\tau = .10$:											
.001								2708	1204	677	301
.005							2158	540	240	135	60
.010					2981	1677	1074	269	120		
.025			4698	2643	1175	661	423				
.050	20596	5149	2289	1288	573	322					
.10	9757	2440	1085	610	272						
.15	6143	1536	683	384	171						
.20	4337	1085	482	272	121						
.30			282	159	71	40	26				
.50				68	31	17	11				
.70					13	8	5				
.90					4	2	2				

Note: All quantities are defined in the three-step method of Section 4.2.