

**EVALUATION OF A THREE-STEP METHOD  
FOR CHOOSING THE NUMBER OF BOOTSTRAP REPETITIONS**

**BY**

**DONALD W. K. ANDREWS AND MOSHE BUCHINSKY**

**COWLES FOUNDATION PAPER NO. 1125**



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS  
YALE UNIVERSITY**

**Box 208281**

**New Haven, Connecticut 06520-8281**

**2006**

**<http://cowles.econ.yale.edu/>**



ELSEVIER

Journal of Econometrics 103 (2001) 345–386

---

---

JOURNAL OF  
Econometrics

---

---

www.elsevier.com/locate/econbase

## Evaluation of a three-step method for choosing the number of bootstrap repetitions

Donald W.K. Andrews<sup>a</sup>, Moshe Buchinsky<sup>b,c,d,\*</sup>

<sup>a</sup>*Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281, USA*

<sup>b</sup>*Department of Economics, Brown University, Box B, Providence, RI 02912, USA*

<sup>c</sup>*National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138, USA*

<sup>d</sup>*INSEE-CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France*

Accepted 24 October 2000

---

### Abstract

This paper provides a variety of Monte Carlo simulations that evaluate the finite-sample performance of the three-step method for choosing the number of bootstrap repetitions, suggested by Andrews and Buchinsky (*Econometrica* 67 (2000) 23–51). The simulations cover bootstrap standard errors, confidence intervals, tests, and  $p$ -values. Three commonly used econometric applications are considered: linear regression, binary probit, and quantile regression. In brief, we find that the three-step method works very well in all of the contexts examined here. We also find that the number of bootstrap repetitions commonly used in econometric applications is much less than needed to achieve accurate bootstrap quantities. © 2001 Elsevier Science S.A. All rights reserved.

*JEL classification:* C12; C13; C14; C15

*Keywords:* Bootstrap; Bootstrap repetitions; Coefficient of excess kurtosis; Confidence interval; Density estimation; Hypothesis test;  $p$ -value; Quantile; Simulation; Standard error estimate

---

---

\* Corresponding author. Tel.: +1-401-863-2951; fax: +1-401-863-1970.

*E-mail address:* moshe\_buchinsky@brown.edu (M. Buchinsky).

## 1. Introduction

Andrews and Buchinsky (2000) consider the problem of choosing the number of bootstrap repetitions  $B$  for a wide variety of bootstrap procedures. They introduce a three-step method for doing so. This method is designed to address the problem that one can obtain a ‘different answer’ from the same data merely by using different simulation draws if  $B$  is too small, but computational costs can be great if  $B$  is chosen to be extremely large. The three-step method of Andrews and Buchinsky (2000) determines  $B$  to attain a specified level of accuracy. In consequence, one can obtain accurate bootstrap quantities with the minimum computational effort. The method is justified by asymptotics as  $B \rightarrow \infty$ .

The primary purpose of this paper is to investigate the finite sample properties of the three-step method. We address the question of whether the three-step method delivers the desired level of accuracy in finite samples. A secondary purpose of this paper, independent of the three-step method, is to determine the magnitudes of  $B$  necessary to obtain different levels of accuracy in a variety of bootstrap situations. The results let one judge whether typical choices for  $B$  used in the literature are appropriate.

We investigate the finite sample properties of the three-step method in a variety of different contexts. We consider bootstrap standard error estimates, symmetric two-sided confidence intervals, tests with a given significance level  $\alpha$ , and  $p$ -values. We consider these bootstrap applications in a linear regression model, a binary probit model, and a quantile regression model. In each model, the observations are independent and identically distributed (iid) and the sample size is taken to be small, only 25 observations. In all cases, we consider the standard nonparametric bootstrap based on the empirical distribution function.

The measure of ‘accuracy’ used by the three-step method is the percentage deviation of the bootstrap quantity of interest based on  $B$  bootstrap repetitions, from the ideal bootstrap quantity for which  $B = \infty$ . For the four bootstrap applications considered here, the bootstrap ‘quantities of interest’ are the standard error estimate, the length of the confidence interval, the critical value of the test, and the  $p$ -value. For example, for standard error estimates, accuracy is measured in terms of the percentage deviation of the bootstrap standard error estimate for a given (finite) value of  $B$ , from the ideal bootstrap standard error estimate.

The percentage deviation of any bootstrap quantity for a given value 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 be less than this bound with a specified probability,  $1 - \tau$ , close to one. The three-step method takes  $pdb$  and  $\tau$  as given and provides a data-dependent method to

determine a value of  $B$ , denoted  $B^*$ , to obtain the desired level of accuracy. Three steps are required because the relevant features of the problem need to be determined in the initial two steps before it is possible to determine a suitable choice of  $B$  in the third step.

In the simulations, we assess the precision of the three-step method as follows. For each simulation, we calculate whether the actual percentage deviation of the bootstrap quantity based on the value of  $B$  selected via the three-step method is less than the percentage deviation bound  $pdb$ . Then, we compare the fraction of cases over all of the Monte Carlo simulations where this is true, denoted the *empirical level*, with the nominal level  $1 - \tau$ . The three-step method performs well if the empirical level is close to  $1 - \tau$ .

The results indicate that in most cases the empirical levels are quite close to the nominal levels. For example, for  $(pdb, \tau) = (10, 0.05)$ , we obtain empirical levels of 0.947, 0.949, and 0.942 for bootstrap standard error estimates in the linear regression model with errors with  $t$  distribution with five degrees of freedom ( $t_5$ ), the binary probit model, and the quantile regression model with  $t_5$  errors, respectively, in comparison to the nominal level of 0.950. For symmetric 90% confidence intervals, the corresponding empirical levels are 0.958, 0.958, and 0.958. For tests with significance level 0.05, the corresponding empirical levels are 0.940, 0.947, and 0.942, respectively. For  $p$ -values with  $p=0.10$ , the corresponding empirical levels are 0.951, 0.944, and 0.947. In general, the empirical levels for the linear and quantile regression models with normal, rather than  $t_5$ , errors are even closer to the 0.950 nominal level.

The simulation results show that the precision of the three-step method does not vary greatly across the different models considered. The results also show that the precision varies somewhat across the different type of bootstrap application considered, with standard errors being the best and  $p$ -values being the worst, but that the variation is not too great.

The simulation results clearly indicate that the precision of the three-step method depends primarily on how tight the  $(pdb, \tau)$  bound is. The smaller the values of  $pdb$  and  $\tau$ , the greater is the required number of bootstrap repetitions  $B^*$ , and the greater is the precision of the three-step method. The reason is that the three-step method is based on asymptotics as  $B \rightarrow \infty$ .

Overall, we conclude that the three-step method works very well over the range of bootstrap applications and models that are considered in the simulations.

We note that the three-step method is applicable in numerous cases that are not considered in this paper. It applies to bootstrap equal-tailed percentile  $t$  confidence intervals, one-sided percentile  $t$  confidence intervals, confidence regions, and bias-correction. It applies to parametric and semiparametric bootstraps for iid and temporally dependent samples, to residual-based regression bootstraps, as well as to nonparametric block bootstraps for temporally

dependent samples. See Andrews and Buchinsky (2000) for details. For want of time and space, we do not consider these cases here.

Next, we discuss the magnitude of the  $B$  values that are needed to obtain accurate bootstrap quantities. In the econometrics literature, it is common for 100 or so bootstrap repetitions to be used. A number of this magnitude is noticeably smaller than the numbers obtained in the simulations. For example, for  $(pdb, \tau) = (10, 0.05)$ , we obtain the following median values (over the simulations) of the  $B^*$  values selected by the three-step method for the same models as discussed above: 287, 207, and 291 for bootstrap standard error estimates, 511, 409, and 543 for symmetric 90% confidence intervals, 767, 828, and 834 for tests with significance level 0.05, and 3580, 3690, and 3622 for  $p$ -values with  $p = 0.10$ . Note that these median  $B^*$  values are very good indicators of the median values of  $B$  that are *necessary* to obtain a  $(pdb, \tau)$  accuracy of *exactly*  $(10, 0.95)$ , because the empirical levels of the three-step method are quite close to the nominal level of 0.95.

These median  $B^*$  values vary considerably across the different bootstrap applications considered and with the specified degree of accuracy  $(pdb, \tau)$  within each application. They also vary somewhat across the different models considered. In the case of  $p$ -values, the level of accuracy given by  $(pdb, \tau) = (10, 0.05)$  may be more than one requires. In this case the large values of  $B$  given above would be replaced by smaller values, when larger  $(pdb, \tau)$  values are specified. Nevertheless, the results indicate that if the specified level of accuracy is desired, then the number of bootstrap repetitions required can be quite large.

We conclude that to obtain results that do not depend on the particular bootstrap simulation draws employed, one needs to use more bootstrap repetitions than is commonly used in the econometrics literature. How many more depends on the type of bootstrap application, the model under consideration, and the desired level of accuracy.

Papers in the literature that are related to the three-step method considered here include Efron and Tibshirani (1986), Hall (1986), Davison and Hinkley (1997, Sections 2.5.2 and 4.2.5), Davidson and MacKinnon (2000), and Andrews and Buchinsky (2000, 2001).

Efron and Tibshirani (1986, Section 9) 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 unknown parameters that are not estimable. Hence, Efron and Tibshirani use their formula to suggest a range of plausible values of  $B$ , rather than a specific value of  $B$ .

Hall (1986) considers *unconditional* coverage probabilities of confidence intervals, i.e., coverage probabilities with respect to the randomness in the data and the bootstrap simulations. The three-step method considered here

focuses on *conditional* coverage probabilities, i.e., coverage probabilities with respect to the randomness in the data conditional on the bootstrap simulations. The reason is that one does not want to be able to obtain ‘different answers’ from the same data due to the use of different simulation draws.

Davison and Hinkley (1997, Section 2.5.2) provide formulae that decompose the variance of bootstrap bias correction estimates, variance estimates, and quantile estimates into the part that is due to simulation and the part that is due to sample variation. They use these formulae to suggest values of  $B$ . Davison and Hinkley (1997, Section 4.5.2) provide some formulae for the effect of  $B$  on the power of a test.

Davidson and MacKinnon (2000) propose a pretesting method of choosing  $B$  for a test with a given significance level  $\alpha$  that aims to ensure that the probability is small that there is a difference between the conclusions of the ideal bootstrap test and the bootstrap test based on  $B$  bootstrap repetitions. In contrast, the three-step method aims to achieve a bootstrap test that has good conditional significance level and power conditional on the simulation randomness by determining an accurate critical value.

Andrews and Buchinsky (2001) provides a three-step method for choosing  $B$  for the  $BC_a$  confidence intervals of Efron (1987). The method is analogous to the three-step method considered here for percentile  $t$  confidence intervals.

The remainder of this paper is organized as follows. Section 2 provides the notation and describes the bootstrap applications of interest, viz., standard error estimates, symmetric two-sided percentile  $t$  confidence intervals, tests for a given significance level  $\alpha$ , and  $p$ -values. Section 3 outlines the three-step method. It also provides a number of tables that illustrate the magnitudes of various quantities that enter the calculations in the three-step method. Section 4 explains the design of the Monte Carlo experiments. Section 5 provides the results of the Monte Carlo experiment. Section 6 provides a brief summary and concluding remarks.

## **2. Notation and description of the bootstrap applications**

In this section, we present the notation used throughout the paper and introduce the bootstrap applications considered in the paper, viz., standard error estimates, confidence intervals, tests for a given significance level, and  $p$ -values.

### *2.1. Notation*

First, we outline the general framework. Suppose that we are interested in some unknown quantity  $\lambda$ . For example,  $\lambda$  could be an *exact* standard error, confidence interval length, critical value, or  $p$ -value. We would like to

estimate  $\lambda$  using an ‘ideal’ bootstrap estimate, denoted  $\hat{\lambda}_\infty$ . Analytic calculation of  $\hat{\lambda}_\infty$  is intractable in most cases, so we use an estimate  $\hat{\lambda}_B$  of  $\hat{\lambda}_\infty$  that is based on a finite number,  $B$ , of bootstrap simulations. The three-step method of Andrews and Buchinsky (2000) specifies a data-dependent method of selecting  $B$  such that  $\hat{\lambda}_B$  is close to  $\hat{\lambda}_\infty$  within a prespecified level of accuracy. We describe this method below.

The observed data are a sample of size  $n$ :  $\mathbf{X} = (X_1, \dots, X_n)'$ . Let  $\mathbf{X}^* = (X_1^*, \dots, X_n^*)'$  be a bootstrap sample of size  $n$  based on the original sample  $\mathbf{X}$ . In this paper, we consider the case where  $\mathbf{X}$  is a sample of iid random vectors and the bootstrap sample  $\mathbf{X}^*$  is an iid sample drawn from the empirical distribution  $\hat{F}$  (i.e., a random sample of size  $n$  drawn from the original sample with replacement). This is the most commonly used bootstrap. Let  $\{\mathbf{X}_b^*: b = 1, \dots, B\}$  denote  $B$  iid bootstrap samples, each with the same distribution as  $\mathbf{X}^*$ . All probability statements and the probability and expectation operators  $P^*$  and  $E^*$ , respectively, refer to the randomness in the iid bootstrap samples  $\{\mathbf{X}_b^*: b = 1, \dots, B\}$  conditional on the observed data  $\mathbf{X}$ .

The accuracy of  $\hat{\lambda}_B$  is measured by the percentage deviation of  $\hat{\lambda}_B$  from  $\hat{\lambda}_\infty$ :

$$100 \frac{|\hat{\lambda}_B - \hat{\lambda}_\infty|}{\hat{\lambda}_\infty}. \quad (1)$$

This percentage deviation is random conditional on the sample  $\mathbf{X}$ , because it depends on the random bootstrap simulations that are used to calculate  $\hat{\lambda}_B$ . Let  $1 - \tau$  denote a probability close to one, such as 0.95. Let  $pdb$  be a bound on the percentage deviation of  $\hat{\lambda}_B$  from  $\hat{\lambda}_\infty$ . The three-step method of Andrews and Buchinsky (2000) is designed to determine  $B$  such that

$$P^* \left( 100 \frac{|\hat{\lambda}_B - \hat{\lambda}_\infty|}{\hat{\lambda}_\infty} \leq pdb \right) \approx 1 - \tau, \quad (2)$$

where  $\approx$  denotes ‘is approximately equal to’.

The three-step method is based on the following asymptotic result:

$$B^{1/2}(\hat{\lambda}_B - \hat{\lambda}_\infty)/\hat{\lambda}_\infty \rightarrow_d N(0, \omega) \quad \text{as } B \rightarrow \infty, \quad (3)$$

where the asymptotic variance  $\omega$  depends on the particular application considered.<sup>1</sup> The three-step method depends on an estimator  $\hat{\omega}_B$  of  $\omega$ . This estimator is based on the bootstrap samples  $\{\mathbf{X}_b^*: b = 1, \dots, B\}$ .

<sup>1</sup> This result holds with probability one with respect to the distribution of the original sample. In the examples in which  $\hat{\lambda}_B$  is a sample quantile, viz., the confidence interval and test for a given significance level examples, this result holds as both  $B \rightarrow \infty$  and  $n \rightarrow \infty$  and it holds with probability one with respect to the distribution of the infinite sequence of random variables that yields the original samples for different values of  $n$ .

In the following subsections, we specify the quantities  $\lambda$ ,  $\hat{\lambda}_\infty$ ,  $\hat{\lambda}_B$ ,  $\omega$ , and  $\hat{\omega}_B$  in each of the applications of interest.

2.2. Standard errors

Let  $\hat{\theta} = \hat{\theta}(\mathbf{X})$  be an estimator of a scalar parameter  $\theta_0$  based on the sample  $\mathbf{X}$ . For standard error estimates, the quantity  $\lambda$  is the standard error, *se*, of  $\hat{\theta}$ :

$$se = (E(\hat{\theta}(\mathbf{X}) - E\hat{\theta}(\mathbf{X}))^2)^{1/2}, \tag{4}$$

where  $E$  denotes expectation with respect to the randomness in  $\mathbf{X}$ .

Let  $\hat{\theta}_b^* = \hat{\theta}(\mathbf{X}_b^*)$  for  $b = 1, \dots, B$  denote  $B$  bootstrap estimates of  $\theta_0$ . The ‘ideal’ bootstrap standard error estimator of *se* is given by

$$\hat{se}_\infty = (E^*(\hat{\theta}_b^* - E^*\hat{\theta}_b^*)^2)^{1/2}. \tag{5}$$

The bootstrap standard error estimator based on  $B$  bootstrap repetitions is

$$\hat{se}_B = \left( \frac{1}{B-1} \sum_{b=1}^B \left( \hat{\theta}_b^* - \frac{1}{B} \sum_{c=1}^B \hat{\theta}_c^* \right)^2 \right)^{1/2}. \tag{6}$$

In this case,  $\hat{\lambda}_\infty = \hat{se}_\infty$  and  $\hat{\lambda}_B = \hat{se}_B$ . Provided  $E^*((\hat{\theta}_b^*)^2) < \infty$ ,  $\lim_{B \rightarrow \infty} \hat{se}_B = \hat{se}_\infty$  almost surely by the law of large numbers.

In this application, the variance of (3) depends on the *coefficient of excess kurtosis*, denoted  $\gamma_2$ , of the bootstrap estimator  $\hat{\theta}_b^*$ .<sup>2</sup> In particular,

$$\omega = (2 + \gamma_2)/4, \text{ where } \gamma_2 = E^*(\hat{\theta}_b^* - \mu)^4 / \hat{se}_\infty^4 - 3 \text{ and } \mu = E^*\hat{\theta}_b^*. \tag{7}$$

A consistent estimator of  $\omega$  is

$$\hat{\omega}_B = (2 + \hat{\gamma}_{2B})/4, \text{ where } \hat{\gamma}_{2B} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\mu}_B)^4 / \hat{se}_B^4 - 3$$

$$\text{and } \hat{\mu}_B = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*. \tag{8}$$

By the law of large numbers and Slutsky’s Theorem, it follows that  $\lim_{B \rightarrow \infty} \hat{\mu}_B = \mu$ ,  $\lim_{B \rightarrow \infty} \hat{\gamma}_{2B} = \gamma_2$ , and  $\lim_{B \rightarrow \infty} \hat{\omega}_B = \omega$  almost surely, provided  $\hat{se}_\infty \neq 0$  and  $E^*((\hat{\theta}_b^*)^4) < \infty$ .

The estimator  $\hat{\omega}_B$  tends to be biased toward zero in small samples, so we also consider the bootstrap *bias-corrected* version of  $\hat{\omega}_B$  as an estimator of  $\omega$ . The iid sample of  $B$  bootstrap estimates of  $\theta_0$  is  $\Theta_B^* = (\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$ . For

<sup>2</sup> If  $\hat{\theta}_b^*$  has a normal distribution then  $\gamma_2 = 0$ , if  $\hat{\theta}_b^*$  has kurtosis greater than that of a normal distribution then  $\gamma_2 > 0$ , and  $\gamma_2 < 0$  otherwise.



present purposes, we think of  $(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$  as being the original sample and  $\hat{\gamma}_{2B}$  as being an estimator based on this sample that we want to bootstrap bias correct. Let  $\hat{G}$  denote the empirical distribution of  $(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$ . Consider  $R$  independent bootstrap samples  $\{\Theta_{Br}^{**}: r = 1, \dots, R\}$ , where each bootstrap sample  $\Theta_{Br}^{**} = (\hat{\theta}_{1r}^{**}, \dots, \hat{\theta}_{Br}^{**})$  is a random sample of size  $B$  drawn from  $\hat{G}$ . The bootstrap bias-corrected estimator  $\hat{\gamma}_{2BR}$  of  $\gamma_2$  for  $R$  bootstrap repetitions is

$$\hat{\gamma}_{2BR} = 2\hat{\gamma}_{2B} - \frac{1}{R} \sum_{r=1}^R \hat{\gamma}_2(\Theta_{Br}^{**}), \quad \text{where}$$

$$\hat{\gamma}_2(\Theta_{Br}^{**}) = \frac{[1/(B-1)] \sum_{b=1}^B (\hat{\theta}_{br}^{**} - (1/B) \sum_{c=1}^B \hat{\theta}_{cr}^{**})^4}{([1/(B-1)] \sum_{b=1}^B (\hat{\theta}_{br}^{**} - (1/B) \sum_{c=1}^B \hat{\theta}_{cr}^{**})^2)^2} - 3. \quad (9)$$

### 2.3. Symmetric two-sided percentile $t$ confidence intervals

Next, we consider symmetric two-sided percentile  $t$  confidence intervals for the scalar parameter  $\theta_0$ . These intervals are symmetric about the estimator  $\hat{\theta}$ . In the models that we consider below, the normalized estimator  $n^{1/2}(\hat{\theta} - \theta_0)$  has an asymptotic normal distribution as  $n \rightarrow \infty$ . Let  $\hat{\sigma} = \hat{\sigma}(\mathbf{X})$  denote a consistent estimator of the asymptotic standard error of  $n^{1/2}(\hat{\theta} - \theta_0)$ . Let

$$T = |n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma}|. \quad (10)$$

Let  $q_{1-\alpha}^0$  denote the  $1-\alpha$  quantile of  $T$ . The ‘theoretical’ symmetric two-sided percentile  $t$  confidence interval with exact confidence level  $100(1-\alpha)\%$  is

$$J_{SY} = [\hat{\theta} - n^{-1/2} \hat{\sigma} q_{1-\alpha}^0, \hat{\theta} + n^{-1/2} \hat{\sigma} q_{1-\alpha}^0]. \quad (11)$$

The quantity  $\lambda$  of interest in this case is  $q_{1-\alpha}^0$ , which is proportional to the length of the confidence interval.

Define  $\hat{\sigma}_b^* = \hat{\sigma}(\mathbf{X}_b^*)$  and  $T_b^* = |n^{1/2}(\hat{\theta}_b^* - \hat{\theta})/\hat{\sigma}_b^*|$  for  $b = 1, \dots, B$ . The  $1-\alpha$  quantile of  $T_b^*$ , denoted  $\hat{q}_{1-\alpha, \infty}$ , is the ideal bootstrap estimate of  $q_{1-\alpha}^0$ . Thus,  $\hat{\lambda}_\infty$  equals  $\hat{q}_{1-\alpha, \infty}$  in this application. The ideal bootstrap symmetric percentile  $t$  confidence interval of approximate confidence level  $100(1-\alpha)\%$  is  $\hat{J}_{SY, \infty} = [\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{q}_{1-\alpha, \infty}, \hat{\theta} + n^{-1/2} \hat{\sigma} \hat{q}_{1-\alpha, \infty}]$ .<sup>3</sup>

Let  $\hat{q}_{1-\alpha, B}$  denote the  $1-\alpha$  sample quantile of the  $B$  bootstrap  $t$  statistics  $\{T_b^*: b = 1, \dots, B\}$  (defined more precisely below). In this application,  $\hat{\lambda}_B$  equals  $\hat{q}_{1-\alpha, B}$ . The bootstrap symmetric percentile  $t$  confidence interval of

<sup>3</sup> The confidence level of this bootstrap confidence interval exhibits higher order improvements over the corresponding confidence level based on the delta method; e.g., Beran (1988) and Hall (1992).

approximate confidence level  $100(1 - \alpha)\%$  based on  $B$  bootstrap repetitions is

$$\hat{J}_{SY,B} = [\hat{\theta} - n^{-1/2} \hat{\sigma} \hat{q}_{1-\alpha,B}, \hat{\theta} + n^{-1/2} \hat{\sigma} \hat{q}_{1-\alpha,B}]. \tag{12}$$

Following Hall (1992, p. 307), for this application, we choose  $B$  so that  $v/(B + 1) = 1 - \alpha$  for some positive integer  $v$ . We consider values of  $\alpha$  that are rational and can be written as

$$\alpha = \alpha_1/\alpha_2 \tag{13}$$

for some positive integers  $\alpha_1$  and  $\alpha_2$  (with no common integer divisors). Then,  $B = \alpha_2 h - 1$  and  $v = (\alpha_2 - \alpha_1)h$  for some positive integer  $h$ . Let  $\{T_{B,b}^*; b = 1, \dots, B\}$  denote the ordered sample of bootstrap  $T$  statistics. Then, for  $B$  and  $\alpha$  as above, the  $1 - \alpha$  sample quantile  $\hat{q}_{1-\alpha,B}$  of  $\{T_b^*; b = 1, \dots, B\}$  is

$$\hat{q}_{1-\alpha,B} = T_{B,v}^*. \tag{14}$$

That is,  $\hat{q}_{1-\alpha,B}$  is the  $v$ th order statistic of  $\{T_b^*; b = 1, \dots, B\}$ .

In this application,  $\omega$  is given by

$$\omega = \alpha(1 - \alpha)/(4z_{1-\alpha/2}^2 \phi^2(z_{1-\alpha/2})), \tag{15}$$

where  $z_{1-\alpha/2}$  and  $\phi(\cdot)$  denote the  $1 - \alpha/2$  quantile and the density function, respectively, of the standard normal distribution. The estimate  $\hat{\omega}_B$  is

$$\hat{\omega}_B = \alpha(1 - \alpha)(1/\hat{g}_B)^2/\hat{q}_{1-\alpha,B}^2, \quad \text{where}$$

$$1/\hat{g}_B = \frac{B}{2\hat{m}_B} (T_{B,v+\hat{m}_B}^* - T_{B,v-\hat{m}_B}^*), \quad v = (B + 1)(1 - \alpha),$$

$$\hat{m}_B = \text{int}(c_\alpha B^{2/3}) \quad \text{and} \quad c_\alpha = \left( \frac{6z_{1-\alpha/2}^2 \phi^2(z_{1-\alpha/2})}{2z_{1-\alpha/2}^2 + 1} \right)^{1/3}. \tag{16}$$

Note that  $1/\hat{g}_B$  is Siddiqui's (1960) estimator of the reciprocal of the density of  $T_b^*$  with a plug-in estimator of the bandwidth parameter, viz.,  $\hat{m}_B$ .<sup>4</sup>

Symmetric confidence intervals are appropriate only if the distribution of the  $t$  statistic upon which the confidence interval is based has a distribution that is approximately symmetric. The asymptotic distribution of the  $t$  statistic is normal in most cases. So, in large samples, its distribution is approximately symmetric. In small samples, however, its distribution may not be approximately symmetric. If the  $t$  statistic has a noticeably asymmetric distribution, then a symmetric confidence interval may be misleading. In such a case, an equal-tailed two-sided confidence interval is more appropriate. Andrews and Buchinsky (2000) describe a three-step for choosing  $B$  for equal-tailed percentile  $t$  confidence intervals. Andrews and Buchinsky (2001) describe

<sup>4</sup> This estimator has been analyzed by Bloch and Gastwirth (1968) and Hall and Sheather (1988).

a three-step for choosing  $B$  for the equal-tailed  $BC_a$  confidence intervals of Efron (1987).

### 2.4. Tests for a given significance level

Here, we consider one-sided tests for a given significance level  $\alpha$ . The null and alternative hypotheses are

$$H_0: \theta_0 = 0 \quad \text{and} \quad H_1: \theta_0 > 0. \tag{17}$$

The test statistic considered in this case is

$$T = n^{1/2}(\hat{\theta} - \theta_0)/\hat{\sigma}, \tag{18}$$

where  $\hat{\theta}$  and  $\hat{\sigma}$  are defined as above. The ‘theoretical’ test of exact significance level  $\alpha$  rejects the null hypothesis if  $T > q_{1-\alpha}^0$ , where  $q_{1-\alpha}^0$  is the  $1-\alpha$  quantile of  $T$  under the null hypothesis. The quantity  $\lambda$  of interest in this case is  $q_{1-\alpha}^0$ .

The bootstrap version of the test statistic depends on the type of resampling used to construct the bootstrap samples. If the bootstrap samples are generated by a method that does not impose the null hypothesis, such as the nonparametric bootstrap, then  $T_b^* = n^{1/2}(\hat{\theta}_b^* - \hat{\theta})/\hat{\sigma}_b^*$ . On the other hand, if the bootstrap samples are generated by a method that imposes the null hypothesis, such as the parametric bootstrap based on  $\theta_0=0$ , then  $T_b^* = n^{1/2}(\hat{\theta}_b^* - 0)/\hat{\sigma}_b^*$ . Let  $\hat{q}_{1-\alpha,\infty}$  denote the  $1-\alpha$  quantile of  $T_b^*$ . The ideal bootstrap test of approximate significance level  $\alpha$  rejects the null hypothesis if  $T > \hat{q}_{1-\alpha,\infty}$ . The estimate  $\hat{\lambda}_\infty$  in this case is  $\hat{q}_{1-\alpha,\infty}$ .

Let  $\hat{q}_{1-\alpha,B}$  denote the  $1-\alpha$  quantile of  $\{T_b^*: b=1,\dots,B\}$ . We take  $B$  and  $\alpha$  as in the previous subsection. Thus,  $\hat{q}_{1-\alpha,B}$  equals  $T_{B,v}^*$ , the  $v$ th order statistic of  $\{T_b^*: b=1,\dots,B\}$ . The bootstrap test of approximate significance level  $\alpha$  based on  $B$  bootstrap repetitions rejects the null hypothesis if

$$T > \hat{q}_{1-\alpha,B}. \tag{19}$$

In this case,  $\hat{\lambda}_B$  is  $\hat{q}_{1-\alpha,B}$ .

The quantity  $\omega$  in this example is

$$\omega = \alpha(1-\alpha)/(z_{1-\alpha}^2 \phi^2(z_{1-\alpha})). \tag{20}$$

The estimator  $\hat{\omega}_B$  is the same as in (16), but with  $c_\alpha$  defined by

$$c_\alpha = \left( \frac{1.5z_{1-\alpha/2}^2 \phi^2(z_{1-\alpha})}{2z_{1-\alpha}^2 + 1} \right)^{1/3}. \tag{21}$$

### 2.5. *p*-values

We now consider a testing problem in which one wants to report a *p*-value. In this case, the quantity  $\lambda$  of interest is the exact *p*-value. The null and

alternative hypotheses are as in (17). Let  $T$  and  $T_b^*$  be defined as in the previous subsection. The ideal bootstrap  $p$ -value and the bootstrap  $p$ -value based on  $B$  bootstrap repetitions are

$$\hat{p}_\infty = P^*(T_b^* > T) \quad \text{and} \quad \hat{p}_B = \frac{1}{B} \sum_{b=1}^B 1(T_b^* > T). \tag{22}$$

In this case,  $\hat{\lambda}_\infty = \hat{p}_\infty$  and  $\hat{\lambda}_B = \hat{p}_B$ . We assume that  $\hat{p}_\infty$  does not equal zero or one.

The variance  $\omega$  and its estimate  $\hat{\omega}_B$  are given by

$$\omega = (1 - \hat{p}_\infty) / \hat{p}_\infty \quad \text{and} \quad \hat{\omega}_B = (1 - \hat{p}_B) / \hat{p}_B. \tag{23}$$

### 3. A three-step method of determining $B$

We now specify the three-step method of Andrews and Buchinsky (2000a) for determining  $B$  to achieve a desired accuracy of  $\hat{\lambda}_B$  for estimating  $\hat{\lambda}_\infty$ . Recall that the desired accuracy is specified by a  $(pdb, \tau)$  combination.

#### 3.1. The method

The three-step method depends on a preliminary estimate  $\omega_1$  of the asymptotic variance  $\omega$  of  $B^{1/2}(\hat{\lambda}_B - \hat{\lambda}_\infty) / \hat{\lambda}_\infty$ . For the applications of Section 2, we use the following:

Standard errors:

$$\omega_1 = 1/2,$$

Symmetric two-sided confidence intervals:

$$\omega_1 = \alpha(1 - \alpha) / (4z_{1-\alpha/2}^2 \phi^2(z_{1-\alpha/2})), \tag{24}$$

Tests for a given significance level:

$$\omega_1 = \alpha(1 - \alpha) / (z_{1-\alpha}^2 \phi^2(z_{1-\alpha})),$$

$p$ -values:

$$\omega_1 = \Phi(T) / (1 - \Phi(T)),$$

where  $z_{1-\alpha}$ ,  $\phi(\cdot)$ , and  $\Phi(\cdot)$  denote the  $1 - \alpha$  quantile, density function, and distribution function, respectively, of the standard normal distribution.<sup>5</sup> These specifications of  $\omega_1$  are based on asymptotics, but the three-step method is

---

<sup>5</sup> The last three formulae for  $\omega_1$  in (24) and the corresponding formulae for  $\omega$  given above are suitable only when  $T$  has an absolute standard normal, standard normal, and standard normal asymptotic distribution, respectively, which is the case considered here. Andrews and Buchinsky (2000) give the appropriate formulae for the general case, which includes the common testing case in which  $T$  has an asymptotic chi-squared distribution.

not too sensitive to their choice, because it uses a finite sample estimate of  $\omega$  in the last step.

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

The three-step method is as follows:

*Step 1.* Given  $\omega_1$ , compute

$$B_1 = \text{int}(10,000z_{1-\tau/2}^2\omega_1/pdb^2) \quad (25)$$

or, if  $\hat{\lambda}_B$  is a  $1 - \alpha$  sample quantile, compute  $B_1 = \alpha_2 h_1 - 1$  and  $v_1 = (B_1 + 1)(1 - \alpha)$ , where  $\alpha = \alpha_1/\alpha_2$  and  $h_1 = \text{int}(10,000z_{1-\tau/2}^2\omega_1/(pdb^2\alpha_2))$ .

*Step 2.* Simulate  $B_1$  bootstrap samples  $\{\mathbf{X}_b^*: b = 1, \dots, B_1\}$  and compute an improved estimate  $\hat{\omega}_{B_1}$  of  $\omega$  using the appropriate formulae given in (8), (16), (21), or (23), with  $B$  replaced by  $B_1$ .

*Step 3.* Compute

$$B_2 = \text{int}(10,000z_{1-\tau/2}^2\hat{\omega}_{B_1}/pdb^2) \quad (26)$$

or, if  $\hat{\lambda}_B$  is a  $1 - \alpha$  sample quantile, compute  $B_2 = \alpha_2 h_2 - 1$ , where  $h_2 = \text{int}(10,000z_{1-\tau/2}^2\hat{\omega}_{B_1}/(pdb^2\alpha_2))$ . Take the desired number of bootstrap repetitions to be  $B^* = \max\{B_2, B_1\}$ .

### 3.2. Justification of the three-step method

The justification of the three-step method is that as  $pdb \rightarrow 0$  (and  $n \rightarrow \infty$  when  $\hat{\lambda}_B$  is a sample quantile), we have

$$P^* \left( 100 \frac{|\hat{\lambda}_{B_2} - \hat{\lambda}_\infty|}{\hat{\lambda}_\infty} \leqslant pdb \right) \rightarrow 1 - \tau. \quad (27)$$

Note that  $B_2$  depends on  $pdb$  in (27) via (26) and  $B_2 \rightarrow \infty$  as  $pdb \rightarrow 0$ .

Eq. (27) implies that the three-step method attains precisely the specified accuracy asymptotically using ‘small  $pdb$ ’ asymptotics when  $\omega \geqslant \omega_1$ . If  $\omega < \omega_1$ , then  $B^* = B_1 > B_2$  with probability that goes to one as  $pdb \rightarrow 0$  (and  $n \rightarrow \infty$  when  $\hat{\lambda}_B$  is a sample quantile) and the accuracy of  $\hat{\lambda}_{B^*}$  for approximating  $\hat{\lambda}_\infty$  exceeds that of  $(pdb, \tau)$ . This is a consequence of the fact that it would be silly to throw away the extra  $B_1 - B_2$  bootstrap estimates that have already been calculated in Step 2.

Because one normally specifies a small value of  $pdb$ , the asymptotic result (27) should be indicative of the relevant non-zero  $pdb$  behavior of the three-step method. The simulation results of Section 5 are designed to examine this. We note that the asymptotics used here are completely analogous to large sample size asymptotics with  $pdb$  driving  $B_2$  to infinity as  $pdb \rightarrow 0$  and  $B_2$  playing the role of the sample size.

For more details on the asymptotic justification, see Andrews and Buchinsky (2000).

Table 1  
 Values of *pdb* and  $B_2$  when  $\tau = 0.05$  for standard errors

(A) Values of <i>pdb</i> as a function of $\hat{\gamma}_{2B}$ and $B$										
$\hat{\gamma}_{2B}$	$B$									
	10	25	50	100	200	350	500	750	1000	2000
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

  

(B) Values of $B_2$ as a function of $\hat{\gamma}_{2B_1}$ and <i>pdb</i>			
$\hat{\gamma}_{2B_1}$	<i>pdb</i>		
	20	10	5
0	48	192	768
1	72	288	1152
2	96	384	1536
3	120	480	1920

### 3.3. Evaluation of the three-step method

#### 3.3.1. Standard errors

Table 1(A) provides the values of *pdb* that correspond to an array of values of  $\hat{\gamma}_{2B}$  and  $B$  when  $\tau = 0.05$ . For example, if  $\hat{\gamma}_{2B} = 0$  (which corresponds to the kurtosis of the normal distribution) and  $B = 50$ , then  $pdb \approx 20$ . That is, with probability approximately 0.95,  $\hat{s}e_B$  is within  $\pm 20\%$  of  $\hat{s}e_\infty$ . Or, with probability approximately 0.95,  $\hat{s}e_\infty$  is within  $\pm 20\%$  of  $\hat{s}e_B$ .

Table 1(A) shows that to obtain very accurate estimates of  $\hat{s}e_\infty$ , say  $pdb = 5$ , one needs quite large values of  $B$ , e.g.,  $B = 750$  when  $\hat{\gamma}_{2B} = 0$  and  $B = 2000$  when  $\hat{\gamma}_{2B} = 3$ . Much smaller values of  $B$  are required to obtain moderate accuracy, say  $pdb = 20$ , e.g.,  $B = 50$  when  $\hat{\gamma}_{2B} = 0$  and  $B = 100$  when  $\hat{\gamma}_{2B} = 2$ .

For illustrative purposes, Table 1(B) provides values of  $B_2$  that correspond to several values of  $\hat{\gamma}_{2B_1}$  and *pdb*, with  $\tau = 0.05$ . The values of *pdb* considered are 20 (moderately accurate), 10 (accurate), and 5 (very accurate). Table 1(B) indicates that the necessary  $B_2$  values increase very quickly as the desired level of accuracy increases.

#### 3.3.2. Symmetric two-sided confidence intervals

In Table 2(A), we provide the values of  $B_1$  that correspond to common values of  $\alpha$ ,  $\tau$ , and *pdb*. Table 2(A) indicates that  $B_1$  increases significantly as  $\tau$  decreases and even more so as *pdb* decreases. For example, the combination  $(\alpha, pdb, \tau) = (0.05, 15, 0.10)$  requires  $B_1 = 119$ . In contrast,

Table 2  
Values of  $B_1$  as a function of  $\alpha$ ,  $\tau$ , and  $pdb^a$

(A) Symmetric confidence intervals											
$\alpha$	0.01			0.05			0.10				
$\tau$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10		
$pdb$											
5	4799	2799	1999	2399	1399	979	2079	1209	849		
10	1199	699	499	599	359	259	519	309	219		
15	599	399	299	279	159	119	239	139	99		

  

(B) Equal-tailed and one-sided confidence intervals												
$\alpha$	0.01			0.025			0.05			0.10		
$\tau$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
$pdb$												
5	6899	3999	2899	4959	2879	2039	4379	2539	1799	4729	2739	1939
10	1799	999	799	1239	719	519	1099	639	459	1189	689	489
15	799	499	399	559	319	239	499	299	199	529	309	219

<sup>a</sup>Note: All quantities are defined in the three-step procedure of Section 3.

$(\alpha, pdb, \tau) = (0.05, 5, 0.01)$  requires  $B_1 = 2399$ . In addition,  $B_1$  increases as  $\alpha$  decreases.

### 3.3.3. One-sided and equal-tailed two-sided confidence intervals

The three-step method of Andrews and Buchinsky (2000) applies to equal-tailed and one-sided confidence intervals, as well as to symmetric two-sided confidence intervals. For comparative purposes, Table 2(B) provides the values of  $B_1$  for equal-tailed and one-sided confidence intervals that correspond to different  $(\alpha, pdb, \tau)$  combinations. The variation in the values of  $B_1$  in Table 2(B), as  $(\alpha, pdb, \tau)$  varies, is very similar to that in Table 2(A) for symmetric confidence intervals, as is expected from the formulae of Andrews and Buchinsky (2000).

Tables 2(A) and (B) indicate that the  $B_1$  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  $B_1$  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  $\tau$  (except for rounding effects from the  $\text{int}(\cdot)$  function). For equal-tailed confidence intervals, this ratio is 2.0 and 2.1 for confidence levels 0.95 and 0.90, respectively. For one-sided confidence intervals, this ratio is 1.8 and 2.3 for confidence levels 0.95 and 0.90, respectively. The reason fewer repetitions are needed for symmetric confidence intervals is that the asymptotic density

Table 3

Values of  $B_1$  for tests with significance level  $\alpha$ , as a function of  $\alpha$  and  $\tau$ , for  $pdb = 10$ , under alternative asymptotic null distributions<sup>a</sup>

$\alpha$	0.01			0.05			0.10		
	$\tau$	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05
A. Absolute N(0, 1):	1199	699	499	599	359	259	519	309	219
B. N(0, 1):	1799	999	799	1099	639	459	1189	689	489
C. $\chi^2_5$ :	1799	999	699	699	419	299	519	299	219
D. $\chi^2_{15}$ :	799	499	399	279	179	119	199	119	89

<sup>a</sup>Note: All quantities are defined in the three-step method of Section 3.

of  $|n^{1/2}(\hat{\theta}_b^* - \hat{\theta})/\hat{\sigma}_b^*|$  is twice as large as that of  $n^{1/2}(\hat{\theta}_b - \hat{\theta})/\hat{\sigma}_b$  at any positive value.

### 3.3.4. Tests with given significance level $\alpha$

To assess the computational burden of the three-step procedure for tests with specified significance levels, Table 3 provides values for  $B_1$  for a variety of  $(\alpha, \tau)$  combinations when  $pdb = 10$  and the asymptotic null distribution of the test statistic is absolute N(0, 1), N(0, 1),  $\chi^2_5$ , and  $\chi^2_{15}$ , where  $\chi^2_d$  denotes a chi-squared distribution with  $d$  degrees of freedom. Formulae for each of these cases are provided in Andrews and Buchinsky (2000). The formulae given in Sections 2.4 and 3.1 correspond to the N(0, 1) case.

Table 3 shows that for tests with absolute N(0, 1) asymptotic null distribution the same number of initial bootstrap repetitions  $B_1$  are needed as for symmetric confidence intervals. For tests with N(0, 1) asymptotic null distribution, noticeably larger  $B_1$  values are required—the ratio of  $B_1$  values for N(0, 1) to absolute N(0, 1) tests is between 1.8 and 2.3, for  $\alpha = 0.05$  or 0.10. For tests with  $\chi^2_5$  asymptotic null distribution, similar  $B_1$  values are required as for absolute N(0, 1) tests—the ratio of  $B_1$  values for  $\chi^2_5$  to absolute N(0, 1) tests is in the range 1.0–1.2 for  $\alpha = 0.05$  or 0.10. For tests with  $\chi^2_{15}$  asymptotic null distribution, noticeably smaller  $B_1$  values are required than for absolute N(0, 1) tests—the ratio of  $B_1$  values for  $\chi^2_{15}$  to absolute N(0, 1) tests is in the range 0.38–0.50 for  $\alpha = 0.05$  or 0.10. Thus, there is considerable variation in suitable values of  $B_1$  for test statistics with different asymptotic null distributions.

In all cases,  $B_1$  increases quickly as  $\alpha$  or  $\tau$  decreases. It is also true that  $B_1$  increases very quickly as  $pdb$  decreases, but Table 3 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.



Table 4  
 Values of  $B_1$  as a function of  $\tau$ ,  $pd_b$ , and  $1 - \Phi(T)$  for  $p$ -values<sup>a</sup>

$1 - \Phi(T)$	$pd_b$										
	5	10	15	20	30	40	50	100	150	200	300
$\tau = 0.025$ :											
0.001								5019	2231	1255	558
0.005						6248	3999	1000	444	250	111
0.010					5526	3109	1989	497	221	124	
0.025			8708	4898	2177	1225	784	196	87	49	
0.050	38,182	9545	4242	2386	1061	597	382	95	42		
0.10	18,086	4521	2010	1130	502	283	181	45	20		
0.15	11,387	2847	1265	712	316	178	114	28			
0.20	8038	2010	893	502	223	126	80	20			
0.30		1172	521	293	130	73	47				
0.50			223	126	56	31	20				
0.70			96	54	24	13					
$\tau = 0.05$ :											
0.001								3838	1706	959	426
0.005						4778	3058	764	340	191	85
0.010				9508	4226	2377	1521	380	169	95	
0.025			6659	3745	1665	936	599	150	67	37	
0.050	29,195	7299	3244	1825	811	456	292	73	32		
0.10	13,829	3457	1537	864	384	216	138	35	15		
0.15	8707	2177	967	544	242	136	87	22			
0.20	6146	1537	683	384	171	96	61	15			
0.30		896	398	224	100	56	36				
0.50			171	96	43	24	15				
0.70			73	41	18	10					
$\tau = 0.10$ :											
0.001								2703	1201	676	300
0.005					5982	3365	2154	538	239	135	60
0.010				6696	2976	1674	1071	268	119	67	
0.025			4690	2638	1172	659	422	106	47	26	
0.050	20,562	5141	2285	1285	571	321	206	51	23		
0.10	9740	2435	1082	609	271	152	97	24	11		
0.15	6133	1533	681	383	170	96	61	15			
0.20	4329	1082	481	271	120	68	43	11			
0.30		631	281	158	70	39	25				
0.50			120	68	30	17	11				
0.70			52	29	13	7					

<sup>a</sup>Note: All quantities are defined in the three-step method of Section 3.

### 3.3.5. $p$ -Values

Table 4 provides representative values of  $B_1$  for the three-step method for  $p$ -values. Three different values of  $\tau$  are considered, viz., 0.025, 0.05,

and 0.10. A range of values of the initial  $p$ -value estimate  $1 - \Phi(T)$  and the accuracy bound  $pdb$  are considered. For clarity, Table 4 only provides  $(1 - \Phi(T), pdb)$  combinations that are of some interest. For example, it is not of interest to consider the combination (0.001, 5), because this combination yields excessive accuracy and, hence, requires an excessively large value of  $B_1$ .

Table 4 indicates that the required magnitude of  $B_1$  depends on the initial  $p$ -value estimate  $1 - \Phi(T)$ . If it is quite small or large, then one does not need a small value of  $pdb$  and the required magnitude of  $B_1$  is not large. On the other hand, if  $1 - \Phi(T)$  is in an intermediate range, such as (0.01, 0.15), then one may want to employ a relatively small value of  $pdb$  and the required magnitude of  $B_1$  may be quite large.

#### 4. Monte Carlo simulation design

In this section, we introduce the design of our simulation experiments. As indicated above, the main purpose of the experiments is to see whether the asymptotic justification of the three-step method in (27) is indicative of finite sample behavior for a range of values of  $(pdb, \tau)$  in several standard econometric models. More specifically, given several  $(pdb, \tau)$  combinations, we want to see how close  $P^*(100|\hat{\lambda}_{B_2} - \hat{\lambda}_{\infty}|/\hat{\lambda}_{\infty} \leq pdb)$  is to  $1 - \tau$ . We focus on  $B_2$  rather than  $B^*$  because the limit in (27) is exactly  $1 - \tau$ , whereas the limit in (27) when  $\hat{\lambda}_{B_2}$  is replaced by  $\hat{\lambda}_{B^*}$  is greater than or equal to  $1 - \tau$ . Of course, our interest ultimately is in the performance of  $B^*$ .

We consider three different models: (i) linear regression; (ii) binary probit; and (iii) quantile regression. For each of these three models, we evaluate the performance of the three-step method for choosing  $B$  for: (i) standard error estimates; (ii) symmetric confidence intervals; (iii) tests for a given significance level  $\alpha$ ; and (iv)  $p$ -values. For the confidence intervals and tests, we consider two values of  $\alpha$ , viz., 0.05 and 0.10.

##### 4.1. Models

###### 4.1.1. Linear regression model

The linear regression model is

$$y_i = x_i' \beta + u_i \quad (28)$$

for  $i = 1, \dots, n$ , where  $n = 25$ ,  $X_i = (y_i, x_i')$  are iid over  $i = 1, \dots, n$ ,  $x_i = (1, x_{1i}, \dots, x_{5i})' \in R^6$ ,  $(x_{1i}, \dots, x_{5i})$  are mutually independent normal random variables,  $x_i$  is independent of  $u_i$ , and  $Eu_i = 0$ . The simulation results for this model are invariant with respect to the means and variances of the regressor vector  $x_i$ , the variance of the error term  $u_i$ , and the value of the

regression parameter  $\beta$ , so we need not be specific as to their values. We consider three error distributions:  $N(0, 1)$ ,  $t_5$ , and  $\chi_5^2$  shifted to have mean zero.

We estimate  $\beta$  by least squares (LS). We focus attention on the bootstrap quantities for the LS estimator of the first slope coefficient. Thus, the parameter  $\theta$  of the previous sections is  $\beta_2$ , the second element of  $\beta$ . Our estimate  $\hat{\sigma}(\mathbf{X})$  of the asymptotic standard error of  $\hat{\theta}$  ( $=\hat{\beta}_2$ ), used to construct the  $t$  statistics, is given by the square root of the (2,2) element of  $\hat{\sigma}_u^2(\sum_{i=1}^n x_i x_i' / n)^{-1}$ , where  $\hat{\sigma}_u^2 = e'e / (n - 6)$  and  $e$  is the  $n$ -vector of LS residuals.

The LS estimator of  $\theta$  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  $\theta$  is zero. However, the crucial parameter for standard error estimates is the coefficient of excess kurtosis,  $\gamma_2$ , of the discrete *bootstrap* distribution of the LS estimator of  $\theta$ . In general, the parameter  $\gamma_2$  depends on the sample and need not equal zero. Nevertheless, the value of  $\gamma_2$  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  $\gamma_2$  will tend to be large for most samples.

To obtain samples for which  $\gamma_2$  is close to zero, we consider normal errors. To obtain samples with larger values of  $\gamma_2$ , we consider the fat-tailed error distributions  $t_5$  and  $\chi_5^2$ . The  $t_5$  and  $\chi_5^2$  distributions have similar tail behavior and generate samples with similar values of  $\gamma_2$ . The  $t_5$  and normal distributions are symmetric, whereas the  $\chi_5^2$  distribution is highly skewed. The results for the  $\chi_5^2$  error distribution are used to determine whether skewness of the error distribution has an impact in finite samples on the performance of the three-step method for determining  $B$ .

#### 4.1.2. Binary probit model

The binary probit model is

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad y_i^* = x_i' \beta + u_i \quad (29)$$

for  $i = 1, \dots, n$ , where  $n = 25$ ,  $X_i = (y_i, x_i')'$  are iid over  $i = 1, \dots, n$ ,  $x_i = (1, x_{1i}, \dots, x_{5i})' \in R^6$ ,  $(x_{1i}, \dots, x_{5i})$  are mutually independent standard normal random variables,  $x_i$  is independent of  $u_i$ ,  $u_i \sim N(0, 1)$ , and  $\beta = (1, 1.5, 1, 0.75, 0.5)'$ .

We estimate  $\beta$  by maximum likelihood (ML) using the Nelder–Meade algorithm. We focus attention on the first slope coefficient, i.e.,  $\theta = \beta_2$ . Our estimate  $\hat{\sigma} = \hat{\sigma}(\mathbf{X})$  of the asymptotic standard error of  $\hat{\theta}$  ( $=\hat{\beta}_2$ ), used to construct the  $t$  statistics, is the (2,2) element of the inverse of the sample information matrix calculated from the second derivative of the likelihood function.

4.1.3. *Quantile regression model*

The quantile regression model is

$$\begin{aligned} y_i &= x_i' \beta + (x_i' \gamma) u_i \\ &= x_i' \delta_q + (x_i' \gamma) u_{qi} \end{aligned} \tag{30}$$

for  $i = 1, \dots, n$ , where

$$\delta_q = \beta + \gamma \xi_q,$$

$u_{qi} = u_i - \xi_q$ ,  $\xi_q$  denotes the  $q$  quantile ( $0 < q < 1$ ) of  $u_i$ ,  $n = 25$ ,  $X_i = (y_i, x_i')$  are iid over  $i = 1, \dots, n$ ,  $x_i = (1, x_{1i}, \dots, x_{5i})' \in R^6$ ,  $(x_{1i}, \dots, x_{5i})$  are mutually independent standard normal random variables,  $x_i$  is independent of  $u_i$ ,  $\text{Median}(u_i) = 0$ , and  $\gamma = (1, 0.75, 0.5, 0.375, 0.25, 0.125)'$ .<sup>6</sup> The conditional  $q$  quantile of  $y_i$  given  $x_i$  is  $x_i' \delta_q$ . The parameter vector of interest is  $\delta_q$ . The simulation results are invariant with respect to  $\beta$ . As in the linear regression model, 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 median zero (again denoted  $\chi_5^2$ ).

We estimate  $\delta_q$  using the Koenker and Bassett (1978) quantile regression (QR) estimator implemented using the Barrodale and Roberts (1973) linear programming algorithm. We focus attention on the first slope coefficient, i.e.,  $\theta = \delta_{q2}$ . We consider two alternative quantiles:  $q = 0.50$  and  $0.75$ . The results do not differ substantially between these two quantiles, so we only report results for  $q = 0.75$ . Our estimator  $\hat{\sigma} = \hat{\sigma}(\mathbf{X})$  of the asymptotic standard error of  $\hat{\theta}$  ( $= \hat{\delta}_{q2}$ ), used to construct the  $t$  statistics, is the square root of the (2, 2) element of the kernel estimator of the asymptotic variance of the quantile regression estimator defined in Buchinsky (1995). The asymptotic variance estimator is defined by  $\hat{\Delta}_{fxx}^{-1} \hat{\Delta}_{xx} \hat{\Delta}_{fxx}^{-1}$ , where

$$\hat{\Delta}_{xx} = \sum_{i=1}^n x_i x_i' / n, \quad \hat{\Delta}_{fxx} = d_n^{-1} \sum_{i=1}^n \exp\{-\hat{u}_{qi}^2 / (2c_n^2)\} 1(\hat{u}_{qi} \geq 0) x_i x_i' / n, \tag{31}$$

$\hat{u}_{qi}$  ( $i = 1, \dots, n$ ) are the residuals from the  $q$ th quantile regression,  $d_n = nc_n \sqrt{2\pi} / 2$ , and  $c_n$  is the kernel bandwidth. The kernel bandwidth was chosen using the least-squares cross-validation method.

4.2. *Design of the experiments*

For each of the above models, we simulate 100 different samples from each error distribution. The reason for considering 100 different samples is that the empirical distribution  $\hat{F}$ , the distributions of  $\hat{\theta}_b^*$  and  $T_b^*$ , and all the

<sup>6</sup> This model was introduced by Koenker and Bassett (1982). For details see, for example, Buchinsky (1995).

quantities  $\gamma_2$ ,  $\widehat{se}_\infty$ ,  $\widehat{q}_{1-\alpha,\infty}$ , and  $\widehat{p}_\infty$ , vary with the sample. For each of the 100 samples drawn (for a given distribution of  $u_i$ ), we compute the estimates  $(\widehat{\theta}, \widehat{\sigma})$  and simulate the ideal bootstrap estimate  $\widehat{\lambda}_\infty$  for each of the first three applications (viz., standard error, confidence interval, and test for given  $\alpha$ ). We simulate  $\widehat{\lambda}_\infty$  using 250,000 bootstrap repetitions (each of sample size 25). Let  $\widehat{\lambda}_{250,000}$  denote the simulated value. For the confidence interval and test applications, we compute  $\widehat{\lambda}_{250,000}$  for two different values of  $\alpha$ , viz., 0.05 and 0.10.

For each experiment, we provide an estimate of the coefficient of variation (CV) of  $\widehat{\lambda}_{250,000}$ , i.e., the ratio of an estimate of the standard error of  $\widehat{\lambda}_{250,000}$  to the value  $\widehat{\lambda}_{250,000}$  itself:

$$CV = \widehat{se}(\widehat{\lambda}_{250,000}) / \widehat{\lambda}_{250,000}. \quad (32)$$

By (3), the asymptotic variance of  $\widehat{\lambda}_B$  as  $B \rightarrow \infty$  is  $\omega \widehat{\lambda}_\infty^2$ . Hence, an estimate of the standard error of  $\widehat{\lambda}_{250,000}$  divided by  $\widehat{\lambda}_{250,000}$  is  $CV = (\widehat{\omega}_{250,000} / 250,000)^{1/2}$ , where  $\widehat{\omega}_B$  is defined in (8), (16), (21), and (23) for each of the applications. The CV values for each case considered are reported in the notes to the tables given below. They range from 0.0017 to 0.0092. This indicates that 250,000 is close enough to infinity to accurately estimate  $\widehat{\lambda}_\infty$ .

We run 2500 Monte Carlo repetitions for each of the 100 samples, for a total of 250,000 repetitions. In each Monte Carlo repetition, we compute  $\widehat{\lambda}_{B_2}$  and  $\widehat{\lambda}_{B^*}$  for each of the four applications (and for each of the two values of  $\alpha$  for the confidence interval and test for given  $\alpha$  applications) following the three-step procedure outlined in Section 3. The calculations are made for several combinations of  $pdb$  (viz., 20%, 10%, and 5%) and  $1 - \tau$  (viz., 0.90, 0.95, and 0.975). For each repetition and each  $(pdb, \tau)$  combination, we determine whether or not the estimate  $\widehat{\lambda}_{B_2}$  satisfies

$$100 \frac{|\widehat{\lambda}_{B_2} - \widehat{\lambda}_{250,000}|}{\widehat{\lambda}_{250,000}} < pdb. \quad (33)$$

We call the fraction of times that this condition is satisfied, out of the 2500 repetitions, the *empirical level* based on  $B_2$  bootstrap repetitions. The empirical level based on  $B_1$  and  $B^*$  bootstrap repetitions are defined analogously. To assess the effect of estimation of the asymptotic variance  $\omega$  on the results, we also consider the empirical level based on  $B_\omega$  bootstrap repetitions. Here,  $B_\omega$  is the number of bootstrap repetitions that is appropriate given the true value of  $\omega$ . That is,  $B_\omega$  is defined as  $B_1$  is defined in (25), but with the preliminary estimate  $\omega_1$  of  $\omega$  replaced by the true value of  $\omega$ . (Of course, in practice  $\omega$  is unknown, so that  $B_\omega$  is not a feasible choice for  $B$ .) For each  $(pdb, \tau)$  combination, each sample, each application, each model, and each error distribution, we compute the empirical levels based on  $B_1$ ,  $B_2$ ,  $B^*$ , and  $B_\omega$  bootstrap repetitions.

The three-step method of Section 3 performs well if the empirical level based on  $B_2$  bootstrap repetitions is close to  $1 - \tau$ , or if the empirical level based on  $B^*$  bootstrap repetitions is close to, or greater than,  $1 - \tau$ .

For  $p$ -values, we are interested in the behavior of the three-step method for a number of different samples that all yield the same ideal bootstrap  $p$  and  $\hat{p}_\infty = 0.10$ . The acceptance–rejection method of choosing samples that yield the same ideal bootstrap  $p$ -value is too computationally intensive to be feasible. Instead, we use a method that considers a different null hypothesis for each sample, which is not ideal, but is computationally tractable. We follow the same procedure as described above, except that we consider a test of  $H_0 : \theta = \tilde{\theta}$  versus  $H_1 : \theta > \tilde{\theta}$ , where  $\tilde{\theta}$  is chosen such that our estimate of  $\hat{p}_\infty$  equals the desired value, say 0.05. Specifically, we compute the bootstrap  $t$  statistics  $T_b^* = (\hat{\theta}_b^* - \tilde{\theta})/\hat{\sigma}_b^*$  for  $b = 1, \dots, 250,000$  and take  $\tilde{\theta}$  such that  $T = (\hat{\theta} - \tilde{\theta})/\hat{\sigma}$  is the 0.95 sample quantile of the 250,000 values of  $T_b^*$ . The empirical levels for the  $p$ -value results could be calculated by the method described in (33). It is much quicker, however, to calculate the empirical levels by using the fact that  $\sum_{b=1}^B (T_b^* > T)$  has a binomial distribution with parameters  $\hat{p}_\infty$  and  $B$ .<sup>7</sup> This is what we do. We repeat all the calculations for the case  $\hat{p}_\infty = 0.10$ .

For the confidence interval application, we also compute the fraction of times over all Monte Carlo simulations and all samples that  $\theta$  falls within the constructed confidence interval based on  $B^*$ . We call this fraction the *empirical unconditional coverage probability*. The empirical unconditional coverage probability based on  $B_2$  bootstrap repetitions is defined analogously. To improve the precision of the empirical unconditional coverage probability, we actually use 250 samples, rather than 100, for the confidence interval application with the linear regression model. (For the QR and probit models, only 100 samples are used because of computational costs.)

## 5. Monte Carlo simulation results

### 5.1. Standard errors

#### 5.1.1. Linear regression model

The results for standard error estimates in the linear regression model are reported in Table 5 for the  $N(0, 1)$  and  $t_5$  error distributions. The numbers reported in Table 5 and all the remaining tables are averages over the 100 samples. (For example, Med is the average median over the 100 samples.) Results for the  $\chi_5^2$  error distribution are almost the same as those for the  $t_5$ . In

<sup>7</sup>This was suggested by an anonymous referee.

consequence, we do not report the  $\chi_5^2$  results. They indicate that asymmetry of the errors is not an important factor for the performance of the three-step method.

Table 5(A) shows that the empirical levels for  $B_2$  and  $B^*$  are very close to the desired levels (given by  $1 - \tau$ ) 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 desired levels. The reason is that the asymptotic approximation improves as  $B_2$  and  $B^*$  increase. Smaller *pdb* values and/or larger  $1 - \tau$  values lead to larger  $B_2$  and  $B^*$  values and, hence, better performance.

In contrast, the empirical levels based on  $B_1$  bootstrap repetitions are noticeably less than the desired levels. This indicates that the use of three steps, in which  $\omega$  is estimated in the first two steps, results in much better performance than that obtained by simply using the first-step value  $B_1$ . This holds true for all of the cases reported in this paper, because the empirical levels based on  $B_1$  bootstrap repetitions are always noticeably less than the desired levels. For brevity, given this clear pattern, we do not discuss the empirical levels for  $B_1$  for any of the other cases considered in the paper, although they are reported in the tables.

The empirical levels based on  $B_\omega$  repetitions are much closer to the desired levels than those based on  $B_1$  repetitions, but they are not as close as those based on  $B_2$  or  $B^*$ . This pattern holds true for all of the cases reported in this paper (although for the confidence interval results the empirical levels for  $B_2$ ,  $B^*$ , and  $B_\omega$  are all slightly greater than the desired values, rather than less). The superior performance of  $B_2$  and  $B^*$  over  $B_\omega$  is due to a better approximation of  $B^{1/2}(\hat{\lambda}_B - \hat{\lambda}_\infty)/\hat{\lambda}_\infty$  by a normal distribution with an estimate of its finite sample variance than by a normal with the true asymptotic variance. These results also indicate the advantage of the three-step method in which unknown quantities are replaced by estimated values. Again, for brevity and given the clear pattern of the results, we do not discuss the empirical levels for  $B_\omega$  for any of the other cases considered in the paper, although they are reported in the tables.

The average of the  $\gamma_2$  values over the 100 samples used in Table 5(A) (computed using 250,000 bootstrap repetitions for each sample) is 0.37. The mean (over 2500 simulation repetitions) of the estimator  $\hat{\gamma}_{2B_1}$  averaged over the 100 samples, as reported in Table 5(A), is markedly lower than 0.37 when  $B_1$  is small (or equivalently, when *pdb* is large). This downward bias of  $\hat{\gamma}_{2B_1}$  leads to  $B_2$  and  $B^*$  values that are smaller than desired. In turn, this leads to empirical levels based on  $B_2$  and  $B^*$  bootstrap repetitions that are

Table 5  
Standard errors using the three-step method with no bias correction—linear regression model<sup>a</sup>

pdb	1 - τ			Empirical level			$\hat{\gamma}_{2B_1}$							
	$B_1$	$B_2$	$B^*$	$B_o$	$B_1$	$B_o$	Mean	Med	Min	Max	Mean	Med	Min	Max
(A) Error distribution $N(0, 1)$														
20	0.900	0.831	0.872	0.861	34	41	38	34	11	225	0.14	-0.07	-1.35	11.29
10	0.900	0.836	0.891	0.897	136	162	161	150	74	1277	0.33	0.19	-0.84	16.99
5	0.900	0.842	0.894	0.896	542	643	639	620	411	4463	0.39	0.27	-0.43	14.85
20	0.950	0.889	0.942	0.944	48	57	56	50	18	390	0.23	0.04	-1.29	14.67
10	0.950	0.899	0.945	0.948	192	228	222	210	118	1835	0.36	0.23	-0.71	17.21
5	0.950	0.908	0.950	0.951	768	911	922	889	621	6948	0.38	0.31	-0.32	15.44
20	0.975	0.914	0.970	0.973	63	75	69	63	28	565	0.27	0.09	-1.19	15.94
10	0.975	0.922	0.972	0.974	251	298	289	278	153	2481	0.37	0.25	-0.66	17.09
5	0.975	0.931	0.974	0.975	1004	1190	1211	1151	832	11,002	0.41	0.33	-0.35	18.22
(B) Error distribution $t_5$														
20	0.900	0.825	0.849	0.860	34	41	47	40	12	280	0.66	0.31	-1.29	14.74
10	0.900	0.830	0.873	0.877	136	162	211	191	86	1591	1.06	0.79	-0.73	22.06
5	0.900	0.838	0.881	0.890	542	643	870	824	514	6915	1.22	1.07	-0.11	22.99
20	0.950	0.880	0.913	0.918	48	57	69	57	20	460	0.78	0.41	-1.15	18.79
10	0.950	0.890	0.931	0.935	192	228	309	274	136	2256	1.13	0.86	-0.55	22.22
5	0.950	0.899	0.940	0.943	768	911	1264	1201	768	9577	1.25	1.08	0.05	23.52
20	0.975	0.905	0.951	0.957	63	75	90	79	33	662	0.89	0.55	-1.05	19.09
10	0.975	0.916	0.964	0.966	251	298	402	372	192	2755	1.14	0.94	-0.46	19.80
5	0.975	0.921	0.970	0.972	1004	1190	1633	1559	1038	11,283	1.26	1.11	0.07	20.91

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample, we carry out 2500 Monte Carlo repetitions.  $\hat{\gamma}_2$  is calculated for each of the 100 samples using 250,000 bootstrap repetitions. The average of the 100  $\hat{\gamma}_2$  values is 0.37 in part (A) and 1.26 in part (B). The average of the  $CV$  values is 0.0017 in panel (A) and 0.0021 in panel (B).



less than  $1 - \tau$  when  $B_1$  is small. For larger values of  $B_1$  (which occur with smaller  $pdb$  values), this bias vanishes and the empirical levels are closer to  $1 - \tau$ . The problem of underestimating  $\gamma_2$  stems from the fact that neither the numerator nor one over the denominator of the estimator  $\hat{\gamma}_{2B_1}$  in (8) is an unbiased estimator of its population counterpart, although both are consistent estimators.

Note that there is significant variation in the values of  $B_2$  over the various  $(pdb, \tau)$  combinations in Table 5(A). The mean values of  $B_2$  are between 38 and 1211. The corresponding values for  $B^*$  are very similar, because  $\gamma_2 > 0$  for all 100 samples,  $\hat{\gamma}_{2B_1}$  is positive or close to zero for the vast majority of repetitions, and  $B^* = B_2$  whenever  $\hat{\gamma}_{2B_1} > 0$ . 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 bound (e.g.,  $pdb=5\%$ ) and a very high probability (e.g.,  $1 - \tau = 0.975$ ), then the number of bootstrap repetitions needed to achieve this level of accuracy is quite large.

Table 5(B) presents the results based on  $t_5$  errors. The average value of  $\gamma_2$  over the 100 samples with  $t_5$  errors is 1.26, which is noticeably larger than the value of 0.37 for normal errors. In Table 5(B), the empirical levels based on  $B_2$  and  $B^*$  bootstrap repetitions are lower than  $1 - \tau$  and lower than their values in Table 5(A). Nevertheless, the same basic pattern is observed as in Table 5(A). That is, the difference between the empirical levels and  $1 - \tau$  are largest when  $B_1$  is small, which corresponds to  $pdb$  being large. When  $B_1$  is small,  $\hat{\gamma}_{2B_1}$  is markedly downward biased and its bias is greater in magnitude than in Table 5(A). This causes  $B_2$  and  $B^*$  to be smaller than desired by a greater magnitude than in Table 5(A).

Overall, the empirical level results of Table 5(B) are not as good as those of Table 5(A). Nevertheless, the three-step method still performs quite well with  $t_5$  errors. The largest deviation of an empirical level based on  $B_2$  repetitions from its asymptotic counterpart is 0.051 and for all other  $(pdb, \tau)$  combinations the deviations are less than half as large. Furthermore, the deviations based on  $B^*$  are smaller than those based on  $B_2$ .

Table 6 reports simulation results for the bias-corrected three-step method for the linear regression model with  $t_5$  errors. That is, the results of Table 6 are analogous to those of Table 5(B) except that the bootstrap bias-corrected estimator  $\hat{\gamma}_{2B_1, R}$  of  $\gamma_2$ , defined in (9), is used in the definition of  $\hat{\omega}_{B_1}$ . We only report results for the  $t_5$  errors because they yield the worst results of the three error distributions considered. 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 trade-off between computational time for our simulation experiment and accuracy of the bootstrap bias-corrected estimator.

Table 6  
Standard errors using the three-step method with bias correction—linear regression model with error distribution  $t_5^a$

$pdb$	$1 - \tau$	Empirical level			$B_2$										
		$B_1$	$B_2$	$B^*$	$B_{\omega}$	$B_1$	$B_{\omega}$	Mean	Med	Min	Max	Mean	Med	Min	Max
20	0.900	0.831	0.876	0.878	0.865	34	55	54	48	12	353	1.04	0.73	-1.33	17.66
10	0.900	0.836	0.889	0.894	0.881	136	222	219	207	82	1721	1.20	0.91	-0.85	22.54
5	0.900	0.842	0.896	0.899	0.894	542	884	902	852	501	7311	1.21	0.95	-0.27	24.22
20	0.950	0.889	0.938	0.942	0.924	48	79	78	74	21	517	1.09	0.80	-1.21	18.52
10	0.950	0.899	0.944	0.947	0.932	192	313	314	287	134	2444	1.20	0.91	-0.66	22.64
5	0.950	0.908	0.948	0.950	0.942	768	1251	1295	1188	747	10,148	1.23	0.97	-0.19	23.47
20	0.975	0.914	0.969	0.972	0.947	63	103	104	94	33	701	1.11	0.81	-1.10	18.50
10	0.975	0.922	0.971	0.974	0.952	251	410	419	397	190	3323	1.22	0.97	-0.56	23.52
5	0.975	0.931	0.973	0.975	0.962	1004	1637	1622	1577	1010	14,112	1.26	1.04	-0.08	24.12

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample we carry out 2500 Monte Carlo repetitions.  $\gamma_2$  is calculated for each of the 100 samples using 250,000 bootstrap repetitions. The average of the 100  $\gamma_2$  values is 1.26. The value of  $R$  for the bias-corrected method is 407. The average of the  $CV$  values is 0.0021, the same as in panel (B) of Table 5.

The results of Table 6 show a significant improvement in the performance of the three-step method when it is augmented by the bias-correction of  $\hat{\gamma}_{2B_1}$ . Most of the empirical levels for  $B_2$  and  $B^*$  are very close to the desired levels. The largest deviation is 0.024 and most deviations are less than 0.010. The results in the ‘Mean’ column for  $\hat{\gamma}_{2B_1,R}$  indicate that the bias of the bias-corrected estimator  $\hat{\gamma}_{2B_1,R}$  is much smaller than that of  $\hat{\gamma}_{2B_1}$  in Table 5(B). The ‘Mean’ numbers of bootstrap repetitions  $B_2$  are larger than in Table 5(B) due to the bias-correction, but the increase is not substantial. Analogous results hold for  $B^*$ .

### 5.1.2. Binary probit model

The results for standard error estimates in the binary probit model are reported in Table 7(A) for the  $N(0,1)$  error distribution. Table 7(A) shows that the empirical levels are very close to the desired levels. Again, this is true even though the bootstrap distribution of the probit estimate with only 25 observations in the sample can be far away from its asymptotic normal distribution. Note that even though the average of the  $\gamma_2$  values over the 100 samples used in Table 7(A) is only 0.071, 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 desired levels. As above, smaller *pdb* values and/or larger  $1 - \tau$  values lead to larger  $B_2$  and  $B^*$  values and, hence, better performance.

The mean (over 2500 simulation repetitions) of the estimator  $\hat{\gamma}_{2B_1}$  averaged over the 100 samples, as reported in Table 7(A), is markedly lower than 0.071 when  $B_1$  is small (or equivalently, when *pdb* is large). This downward bias of  $\hat{\gamma}_{2B_1}$  leads to  $B_2$  and  $B^*$  values that are smaller than desired. In turn, this leads to empirical levels based on  $B_2$  and  $B^*$  bootstrap repetitions that are farther away from  $1 - \tau$  when  $B_1$  is small. As in the linear regression model, when  $B_1$  is larger, this bias vanishes and the empirical levels are closer to  $1 - \tau$ . Also, note that there is significant variation in the values of  $B_2$  over the various (*pdb*,  $\tau$ ) combinations.

Overall, the empirical level results of Table 7(A) are quite close to the desired levels. The largest deviation of an empirical level based on  $B_2$  repetitions from the desired level is 0.015 and for all other (*pdb*,  $\tau$ ) combinations the deviations are much smaller.

Table 7(B) reports simulation results for the bias-corrected three-step method for the binary probit model. As in the linear regression model, the number of repetitions,  $R$ , used in the bootstrap bias correction is taken to be 407.

The results of Table 7(B) show a significant improvement in the performance of the three-step method when it is augmented by the bias-correction of  $\hat{\gamma}_{2B_1}$ . Most of the empirical levels are very close to the desired levels.

Table 7  
Standard errors—binary probit model<sup>a</sup>

pdb	Empirical level				$\hat{\gamma}_{2B_1}$											
	$B_1$	$B_2$	$B^*$	$B_o$	$B_1$	$B_o$	$B_2$	$B_2$	Mean	Med	Min	Max				
<i>(A) Using the three-step method with no bias correction</i>																
20	0.900	0.842	0.901	0.924	0.895	34	36	35	30	9	344	344	-0.49	-0.21	-1.41	12.22
10	0.900	0.833	0.892	0.915	0.884	136	141	142	129	59	1637	1637	-0.25	0.02	-1.18	16.32
5	0.900	0.821	0.885	0.899	0.880	542	562	576	538	331	8301	8301	-0.06	0.08	-0.75	22.87
20	0.950	0.900	0.952	0.959	0.937	48	50	50	41	15	455	455	-0.40	-0.14	-1.27	11.42
10	0.950	0.891	0.938	0.956	0.932	192	199	197	181	91	1968	1968	-0.21	0.02	-1.05	14.57
5	0.950	0.888	0.937	0.948	0.929	768	796	814	768	483	8594	8594	-0.02	0.12	-0.71	17.55
20	0.975	0.921	0.971	0.977	0.960	63	66	65	56	22	631	631	-0.35	-0.11	-1.32	12.76
10	0.975	0.917	0.969	0.972	0.955	251	260	264	242	125	5141	5141	-0.14	0.06	-0.96	26.16
5	0.975	0.917	0.968	0.971	0.962	1004	1040	1053	1016	657	9398	9398	-0.04	0.13	-0.67	15.32
<i>(B) Using the three-step method with bias correction</i>																
20	0.900	0.842	0.903	0.904	0.895	34	36	43	40	10	361	361	0.054	0.061	-1.30	15.65
10	0.900	0.833	0.902	0.903	0.884	136	141	154	144	53	1721	1721	0.061	0.064	-1.21	16.43
5	0.900	0.821	0.898	0.900	0.880	542	562	596	571	319	8200	8200	0.067	0.071	-1.10	19.46
20	0.950	0.900	0.945	0.948	0.947	48	50	60	52	17	467	467	0.054	0.065	-1.31	11.88
10	0.950	0.891	0.947	0.949	0.932	192	199	223	207	91	2023	2023	0.061	0.067	-1.18	14.07
5	0.950	0.888	0.949	0.950	0.929	768	796	798	772	501	8401	8401	0.070	0.074	-1.10	16.33
20	0.975	0.921	0.968	0.971	0.970	63	66	74	67	24	692	692	0.064	0.066	-1.21	10.43
10	0.975	0.917	0.971	0.973	0.960	251	260	274	258	131	4985	4985	0.066	0.068	-1.11	11.44
5	0.975	0.917	0.973	0.975	0.968	1004	1040	1102	1029	621	9537	9537	0.069	0.072	-0.98	15.24

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample, we carry out 2500 Monte Carlo repetitions.  $\hat{\gamma}_2$  is calculated for each of the 100 samples using 250,000 bootstrap repetitions. The average of the 100  $\hat{\gamma}_2$  values is 0.071. The value of  $R$  for the bias-corrected method is 407. The average of the  $CV$  values is 0.0028.

The largest deviation for  $B_2$  or  $B^*$  is 0.007 and most are less than half as large. The results in the ‘Mean’ column for  $\hat{\gamma}_{2B_1R}$  indicate that the bias of the bias-corrected estimator  $\hat{\gamma}_{2B_1R}$  is much smaller than that of  $\hat{\gamma}_{2B_1}$  in Table 7(A). As a consequence, the ‘Mean’ numbers of bootstrap repetitions  $B_2$  (and  $B^*$ ) are somewhat larger in Table 7(B) than in Table 7(A).

### 5.1.3. *Quantile regression model*

The results for standard error estimates in the quantile regression model are reported in Tables 8 and 9.<sup>8</sup> The empirical levels are a bit lower and farther from the desired levels in Tables 8 and 9 than in Tables 5 and 6 for the linear regression model, but the basic pattern of the tables is quite similar. In particular, the empirical levels are closer to the desired levels with  $N(0,1)$  errors than with  $t_5$  errors. The estimator  $\hat{\gamma}_{2B_1}$  is downward biased, especially with  $t_5$  errors. There is significant variation in the values of  $B_2$  over the various  $(p, \delta, \tau)$  combinations. When bias correction is introduced the performance of the three-step method improves significantly. As seen in Table 9, the empirical levels of the bias-corrected three-step method are quite close to the desired levels, with the largest deviation being 0.030 for  $B_2$  and 0.021 for  $B^*$ .

Based on the results presented above for the three econometric models, we conclude that the bias-corrected three-step method yields a noticeable improvement over the three-step method in cases where  $\gamma_2$  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  $\hat{s}e_{B^*}$  whenever  $\hat{\theta}$  is difficult to compute. Thus, we recommend the use of the bias-corrected three-step method in most cases.

## 5.2. *Symmetric confidence intervals*

### 5.2.1. *Linear regression model*

The results for symmetric confidence intervals in the linear regression model are reported in Table 10 for the  $N(0,1)$  and  $t_5$  error distributions. The numbers reported in this table are averages over 250 samples. The results for the  $\chi_5^2$  error distribution are very similar to those given in Table 10(B) for the  $t_5$  error distribution in terms of both the empirical levels obtained and the number of bootstrap repetitions  $B^*$  needed. These results show that the high skewness of the  $\chi_5^2$  error distribution does not have any effect

<sup>8</sup> As stated above, all results reported for the quantile regression model are for the case  $q = 0.75$ .

Table 8  
Standard errors using the three-step method with no bias correction—quantile regression model<sup>a</sup>

pdb	Empirical level										$\hat{\gamma}_{2B_1}$				
	$1 - \tau$	$B_1$	$B_2$	$B^*$	$B_o$	$B_1$	$B_o$	$B_2$			Mean	Med	Min	Max	
(A) Error distribution $N(0,1)$															
20	0.900	0.824	0.868	0.880	0.860	34	43	42	37	13	265	0.20	0.12	-1.26	14.22
10	0.900	0.829	0.874	0.882	0.869	136	174	173	166	79	1376	0.33	0.24	-0.98	18.39
5	0.900	0.837	0.892	0.894	0.875	542	697	694	642	434	4644	0.50	0.35	-0.79	28.25
20	0.950	0.898	0.931	0.939	0.920	48	62	64	54	21	421	0.31	0.19	-1.07	12.09
10	0.950	0.904	0.936	0.940	0.933	192	247	255	244	120	1931	0.42	0.32	-0.69	15.28
5	0.950	0.918	0.947	0.948	0.940	768	987	988	942	635	6954	0.57	0.43	-0.44	18.66
20	0.975	0.925	0.965	0.970	0.952	63	81	82	74	31	577	0.36	0.25	-1.02	11.07
10	0.975	0.931	0.971	0.973	0.962	251	323	309	288	172	2483	0.44	0.36	-0.61	13.56
5	0.975	0.939	0.975	0.975	0.969	1004	1290	1277	1189	833	9565	0.60	0.49	-0.22	15.33
(B) Error distribution $t_5$															
20	0.900	0.812	0.833	0.842	0.821	34	61	56	46	18	319	0.79	0.35	-1.05	16.83
10	0.900	0.817	0.856	0.862	0.845	136	246	233	201	95	1677	0.98	0.70	-0.89	19.66
5	0.900	0.828	0.886	0.891	0.875	542	981	895	859	523	5969	1.29	1.04	-0.46	22.09
20	0.950	0.861	0.909	0.913	0.899	48	87	79	71	23	475	0.81	0.38	-1.01	17.81
10	0.950	0.870	0.915	0.918	0.903	192	346	324	311	134	2332	1.09	0.83	-0.77	20.64
5	0.950	0.884	0.941	0.944	0.928	768	1390	1288	1208	754	7576	1.38	1.18	-0.20	24.05
20	0.975	0.908	0.944	0.949	0.934	63	114	92	81	27	669	0.98	0.87	-0.92	18.32
10	0.975	0.919	0.960	0.965	0.946	251	454	393	365	190	2831	1.20	1.08	-0.58	22.22
5	0.975	0.927	0.969	0.973	0.964	1004	1817	1633	1557	1089	9421	1.56	1.38	0.10	25.77

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample, we carry out 2500 Monte Carlo repetitions.  $\hat{\gamma}_2$  is calculated for each of the 100 samples using 250,000 bootstrap repetitions. The average of the 100  $\hat{\gamma}_2$  values is 0.57 in panel (A) and 1.62 in panel (B). The average of the  $CV$  values is 0.0031 in panel (A) and 0.0042 in panel (B).

Table 9  
Standard errors using the three-step method with bias correction—quantile regression model with error distribution  $t_5^a$

$p/db$	$1 - \tau$	Empirical level				$B_2$							$\hat{\tau}_{2B,R}$			
		$B_1$	$B_2$	$B^*$	$B_o$	$B_1$	$B_o$	Mean	Med	Min	Max	Mean	Med	Min	Max	
20	0.900	0.813	0.870	0.879	0.821	34	61	58	49	19	344	1.22	1.09	-1.12	19.13	
10	0.900	0.817	0.877	0.883	0.845	136	246	236	210	100	1690	1.35	1.18	-0.99	22.90	
5	0.900	0.827	0.893	0.897	0.875	542	981	911	887	560	6226	1.53	1.39	-0.50	25.39	
20	0.950	0.860	0.929	0.937	0.899	48	87	70	62	25	481	1.30	1.19	-1.04	20.74	
10	0.950	0.872	0.936	0.942	0.903	192	346	320	291	140	2377	1.43	1.24	-0.67	23.63	
5	0.950	0.885	0.940	0.947	0.928	768	1390	1255	1228	807	7721	1.58	1.46	-0.11	25.26	
20	0.975	0.907	0.953	0.959	0.934	63	114	99	83	30	695	1.35	1.23	-0.81	20.46	
10	0.975	0.917	0.960	0.969	0.946	251	454	420	382	208	2877	1.46	1.35	-0.42	24.99	
5	0.975	0.928	0.971	0.975	0.964	1004	1822	1625	1569	1124	9512	1.64	1.51	0.19	26.93	

<sup>a</sup>Note: See note for Table 8. The value of  $R$  for the bias-corrected method is 407. The average of the  $CI^*$  values is 0.0042.

Table 10  
Symmetric confidence intervals—linear regression model<sup>a</sup>

$1 - \alpha$	$pdb$	$1 - \tau$	Empirical level				$B_2$					
			$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) <i>Error distribution</i> N(0,1)												
0.900	15	0.900	0.842	0.942	0.949	0.951	99	255	257	218	27	1830
0.900	10	0.900	0.847	0.924	0.931	0.932	219	401	393	367	70	1491
0.900	5	0.900	0.851	0.907	0.912	0.917	849	1348	1313	1280	477	3014
0.900	15	0.950	0.911	0.967	0.971	0.977	139	319	311	271	46	1513
0.900	10	0.950	0.919	0.958	0.966	0.965	309	535	524	492	121	1649
0.900	5	0.950	0.921	0.952	0.955	0.960	1209	1839	1819	1776	751	3835
0.950	15	0.900	0.833	0.948	0.953	0.961	119	573	563	467	31	16,404
0.950	10	0.900	0.839	0.944	0.949	0.958	259	763	751	649	100	4309
0.950	5	0.900	0.845	0.913	0.917	0.922	979	1949	1922	1837	588	5132
0.950	15	0.950	0.910	0.987	0.989	0.994	159	1242	1209	806	70	35,314
0.950	10	0.950	0.922	0.968	0.970	0.979	359	1419	1371	1345	149	3837
0.950	5	0.950	0.930	0.950	0.955	0.962	1399	2653	2601	2522	935	6101
(B) <i>Error distribution</i> $t_5$												
0.900	15	0.900	0.839	0.939	0.943	0.959	99	298	271	234	28	1972
0.900	10	0.900	0.844	0.924	0.927	0.940	219	439	414	381	77	1509
0.900	5	0.900	0.847	0.906	0.911	0.922	849	1422	1377	1334	501	3187
0.900	15	0.950	0.907	0.966	0.969	0.983	139	349	333	282	47	1690
0.900	10	0.950	0.913	0.954	0.958	0.975	309	568	551	511	122	1789
0.900	5	0.950	0.920	0.952	0.954	0.961	1209	1992	1909	1840	788	4048
0.950	15	0.900	0.834	0.946	0.951	0.968	119	605	582	521	32	18,401
0.950	10	0.900	0.838	0.944	0.949	0.964	259	822	804	712	102	4445
0.950	5	0.900	0.842	0.920	0.926	0.928	979	2111	2045	1933	629	5344
0.950	15	0.950	0.911	0.989	0.989	0.993	159	1313	1271	1042	74	39,203
0.950	10	0.950	0.927	0.969	0.969	0.982	359	1498	1427	1298	161	3962
0.950	5	0.950	0.933	0.955	0.957	0.969	1399	2888	2812	2721	1017	6548

<sup>a</sup>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 2500 Monte Carlo repetitions. The average of the *CV* values is 0.0043 in panel (A) and 0.0067 in panel (B).

on the performance of the three-step method. For brevity, we do not report these results.

Table 10(A) shows that the empirical levels are higher than the corresponding  $1 - \tau$  values for the experiments with the N(0,1) error distribution. With relatively low *pdb* (e.g., *pdb*=5), the empirical levels are quite close to the desired levels. Table 10(A) indicates that the performance of the three-step method is determined by the number of bootstrap repetitions,  $B_2$  or  $B^*$ , employed. The  $(\alpha, pdb, \tau)$  combinations that yield the best results are those that induce a relatively large number of bootstrap repetitions. For example, for



the (0.10, 5, 0.10) combination, the median  $B_2$  value is 1280, while for the combination (0.10, 15, 0.10), it is only 218. As a result, the empirical level for the former case is 0.907, which is quite close to 0.900, while for the latter it is 0.942.

The empirical levels are closer to the desired levels for the confidence intervals with lower confidence level  $1 - \alpha$ . This occurs because it is more difficult to estimate the 0.95 quantile of  $T^*$  needed for a 95% confidence interval than to estimate the 0.90 quantile of  $T^*$  needed for a 90% confidence interval.

Table 10(B) reports the results from the Monte Carlo simulations with the  $t_5$  error distribution. The general picture revealed by Table 10(B) is very similar to that of Table 10(A). The empirical levels are comparable to those reported in Table 10(A). They are somewhat higher than the desired levels. The most pronounced difference between the two sets of experiments is that for all  $(\alpha, pdb, \tau)$  combinations, the number of bootstrap repetitions  $B^*$  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  $N(0,1)$  distribution itself.

Lastly, we consider the empirical unconditional coverage probabilities. In all cases, they are the same whether based on  $B_2$  or  $B^*$  bootstrap repetitions. For normal errors, they are in the range 0.907–0.909 for all cases where  $\alpha = 0.900$  and in the range 0.954–0.956 for all cases where  $\alpha = 0.950$ . For  $t_5$  errors, they are in the range 0.900–0.902 for all cases where  $\alpha = 0.900$  and in the range 0.951–0.953 for all cases where  $\alpha = 0.950$ . Thus, the empirical unconditional coverage probabilities are extremely close to the desired levels. 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 conditional coverage probability given the bootstrap simulation randomness is close to the desired coverage probability and whose length is close to that of the ideal bootstrap confidence interval, one does need to employ a relatively large number of bootstrap repetitions.

### 5.2.2. Binary probit model

The results for symmetric confidence intervals in the binary probit model are reported in Table 11 for the  $N(0,1)$  error distribution. Table 11 shows that, as in the linear regression model, the empirical levels are somewhat higher than the corresponding  $1 - \tau$  values. Nevertheless, with relatively low  $pdb$  (e.g.,  $pdb = 5$ ), the empirical levels are quite close to the desired levels. Overall, the empirical levels are somewhat closer to the desired levels in the binary probit model than in the linear regression model.

Table 11  
Symmetric confidence intervals—binary probit model<sup>a</sup>

$1 - \alpha$	$pdb$	$1 - \tau$	Empirical level				$B_2$					
			$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) Error distribution N(0,1)												
0.900	15	0.900	0.853	0.937	0.942	0.944	99	155	146	137	28	1908
0.900	10	0.900	0.844	0.920	0.923	0.932	219	297	291	259	70	2087
0.900	5	0.900	0.833	0.904	0.906	0.910	849	937	924	866	460	3076
0.900	15	0.950	0.902	0.959	0.961	0.965	139	204	193	178	49	1506
0.900	10	0.950	0.893	0.955	0.958	0.962	309	456	445	409	128	1722
0.900	5	0.950	0.884	0.951	0.952	0.955	1209	1333	1298	1232	578	3308
0.950	15	0.900	0.841	0.935	0.942	0.943	119	168	157	144	37	1625
0.950	10	0.900	0.832	0.930	0.934	0.939	259	320	300	277	75	1876
0.950	5	0.900	0.819	0.906	0.908	0.911	979	1141	1102	1032	508	3143
0.950	15	0.950	0.909	0.975	0.978	0.985	159	231	212	197	101	1549
0.950	10	0.950	0.902	0.956	0.959	0.976	359	432	413	384	162	2011
0.950	5	0.950	0.899	0.952	0.954	0.963	1399	1533	1457	1401	599	3018

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample we carry out 2500 Monte Carlo repetitions. The average of the CV values is 0.0061.

Table 11 indicates that, just as in Table 10, the  $(\alpha, pdb, \tau)$  combinations that yield the best results are those that induce a relatively large number of bootstrap repetitions. As in the linear regression model, the empirical levels are closer to the desired levels for the confidence intervals with lower confidence level  $1 - \alpha$ , for the same reason as discussed above.

The empirical unconditional coverage probabilities for the binary probit model are the same whether based on  $B_2$  or  $B^*$  bootstrap repetitions. They are in the range 0.904–0.905 for all cases where  $\alpha = 0.900$  and in the range 0.952–0.954 for all cases where  $\alpha = 0.950$ .

### 5.2.3. Quantile regression model

The results for symmetric confidence intervals for the quantile regression model are reported in Table 12 for the N(0,1) and  $t_5$  error distributions. The results are essentially the same as those obtained for the linear regression model. In particular, the empirical levels for both the experiment with N(0,1) error, reported in Table 12(A), and the experiment with  $t_5$  error, reported in Table 12(B), are somewhat above the desired levels. For tighter bounds (e.g.,  $(pdb, \tau) = (5, 0.05)$  or  $(10, 0.05)$ ), however, the empirical levels are quite close to  $1 - \tau$ .

The most important difference between the experiments reported in Table 12 and those reported in Table 10 is in the eventual number of bootstrap repetitions. The ‘Mean’ and ‘Median’ numbers of bootstrap repetitions are

Table 12  
Symmetric confidence intervals—quantile regression model<sup>a</sup>

$1 - \alpha$	$pdb$	$1 - \tau$	Empirical level				$B_2$					
			$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) Error distribution N(0,1)												
0.900	15	0.900	0.838	0.943	0.948	0.949	99	273	266	235	32	1966
0.900	10	0.900	0.818	0.927	0.930	0.934	219	422	409	389	84	2064
0.900	5	0.900	0.798	0.903	0.906	0.914	849	1334	1302	1264	477	3245
0.900	15	0.950	0.914	0.966	0.972	0.975	139	354	344	292	51	2016
0.900	10	0.950	0.905	0.956	0.962	0.966	309	574	555	521	124	2264
0.900	5	0.950	0.888	0.950	0.952	0.959	1209	2039	1981	1905	759	3898
0.950	15	0.900	0.837	0.953	0.958	0.963	119	634	590	411	50	12,875
0.950	10	0.900	0.821	0.947	0.951	0.950	259	854	804	722	109	4701
0.950	5	0.900	0.807	0.915	0.918	0.919	979	2091	2023	1913	628	5413
0.950	15	0.950	0.922	0.970	0.974	0.977	159	1422	1399	921	72	28,354
0.950	10	0.950	0.912	0.959	0.963	0.963	359	1978	1876	955	176	4832
0.950	5	0.950	0.899	0.950	0.953	0.957	1399	2939	2804	2521	985	6689
(B) Error distribution $t_5$												
0.900	15	0.900	0.830	0.940	0.945	0.954	99	333	302	258	30	1954
0.900	10	0.900	0.808	0.925	0.933	0.939	219	476	445	409	84	2021
0.900	5	0.900	0.787	0.904	0.907	0.918	849	1555	1500	1433	556	3222
0.900	15	0.950	0.909	0.961	0.967	0.981	139	384	352	320	55	1843
0.900	10	0.950	0.905	0.953	0.958	0.972	309	619	591	543	142	2181
0.900	5	0.950	0.879	0.948	0.952	0.964	1209	2082	2033	1911	830	4166
0.950	15	0.900	0.828	0.958	0.966	0.969	119	654	622	541	38	7035
0.950	10	0.900	0.803	0.952	0.957	0.959	259	873	822	721	108	4831
0.950	5	0.900	0.799	0.908	0.914	0.921	979	2189	2128	2021	676	6032
0.950	15	0.950	0.901	0.969	0.975	0.981	159	1528	1478	1199	83	9309
0.950	10	0.950	0.891	0.961	0.964	0.968	359	2097	1973	1634	174	4102
0.950	5	0.950	0.881	0.951	0.954	0.961	1399	3112	2946	2619	1162	7218

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample we carry out 2500 Monte Carlo repetitions. The average of the CV values is 0.0052 in panel (A) and 0.0071 in panel (B).

somewhat higher for the latter case. It is worth noting that for the 0.50 quantile estimator (not reported here for brevity) the numbers of bootstrap repetitions are comparable to those for the LS estimator in the linear regression model.

The empirical unconditional coverage probabilities for the quantile regression model are the same whether based on  $B_2$  or  $B^*$  bootstrap repetitions. For the normal error case, they are in the range 0.907–0.909 for all cases where  $\alpha = 0.900$  and in the range 0.953–0.955 for all cases where  $\alpha = 0.950$ . For the  $t_5$  error case, they are in the range 0.905–0.906 for all cases where  $\alpha = 0.900$  and in the range 0.949–0.951 for all cases where  $\alpha = 0.950$ .

Table 13  
 Tests with given significance level  $\alpha$ —linear and quantile regression models<sup>a</sup>

$1 - \alpha$	$pdb$	$1 - \tau$	Empirical level				$B_2$					
			$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) <i>Linear regression model with error distribution <math>t_5</math></i>												
0.950	15	0.900	0.810	0.855	0.859	0.847	197	276	264	243	84	411
0.950	10	0.900	0.825	0.868	0.871	0.856	446	572	552	533	181	1102
0.950	5	0.900	0.832	0.891	0.897	0.879	1785	2397	2336	2279	1099	3169
0.950	15	0.950	0.891	0.926	0.933	0.922	281	405	381	360	101	487
0.950	10	0.950	0.901	0.936	0.940	0.930	633	827	802	767	338	1437
0.950	5	0.950	0.915	0.948	0.949	0.939	2535	3227	3108	3067	1389	4354
0.900	15	0.900	0.822	0.848	0.855	0.834	213	276	254	232	95	381
0.900	10	0.900	0.833	0.866	0.872	0.853	480	577	540	519	201	1012
0.900	5	0.900	0.841	0.883	0.889	0.870	1924	2309	2223	2156	1289	2933
0.900	15	0.950	0.899	0.920	0.926	0.910	303	371	358	345	109	454
0.900	10	0.950	0.908	0.932	0.937	0.922	682	792	775	734	364	1305
0.900	5	0.950	0.914	0.937	0.943	0.930	2733	3224	3032	2914	1501	4124
(B) <i>Quantile regression model with error distribution <math>t_5</math></i>												
0.950	15	0.900	0.803	0.849	0.854	0.842	197	299	273	261	94	515
0.950	10	0.900	0.811	0.873	0.877	0.870	446	613	590	563	189	1141
0.950	5	0.900	0.823	0.897	0.899	0.878	1785	2613	2516	2409	1102	3447
0.950	15	0.950	0.871	0.928	0.931	0.921	281	440	422	382	112	591
0.950	10	0.950	0.891	0.938	0.942	0.928	633	908	869	834	329	1519
0.950	5	0.950	0.903	0.950	0.950	0.932	2535	3502	3288	3200	1388	4433
0.900	15	0.900	0.842	0.841	0.847	0.829	213	286	280	255	102	407
0.900	10	0.900	0.851	0.856	0.862	0.842	480	599	573	539	204	1110
0.900	5	0.900	0.876	0.879	0.882	0.866	1924	2424	2312	2224	1256	3042
0.900	15	0.950	0.879	0.917	0.923	0.909	303	404	384	358	118	484
0.900	10	0.950	0.894	0.921	0.924	0.919	682	827	791	760	339	1362
0.900	5	0.950	0.903	0.938	0.944	0.928	2733	3312	3166	3089	1476	4290

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample we carry out 2500 Monte Carlo repetitions. The average of the CV values is 0.0081 in panel (A) and 0.0092 in panel (B).

### 5.3. Tests with given significance level

In this section, we evaluate the performance of the three-step method in testing situations with a prespecified level of significance  $\alpha$ . For brevity, we present results only for the  $t_5$  error distribution. The results for the normal error distribution are quite similar.

#### 5.3.1. Linear regression model

The results for the linear regression model are reported in Table 13(A). Table 13(A) indicates that the empirical levels are somewhat lower than the

desired levels. With moderate bounds on  $(pdb, \tau)$ , quite reasonable results are obtained. For example, for  $(\alpha, pdb, \tau) = (0.05, 10, 0.05)$ , the empirical level based on  $B_2$  is just 0.014 below the desired 0.950 level. The ‘Mean’ number of bootstrap repetitions required to achieve this is rather moderate, only 802 bootstrap repetitions.

Tighter bounds on  $(pdb, \tau)$  increase the number of bootstrap repetitions  $B_2$  and  $B^*$  and, hence, significantly increase the precision of the three-step method. For example, with  $(\alpha, pdb, \tau) = (0.05, 5, 0.05)$ , the empirical level is almost identical to the desired level. But, in this case, the number of bootstrap repetitions  $B_2$  is 3108 on average.

The number of bootstrap repetitions varies dramatically over the  $(\alpha, pdb, \tau)$  combinations. For example, the maximum number of bootstrap repetitions (over the 100 samples and the 2500 simulations for each sample) for the  $(0.10, 15, 0.10)$  combination is lower than the minimum number of bootstrap repetitions for the  $(0.05, 5, 0.05)$  combination.

Overall, the three-step method seems to perform quite well and does not impose excessive computational costs. A high degree of accuracy can be achieved, though at the cost of employing a large number of bootstrap repetitions.

### 5.3.2. *Quantile regression model*

The results for tests with given  $\alpha$  in the quantile regression model are reported in Table 13(B). Table 13(B) shows that, in general, the results for the linear regression model also hold for the quantile regression model. The empirical levels are somewhat lower than the desired levels and are somewhat lower than those for the linear regression model.

For moderate bounds on  $(pdb, \tau)$ , quite reasonable precision is obtained. For example, for  $(\alpha, pdb, \tau) = (0.05, 10, 0.05)$ , the empirical level for  $B_2$  repetitions is 0.938, whereas the desired value is 0.950. The ‘Mean’ number of bootstrap repetitions required to achieve this, 869, is slightly larger than in the linear regression model.

Overall, the three-step method seems to perform quite well for tests with given  $\alpha$  in the quantile regression model.

### 5.3.3. *Binary probit model*

The results for tests with given  $\alpha$  in the binary probit model are reported in Table 14. This table shows that, by and large, the results of the previous two models hold for the binary probit as well. The empirical levels are somewhat lower than the desired levels and are somewhat lower than those for the linear regression model. Nevertheless, they are as good as, and in most cases better than, those for the quantile regression model.

For moderate bounds on  $(pdb, \tau)$ , quite reasonable precision is obtained. For example, for  $(\alpha, pdb, \tau) = (0.05, 10, 0.05)$ , the empirical level for  $B_2$

Table 14  
 Tests with given significance level  $\alpha$ —binary probit model<sup>a</sup>

$1 - \alpha$	$pdb$	$1 - \tau$	Empirical level				$B_2$					
			$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
0.950	15	0.900	0.806	0.854	0.865	0.844	197	289	265	251	92	517
0.950	10	0.900	0.814	0.879	0.884	0.873	446	600	568	536	180	1169
0.950	5	0.900	0.827	0.891	0.898	0.879	1785	2504	2445	2389	1120	3431
0.950	15	0.950	0.875	0.932	0.937	0.922	281	429	411	376	113	592
0.950	10	0.950	0.894	0.940	0.947	0.930	633	868	851	828	330	1540
0.950	5	0.950	0.913	0.946	0.949	0.934	2535	3388	3229	3166	1378	4455
0.900	15	0.900	0.847	0.846	0.853	0.832	213	280	266	250	111	428
0.900	10	0.900	0.854	0.859	0.866	0.846	480	584	561	532	220	1123
0.900	5	0.900	0.879	0.881	0.891	0.868	1924	2374	2270	2188	1219	3060
0.900	15	0.950	0.883	0.916	0.924	0.910	303	391	372	348	120	505
0.900	10	0.950	0.901	0.924	0.933	0.920	682	803	780	751	327	1399
0.900	5	0.950	0.913	0.941	0.950	0.930	2733	3269	3127	3079	1488	4333

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample we carry out 2500 Monte Carlo repetitions. The average of the *CV* values is 0.0087.

repetitions is 0.940, whereas the desired value is 0.950. The ‘Mean’ number of bootstrap repetitions  $B_2$  required to achieve this, 851, is slightly larger than in the linear regression model, but smaller than in the quantile regression model.

Overall, the three-step method performs quite well for tests with given  $\alpha$  in the binary probit model.

#### 5.4. *p-Values*

In this section, we evaluate the performance of the three-step method in testing situations where one is interested in computing a *p*-value. Again, for brevity, we only report results for the  $t_5$  error distribution. The results for the normal error distribution are similar.

##### 5.4.1. *Linear regression model*

The results for *p*-values in the linear regression model are reported in Table 15. Table 15(A) provides results for samples in which the asymptotic *p*-value is approximately 0.05. Table 15(B) reports results when the asymptotic *p*-value is approximately 0.10.

Table 15(A) indicates that the three-step bootstrap method leads to empirical levels that are somewhat lower than the desired levels. For tighter

Table 15  
*p*-Values—linear regression model with error distribution  $t_5^a$

<i>pdb</i>	$1 - \tau$	Empirical level				$B_2$					
		$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) <i>p</i> -Value = 0.05											
20	0.900	0.813	0.838	0.842	0.820	1285	1429	1404	1347	981	7017
15	0.900	0.821	0.860	0.865	0.832	2285	2453	2399	2342	1476	9112
10	0.900	0.854	0.884	0.893	0.853	5141	5360	5249	5162	2769	12,223
20	0.950	0.872	0.894	0.902	0.861	1825	2090	1969	1896	1201	7911
15	0.950	0.887	0.921	0.926	0.873	3244	3706	3570	3485	1702	11,087
10	0.950	0.905	0.941	0.948	0.901	7299	8009	7784	7646	2888	16,290
20	0.975	0.901	0.953	0.958	0.905	2386	2602	2478	2402	1701	9037
15	0.975	0.918	0.964	0.969	0.927	4242	4735	4474	4366	2099	11,603
10	0.975	0.928	0.974	0.975	0.946	9545	10222	9986	9802	4798	19,033
(B) <i>p</i> -Value = 0.10											
20	0.900	0.820	0.852	0.859	0.832	609	729	712	664	304	1104
15	0.900	0.832	0.867	0.871	0.841	1082	1209	1188	1101	466	1499
10	0.900	0.861	0.898	0.904	0.864	2435	2700	2625	2562	1612	3483
20	0.950	0.884	0.910	0.918	0.894	864	992	935	870	404	1262
15	0.950	0.898	0.923	0.926	0.903	1537	1822	1722	1673	622	1789
10	0.950	0.911	0.948	0.951	0.929	3457	3917	3767	3580	2002	4555
20	0.975	0.909	0.961	0.966	0.930	1130	1309	1240	1176	704	1580
15	0.975	0.926	0.967	0.970	0.941	2010	2316	2201	2153	832	2900
10	0.975	0.938	0.974	0.975	0.954	4521	4801	4598	4509	3566	5884

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample, we carry out 2500 Monte Carlo repetitions. The average of the *CV* values is 0.0079 in panel (A) and 0.0057 in panel (B).

bounds, i.e., small *pdb* and  $\tau$ , the performance is quite good. This comes at the cost, however, of a large number of bootstrap repetitions. For example, for  $(pdb, \tau) = (10, 0.05)$ , the ‘Mean’ number of repetitions is 7784, while the ‘Median’ number is only slightly less.

The general picture revealed by the results reported in Table 15(B) is similar to that of Table 15(A). There are, however, two significant differences. First, the number of bootstrap repetitions is considerably lower in Table 15(B) than in Table 15(A). Second, the empirical levels in the Table 15(B) are closer to the desired levels. This is a consequence of the fact that it is much harder to estimate smaller *p*-values. Fortunately, there are many situations in which one should not care whether or not the *p*-value is estimated precisely. For example, if the *p*-value is 0.001, one can estimate it with relatively low accuracy.

Table 16  
*p*-Values—binary probit model<sup>a</sup>

<i>pd</i> <i>b</i>	$1 - \tau$	Empirical level				$B_2$					
		$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) <i>p</i> -Value = 0.05											
20	0.900	0.807	0.828	0.840	0.820	1285	1432	1402	1370	892	7216
15	0.900	0.817	0.844	0.859	0.831	2285	2416	2384	2317	1431	9215
10	0.900	0.848	0.880	0.882	0.852	5141	5221	5191	5101	2407	11,995
20	0.950	0.861	0.892	0.899	0.864	1825	1922	1877	1826	1111	7958
15	0.950	0.880	0.920	0.919	0.888	3244	3604	3563	3488	1472	11,692
10	0.950	0.900	0.940	0.939	0.904	7299	7904	7807	7661	2791	16,731
20	0.975	0.892	0.947	0.954	0.929	2386	2514	2476	2406	1616	9190
15	0.975	0.914	0.961	0.961	0.942	4242	4448	4407	4238	1905	12,005
10	0.975	0.921	0.974	0.975	0.966	9545	10,029	9996	9882	4701	18,388
(B) <i>p</i> -Value = 0.10											
20	0.900	0.818	0.848	0.850	0.839	609	711	701	671	369	1232
15	0.900	0.828	0.864	0.869	0.849	1082	1199	1188	1128	451	1616
10	0.900	0.861	0.892	0.899	0.862	2435	2617	2597	2502	1509	3718
20	0.950	0.880	0.901	0.909	0.891	864	949	930	887	391	1351
15	0.950	0.895	0.917	0.922	0.902	1537	1738	1701	1645	617	1992
10	0.950	0.908	0.940	0.944	0.919	3457	3821	3768	3690	1985	4711
20	0.975	0.902	0.951	0.956	0.928	1130	1243	1207	1158	693	1601
15	0.975	0.920	0.959	0.960	0.939	2010	2221	2178	2090	804	3040
10	0.975	0.931	0.968	0.971	0.957	4521	4597	4503	4399	3519	5987

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample, we carry out 2500 Monte Carlo repetitions. The average of the *CV* values is 0.0087 in panel (A) and 0.0066 in panel (B).

Overall, the three-step method seems to perform reasonably well. The drawback is that it requires rather large number of bootstrap repetitions to achieve precise results.

### 5.4.2. Binary probit model

The results for *p*-values in the binary probit model are reported in Table 16. The results in this table are quite similar to those obtained for the linear regression model. In general, the empirical values are lower than the desired levels. This is especially true for the the lower *p*-value, 0.05. While the results for the higher *p*-value, 0.10, are better than those for the lower *p*-value, the empirical levels are still significantly below the desired levels.



Table 17  
*p*-Values—quantile regression model with error distribution  $t_5^a$

<i>pdb</i>	$1 - \tau$	Empirical level						$B_2$			
		$B_1$	$B_2$	$B^*$	$B_\omega$	$B_1$	$B_\omega$	Mean	Med	Min	Max
(A) <i>p</i> -Value = 0.05											
20	0.900	0.804	0.825	0.832	0.816	1285	1452	1424	1366	981	7145
15	0.900	0.811	0.847	0.853	0.825	2285	2501	2402	2362	1478	9099
10	0.900	0.847	0.877	0.879	0.860	5141	5406	5287	5200	2688	12,243
20	0.950	0.863	0.891	0.897	0.871	1825	2053	1987	1922	1205	7917
15	0.950	0.871	0.918	0.922	0.902	3244	3791	3635	3567	1589	11,425
10	0.950	0.898	0.937	0.941	0.921	7299	8111	7819	7730	2777	16,559
20	0.975	0.890	0.947	0.951	0.930	2386	2660	2521	2471	1654	9087
15	0.975	0.907	0.959	0.964	0.939	4242	4693	4495	4399	2026	11,764
10	0.975	0.916	0.974	0.975	0.949	9545	10,391	10,113	9863	4702	19,401
(B) <i>p</i> -Value = 0.10											
20	0.900	0.810	0.844	0.853	0.828	609	745	708	666	319	1123
15	0.900	0.821	0.860	0.869	0.839	1082	1262	1208	1122	470	1466
10	0.900	0.854	0.890	0.898	0.860	2435	2706	2630	2587	1623	3456
20	0.950	0.867	0.904	0.910	0.872	864	1004	954	892	406	1245
15	0.950	0.890	0.916	0.922	0.896	1537	1814	1743	1698	641	1893
10	0.950	0.900	0.943	0.947	0.912	3457	3955	3816	3622	1992	4516
20	0.975	0.900	0.957	0.962	0.918	1130	1302	1255	1197	717	1578
15	0.975	0.917	0.964	0.967	0.929	2010	2398	2222	2156	844	2935
10	0.975	0.925	0.972	0.974	0.940	4521	4781	4600	4538	3589	5888

<sup>a</sup>Note: The reported numbers are the averages over the simulations performed for 100 samples, each of which consists of 25 observations. For each sample, we carry out 2500 Monte Carlo repetitions. The average of the *CV* values is 0.0092 in panel (A) and 0.0073 in panel (B).

### 5.4.3. Quantile regression model

The results for *p*-values in the quantile regression model are reported in Table 17. As Table 17 shows, the results for the quantile regression model are almost identical to those reported in Table 15 for the linear regression model. In general, the empirical values are lower than the desired levels. For the lower *p*-value, of 0.05, the empirical values are somewhat lower than those for the linear regression model. For the higher *p*-value, of 0.10, however, the results for the quantile regression model are quite similar to those of the linear regression model.

## 6. Summary and conclusions

This paper focuses on practical aspects of choosing the number of bootstrap repetitions in a variety of situations commonly encountered in the empirical

econometric literature. A three-step method for determining the number of bootstrap repetitions in these situations is introduced in Andrews and Buchinsky (2000). The current paper investigates the small sample properties of this method in a number of econometric models, viz., linear regression, binary probit, and quantile regression. The paper investigates the method's performance for a variety of statistical procedures, viz., computation of standard errors, confidence intervals, tests, and  $p$ -values. In all cases considered, the three-step method provides reasonable results. That is, the three-step method comes quite close to achieving the desired precision in estimating the ideal bootstrap quantity of interest.

The three-step method specifies the starting values for the number of bootstrap repetitions,  $B_1$ , based on the asymptotic distribution of the relevant bootstrap statistic. We find that the final number of bootstrap repetitions,  $B^*$ , is usually noticeably larger than the initial value. We also find that  $B^*$  varies considerably across different statistical procedures, such as standard errors, confidence intervals, etc., and across different samples for any given statistical procedure. These findings indicate that it is important to choose the number of bootstrap repetitions in a manner that takes into account the details of the situation at hand, as the three-step method does, rather than to apply some rough rule of thumb.

We find that in almost all cases the number of bootstrap repetitions chosen by the three-step method is quite large relative to what is found in empirical applications in the literature. A common number of bootstrap repetitions reported in the literature is around 100. The experiments performed here show that in many cases this number is not sufficient to estimate statistical quantities with a high degree of accuracy. This is rather disconcerting, because it indicates that statistical inferences in the literature may depend on the particular bootstrap samples that were drawn. The three-step method that we investigate here resolves this problem and is shown to work quite well for the situations examined.

### **Acknowledgements**

The authors thank Ariel Pakes, Jim Powell, Quang Vuong, and five anonymous referees for helpful comments; and Glenna Ames for typing part of the manuscript. D.W.K. Andrews acknowledges the research support of the National Science Foundation via grant numbers SBR-9410975 and SBR-9730277. M. Buchinsky acknowledges the research support of the National Science Foundation via grant number SBR-9320386 and the Alfred P. Sloan Foundation via a Research Fellowship.

## References

- Andrews, D.W.K., Buchinsky, M., 2000. A three-step method for choosing the number of bootstrap repetitions. *Econometrica* 67, 23–51.
- Andrews, D.W.K., Buchinsky, M., 2001. On the number of bootstrap repetitions for BCa confidence intervals. *Econometric Theory* 16, forthcoming.
- Barrodale, I., Roberts, F., 1973. An improved algorithm for discrete  $l_1$  linear approximation. *SIAM Journal of Numerical Analysis* 10, 839–848.
- Beran, R., 1988. Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–697.
- Bloch, D.A., Gastwirth, J.L., 1968. On a simple estimate of the reciprocal of the density function. *Annals of Mathematical Statistics* 39, 1083–1085.
- Buchinsky, M., 1995. Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study. *Journal of Econometrics* 68, 303–338.
- Davidson, R., MacKinnon, J.G., 2000. Bootstrap tests: How many bootstraps? *Econometric Reviews* 19, 55–68.
- Davison, A.C., Hinkley, D.V., 1997. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, UK.
- Efron, B., 1987. Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association* 82, 171–200.
- Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1, 54–77.
- Hall, P., 1986. On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics* 14, 1453–1462.
- Hall, P., 1992. *The bootstrap and edgeworth expansion*. Springer, New York, NY.
- Hall, P., Sheather, S.J., 1988. On the distribution of a studentized quantile. *Journal of the Royal Statistical Society Series B* 50, 381–391.
- Koenker, R., Bassett, W., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R., Bassett, W., 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61.
- 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.