The bootstrap
Joel Horowitz

Report by Jeroen Kerkhof, Tilburg University[¤]

Abstract

In this paper, we investigate the bootstrap. This is a statistical technique to calculate the distribution of statistics (e.g. estimators or test statistics). We derive and exemplify the possibilities and virtues of the bootstrap.

1 Introduction

A well-known problem in econometrics is ¯nding the distribution of an estimator or test statistic in a ¯nite sample. Usually, applied econometricians rely on ¯rst-order asymptotic theory to solve this problem. This can be quite inaccurate, especially in small samples. In this paper we therefore investigate the bootstrap, which is an alternative method for estimating the distribution of an estimator or test statistic in a ¯nite sample by resampling the data. The main idea of the bootstrap is treating the data as if they were the whole population. As we show in the remainder, the bootstrap has two important advantages compared to ¯rst-order asymptotic theory. First, it substitutes the often di±cult mathematical analysis of ¯rst-order asymptotic theory by computer power. Second, under some mild regularity conditions the approximation obtained by the bootstrap is at least as good as that obtained by ¯rst-order asymptotic theory. The set-up of this paper is as follows. We start with the most well-known model in econometrics, the classical regression model to get a °avour for the possibilities of the bootstrap. In Section 3; we discuss the general problem. Subsequently, the theory behind the bootstrap is explained. We discuss some selected topics. Finally, we illustrate the presented theory with simulation studies and conclude.

[¤]Correspondence to: Jeroen Kerkhof, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. I would like to thank Laurens Swinkels for the various discussions to enlarge our understanding of the bootstrap.

2 Motivating example: Classical linear regression

The most often used model in econometrics is without any doubt the classical regression model. It therefore seems a good starting point to illustrate the possible virtues of the bootstrap within this model. Recollecting the set-up of the classical regression model. We have available a data set containing $\{(y_1; x_1); (y_2; x_2); \ldots; (y_n; x_n)\}$. We assume that the data are a realization of a random sample $\{(Y_1; X_1); (Y_2; X_2); \ldots; (Y_n; X_n)\}$ from an unknown distribution F_0. Our model is the classical regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i. \tag{1}$$

with an i.i.d. sequence $\{\varepsilon_i\}$ such that

$$\begin{aligned} \mathbb{E}(\varepsilon_i) &= 0 \\ \mathbb{E}(X_i \varepsilon_i) &= 0 \\ \mathbb{E}\left(\varepsilon_i^2\right) &= \sigma^2. \end{aligned} \tag{2}$$

We estimate $(\beta_0; \beta_1)$ by the OLS estimator b_{OLS}[1]

$$b_{OLS} = (X'X)^{-1} X'Y \tag{3}$$

where

$$X = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \text{ and } Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

and σ^2 by s^2 [2]

$$s^2 = \frac{1}{n-2}(Y - Xb)'(Y - Xb) \tag{4}$$

We would like to know whether X_i helps in explaining Y_i. Stated more formally, we would like to test

$$H_0 : \beta_1 = 0 \text{ vs. } H_0 : \beta_1 \neq 0 \tag{5}$$

To perform this test, we use a test statistic $T_n = T_n(X_1; \ldots; X_n)$:

$$T_n(X_1; \ldots; X_n) = \sqrt{n}\frac{b_1}{s(X'X)^{-1}_{22}} \tag{6}$$

---

[1] We also denote the OLS estimate $b_{OLS} = (x'x)^{-1} x'y$ by b_{OLS}. where $x; y$ denote realizations of $X; Y$; respectively. It will be clear from the context which is meant.

[2] Here also s^2 denotes both the estimator and the estimate for σ^2.

If we assume that $\mathcal{L}(\varepsilon_i) = N(0, \sigma^2)$[3], we have that under some regularity conditions

$$\mathcal{L}(b_{OLS}) = N\left(\beta, \operatorname{plim} \sigma^2 (X'X)^{-1}\right)$$

Standard statistical books give that

$$\mathcal{L}(T_n) = t_{n-2}.$$

We have found the distribution for the test statistic T_n in a finite sample. To achieve this result, we had to make distributional assumptions on the errors, ε_i. Often, we are not too sure about the distribution of the errors and would like to relax this assumption to the assumptions in (2). The traditional approach is to use asymptotic theory. It is well known that if n is large,

$$\mathcal{L}(b_{OLS} | X = x) \approx N\left(\beta, s^2 (x'x)^{-1}\right).$$

If we are willing to make the assumption that our sample is large enough, we find

$$\mathcal{L}(T_n) \approx N(0, 1)$$

and can base inference on this distribution. But how do we know when n is large enough? There is no clear-cut answer to this question. Usually, we design our test statistic such that it rejects a correct null hypothesis with probability α[4]. Since, the distribution of the test statistic is only approximate, it could be the case that a true null hypothesis is rejected far more often or less than wanted. Therefore, we would like to get a better approximation of $\mathcal{L}(T_n)$. Suppose, that instead of just one realization of $\{(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)\}$ we have N. This gives us N realizations of the test statistic as well. Now, if N is large, we get a good approximation of the distribution of T_n. It would be therefore be very useful to have more realizations of $\{(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)\}$. Since F_0 is unknown, we cannot draw from it. What we can do is draw n times from our initial data with replacement and create artificial realizations of size n. This is basically the bootstrap. In Section 6.2, we see that in this example the bootstrap provides us with a better approximation of the finite sample distribution of the test statistic than asymptotic theory. Before we get there, we first look at the general problem trying to understand why the bootstrap provides us with such a better approximation. In the next section, we take a look at the more general problem.

[3] $\mathcal{L}(X)$ denotes the law (or probability distribution) of X.

[4] $\alpha = 0.05$ is most often used.

3 General problem

Assume that the data set $\{x_1; x_2; \ldots; x_n\}$ that we have available is a realization of a random sample $\{X_1; X_2; \ldots; X_n\}$ from a probability distribution with cumulative distribution function (CDF) F_0. F_0 may belong to a finite- or infinite-dimensional family $\mathcal{F}$. If F_0 belongs to a finite-dimensional family $\mathcal{F}(\theta)$ indexed by a parameter θ whose population value is θ_0, we write $F_0(x;\theta_0) = \mathbb{P}_{\theta_0}(X_i \leq x)$[5] and $F(x;\theta) = \mathbb{P}_{\theta}(X_i \leq x)$ for a general member of $\mathcal{F}(\theta)$. Let $T_n = T_n(X_1; X_2; \ldots; X_n)$ denote the test statistic of interest for testing a hypothesis about a population parameter θ or a function $r(\theta)$ of θ. Let $G_n(t; F_0) \equiv \mathbb{P}_{\theta_0}(T_n \leq t)$ denote the exact finite-sample CDF of T_n under the null hypothesis H_0.

The α-level critical $z_{n\alpha}$ value solves

$$G_n(z_{n\alpha}; F) = 1 - \alpha \tag{7}$$

in the case of a one-sided test and

$$G_n(z_{n\alpha}; F) - G_n(-z_{n\alpha}; F) = 1 - \alpha \tag{8}$$

in the case of a two-sided test. In the remainder of the paper, we concentrate on the one-sided test, since this illustrates the virtues of the bootstrap without loss of generality. Furthermore, the theory of the two-sided test is easily recovered from the theory of the one-sided test.

In most applications, $G_n(\cdot; F)$ depends on F, which is unknown, and therefore (7) cannot be solved. In some special cases, $G_n(\cdot; F)$ does not depend on F, in which case we call T_n a pivotal statistic. A well-known example is the usual t-statistic corresponding to the estimate of a mean from a standard normal distribution. F is a normal distribution and from this is it follows that G_n has a t-distribution with $n - 1$ degrees of freedom. Pivots typically exist under strong distributional assumptions (to be more precise, you have to make distributional assumptions in order to obtain a pivot). If T_n is not pivotal (the usual case), we cannot solve (7). Therefore, we have to find an approximation for $z_{n\alpha}$. One way to handle the problem of approximating the critical values of the test-statistic T_n is resorting to asymptotic distribution theory. An alternative approach, the bootstrap, has been introduced by Efron (1979). First, we discuss the traditional approach using asymptotic theory. Second, we investigate the bootstrap. Finally, we compare the two methods.

[5] When it is clear from the context which measure is used, we will supress the subscripts.

3.1 Asymptotic distribution theory

Though few statistics are pivotal, most statistics used in econometrics are asymptotically pivotal. A statistic is asymptotically pivotal if its asymptotic distribution does not depend on unknown population parameters. This distribution is often the standard normal distribution or χ^2_2 distribution. More formally, let

$$G_\infty(t;F) \equiv \lim_{n\to\infty} G_n(t;F). \tag{9}$$

If T_n is asymptotically pivotal,we have

$$G_\infty(t;F) = G_\infty(t). \tag{10}$$

If n is sufficiently large, $G_\infty(t;F)$ is arbitrarily close to $G_n(t;F)$ as can be seen from (9). Furthermore, (10) states that G_∞ does not depend on F if T_n is asymptotically pivotal, so the unknown $G_n(t;F)$ can be approximated by $G_\infty(t)$. We can now get approximate solutions for (7) by replacing it with

$$G_\infty(z_{\infty\alpha}) = 1 - \alpha. \tag{11}$$

For a given level α, $G_\infty(\cdot)$ can be inverted and thereby the asymptotic critical value $z_{\infty\alpha}$ can be recovered to serve as an approximation of $z_{n\alpha}$.

3.2 The bootstrap

The bootstrap provides an alternative approximation to the finite-sample distribution of a test statistic T_n. Where first-order asymptotic approximations replaces the unknown distribution function G_n by G_∞ the bootstrap replaces the unknown CDF F_0 with a known estimator which we denote F_n. Two possible estimators F_n are:

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

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{(-\infty,x)}(X_i) \tag{12}$$

where $\mathbf{I}_{(A)}(x)$ denotes an indicator function[6].

[6] $\mathbf{I}_{(A)}(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$

- A parametric estimator of F_0: If $F_0(\cdot) = F(\cdot; \theta_0)$ for some finite dimensional θ_0, consistently estimated by θ_n; and $F(\cdot; \theta)$ is a continuous function of θ in the neighborhood of θ_0; then $F(x; \theta_n) \to F(x; \theta_0)$ for all x as $n \to \infty$.

Regardless of the choice of F_n the bootstrap estimator of $G_n(\cdot; F_0)$ is $G_n(\cdot; F_n)$. Usually, $G_n(\cdot; F_n)$ also cannot be evaluated analytically. Using a Monte-Carlo simulation drawing samples from F_n we can, however, estimate it with arbitrary accuracy. Therefore, the bootstrap is usually implemented using a Monte-Carlo simulation. The essential characteristic of the bootstrap is, however, the use of F_n to approximate F_0 in $G_n(\cdot; F_0)$, not the Monte-Carlo method used to evaluate $G_n(\cdot; F_n)$.

The Monte-Carlo procedure for estimating $G_n(t; F_n)$ can be implemented in the following way:

1. Generate L bootstrap samples of size n by sampling the distribution corresponding to F_n randomly. Denote the realizations of the random samples $\{\{X_1^*, X_2^*, \ldots, X_n^*\}_j, j = 1, \ldots, L\}$ by $\{\{x_1^*, x_2^*, \ldots, x_n^*\}_j, j = 1, \ldots, L\}$. If F_n is the EDF of the estimation data set, then the bootstrap sample can be obtained by sampling the estimation data with replacement.
2. Compute the statistics $\left\{T_{n,j}^* \equiv T_{n,j}(x_1^*, x_2^*, \ldots, x_n^*), j = 1, \ldots, L\right\}$.
3. Use the result of many replications of step 1 and 2 to compute the EDF of T_n^* (e.g. the proportion of T_n^* values smaller than t).

The bootstrap version of (7) is

$$G_n(z_{n\alpha}^*; F_n) = 1 - \alpha. \tag{13}$$

Inverting the EDF, F_n, of T_n^* we get the bootstrap critical value $z_{n\alpha}^*$ which serves as the bootstrap approximation for $z_{n\alpha}$.

3.3 Asymptotic theory vs the bootstrap

Recollecting the results of the two subsections above; we found two different approximations of the unknown distribution $G_n(t; F_0)$:

1. $G_\infty(t; F_0)$ by using asymptotic theory.

2. $G_n(t; F_n)$ by using the bootstrap.

This gives rise to two natural questions:

- What is the relation between $G_\infty(t; F_0)$ and $G_n(t; F_n)$?
- What is the relation between $G_\infty(t; F_0)$, $G_n(t; F_n)$ and $G_n(t; F_0)$?

For n sufficiently large $G_\infty(t; F_0)$ and $G_n(t; F_0)$ will be close. For the bootstrap, we need in addition n to be large in order for F_n to be close to F. If $G_n(\cdot; F_n)$ and $G_\infty(\cdot; F_n)$ are "continuous" functions in their second arguments near F, we expect that $G_n(t; F_n) - \!\!\to G_\infty(t; F_0)$ a.s. as $n \to \infty$: This preferable property is roughly speaking consistency of the bootstrap. In the next section, consistency is defined more precise.

4 Theory of the bootstrap

In the previous section, we discussed the main idea of the bootstrap and its implementation. In this section, we become a bit more formal and discuss the theory underlying the bootstrap. We start by defining the concept of consistency, which was already informally touched upon in the last section. If we know that the bootstrap is consistent, we can apply it and thereby avoid the often complicated first-order asymptotic mathematics. This is the first advantage of the bootstrap. Furthermore, we investigate the accuracy of the bootstrap. We show that using the bootstrap we can get results that are better than using first-order asymptotic theory. This shows the second advantage of the bootstrap.

4.1 Consistency of the bootstrap

Since $G_n(t; F_0)$ is the distribution of interest and we use $G_n(t; F_n)$ as an approximation for it, we are interested in the difference $|G_n(t; F_0) - G_n(t; F_n)|$. A natural requirement is to have that $|G_n(t; F_0) - G_n(t; F_n)|$ is small if n is large. In order to investigate this, we need a metric (measure of distance) and a limiting concept as $n \to \infty$. As metric for the "distance" between two distributions G_1 and G_2, we use

$$d(G_1; G_2) \equiv \sup_t |G_1(t) - G_2(t)|$$

So, the distance between $G_n(t; F_n)$ and $G_\infty(t; F)$ is defined as $\sup_t |G_n(t; F_n) - G_\infty(t; F)|$: We can now define consistency.

Definition 1 Let $\mathbb{P}_n$ denote the joint distribution function of a random sample $\{X_1, X_2, \ldots, X_n\}$. The bootstrap estimator $G_n(\cdot, F_n)$ is said to be consistent if for every $\varepsilon > 0$ and each $F_0 \in \mathcal{F}$

$$\lim_{n \to \infty} \mathbb{P}_n(\sup_t |G_n(t, F_n) - G_\infty(t, F)| > \varepsilon) = 0 \tag{14}$$

The metric for the space of distributions $\rho_n = d(F_n, F_0)$ is the "distance" between F_n and F_0.

It is rather difficult to check consistency straight from the definition. A useful theorem with sufficient conditions for consistency is provided by Beran and Ducharme (1991).

Theorem 2 $G_n(\cdot, F_n)$ is consistent if for any $\varepsilon > 0$ and $F_0 \in \mathcal{F}$:

1. $\mathbb{P}_n(\rho(F_n, F_0) > \varepsilon) \to 0$ as $n \to \infty$ (note: condition on F and not G)
2. $G_\infty(t, F)$ is a continuous function of t for each F.
3. For any t and any sequence of distribution functions H_n such that $\rho(H_n, F) \to 0$ and $G_n(t, H_n) \to G_\infty(t, F)$ as $n \to \infty$. This condition is called the "triangular array convergence condition".

The theorem by Beran and Ducharme (1991) gives sufficient conditions for consistency, not necessary conditions. The next theorem by Mammen (1992) gives necessary and sufficient conditions for consistency, but for a smaller class of test statistics, namely the class of linear test statistics.

Theorem 3 Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a population. For a sequence of functions g_n and sequences of numbers t_n and σ_n, define $\bar{g}_n = \frac{1}{n}\sum_{i=1} g_n(X_i)$ and $T_n = (\bar{g}_n - t_n)/\sigma_n$. For the bootstrap sample $\{X_1^*, X_2^*, \ldots, X_n^*\}$ define $\bar{g}_n^* = \frac{1}{n}\sum_{i=1} g_n(X_i^*)$ and $T_n^* = (\bar{g}_n^* - \bar{g}_n)/\sigma_n$. Let $G_n(\tau) = \mathbb{P}(T_n \leq \tau)$ and $G_n^*(\tau) = \mathbb{P}^*(T_n^* \leq \tau)$. Where $\mathbb{P}^*$ is the probability distribution induced by bootstrap sampling. Then, G_n^* consistently estimates G_n if and only if $T_n \xrightarrow{d} N(0, 1)$.

To see that the restriction to linear test statistics is not that restrictive, we illustrate an application of Mammen's theorem via an $M-$estimator (see e.g. Van der Vaart, 1998). θ is called an $M-$estimator if it is a solution to

$$\max_\theta \frac{1}{n} \sum_{i=1} q(X_i, \theta) = \max_\theta Q_n(\theta)$$

FOC :

$$
\begin{aligned}
\frac{\partial}{\partial\theta} Q_n(\theta_n) &= 0 \\
\frac{\partial}{\partial\theta} Q_n(\theta_n) &= \frac{\partial}{\partial\theta} Q_n(\theta_0) + \frac{\partial}{\partial\theta\partial\theta'} Q_n\left(\tilde{\theta}\right)(\theta_n - \theta_0)
\end{aligned}
\tag{15}
$$

with $\tilde{\theta}$ between θ_0 and θ_n $\Rightarrow$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1} \frac{\partial q(X_i;\theta_0)}{\partial\theta} &= \frac{1}{n}\sum_{i=1} \frac{\partial^2 q(X_i;\tilde{\theta})}{\partial\theta\partial\theta'}(\theta_n - \theta_0) \\
\frac{1}{\sqrt{n}}\sum_{i=1} \frac{\partial q(X_i;\theta_0)}{\partial\theta} &= \frac{1}{n}\sum_{i=1} \frac{\partial^2 q(X_i;\tilde{\theta})}{\partial\theta\partial\theta'}\sqrt{n}(\theta_n - \theta_0)
\end{aligned}
\tag{16}
$$

Since

$$
\frac{1}{n}\sum_{i=1} \frac{\partial^2 q(X_i;\tilde{\theta})}{\partial\theta\partial\theta'} \to A \text{ (non-stochastic) a.s.} \tag{17}
$$

we get

$$
\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1} A^{-1}\frac{\partial q(X_i;\theta_0)}{\partial\theta} + o\left(n^{-1/2}\right) \tag{18}
$$

We see that an $M-$estimators is asymptotically linear and therefore Mammen's theorem can be applied. Special cases of $M-$estimators are OLS ($q(x) = x^2$); Generalized Method of Moments (GMM), and Maximum Likelihood (ML), and Mammen's theorem can be applied to all these estimators.

Though the bootstrap is consistent in many often used situations, it can be inconsistent. A few examples of bootstrap inconsistency are

- Heavy tailed distributions

  Suppose X has a Cauchy distribution

  Define $\frac{1}{n}\sum_{i=1} X_i \xrightarrow{d} Z$, which is Cauchy distributed. The Cauchy distribution has no mean, therefore, we cannot use the Central Limit Theorem (CLT) or even the law of large numbers.

- Distribution of the sample average squared $\left(\overline{X}\right)^2$

  $\overline{X} = \frac{1}{n}\sum_{i=1} X_i; \qquad \mathbb{E}X_i = \mu$

  Now, distinguish two cases

– $\mu \neq 0$:

$$
\begin{aligned}
\left(\overline{X}\right)^2 &= \left[\frac{1}{n}\sum_{i=1}(X_i - \mu) + \mu\right]^2 \\
&= \mu^2\left[1 + \frac{1}{n\mu}\sum_{i=1}(X_i - \mu)\right]^2 \\
&= \mu^2\left[1 + \frac{2}{n\mu}\sum_{i=1}(X_i - \mu) + \left(\frac{1}{n\mu}\sum_{i=1}(X_i - \mu)\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{n}\left(\left(\overline{X}\right)^2 - \mu^2\right) &= \frac{2}{\mu\sqrt{n}}\sum_{i=1}(X_i - \mu) + \sqrt{n}\left(\frac{1}{n\mu}\sum_{i=1}(X_i - \mu)\right)^2 \\
&\xrightarrow{d} N(0;1)
\end{aligned}
$$

– $\mu = 0$:

$$
\begin{aligned}
L\left(n\left(\frac{1}{n}\sum_{i=1}X_i\right)^2\right) &= L\left(\frac{1}{n}\left(\sum_{i=1}X_i\right)^2\right) \\
&= \frac{1}{n}\left[N\left(0;\sqrt{n}\right)\right]^2 \\
&= \chi_1^2
\end{aligned}
$$

The second case violates the 3rd condition of Beran and Ducharme's (1991) theorem. This does not necessarily imply that the bootstrap is inconsistent at $\mu = 0$; but it can be shown that the bootstrap is inconsistent in this particular case.

Other examples are the estimation of a parameter on the boundary of the parameter space, the estimation of the maximum, and an AR(1) model with a unit root.

What can we do when the bootstrap is inconsistent? We can use the subsampling procedure introduced by Politis and Romano (1994a). Let the sample size be n. Now, let $m < n$. The statistic $T_n(X_1, \ldots, X_n) = \rho(n) \cdot [t_n(X_1, \ldots, X_n) - \mu]$. E.g. if $T_n = \sqrt{n}(\overline{X} - \mu)$ then $\rho(n) = \sqrt{n}$, $t_n = \frac{1}{n}\sum_{i=1} X_i$, and $\mu = \mathbb{E}(X_i)$. Suppose T_n has a non-degenerate asymptotic distribution. Consider subsets of the data, consisting of m observations. There are $\binom{n}{m} = N_m$ different subsets. Define the subset estimator for $G_n(\tau, F)$:

$$
G_{nm}(\tau) = \frac{1}{N_m}\sum_{k=1}^{N_m} 1\left[\rho(n) \cdot [t_{nm} - t_n] \leq \tau\right]; \tag{19}
$$

where t_{nm} denotes the version of t_n for the subset.

Now, from Politis and Romano (1994a) it follows: If T_n has a nondegenerate asymptotic distribution, then G_{nm} estimates G_∞ consistently if $m \to \infty$; $n \to \infty$; and $\frac{m}{n} \to 0$.

If subsampling is always consistent, why use the bootstrap at all? When the bootstrap is consistent, the bootstrap estimator of $G_\infty(t;F)$ and $G_n(t;F)$ is (much) more accurate than the subsampling estimator. So, it is not wise to use subsampling when the regular bootstrap estimator is consistent.

4.2 Accuracy of the bootstrap

In the previous section, we defined the concept of consistency and some described conditions under which the bootstrap yields a consistent estimator of the distribution of the distribution of a statistic. This means, roughly speaking, that the bootstrap gets the asymptotic distribution right, at least if the sample size is large enough. Therefore we can use the bootstrap when we know that it is consistent and thereby avoid the often complicated mathematics of asymptotic theory. The bootstrap, however, often does more than just get the asymptotic distribution right. In a large number of relevant cases in econometrics, it provides us with a higher-order asymptotic approximation to the distribution of a statistic.

4.2.1 Bias reduction

We start by showing how the bootstrap can provide a reduction in the bias of an estimator. Consider a random variable X, with $\mathbb{E}(X) = \mu$. We want to estimate $\theta_0 = g(\mu)$, with g a known function (and μ an unknown parameter). We estimate μ by $\overline{X} = \frac{1}{n}\sum_{i=1} X_i$ and θ_0 by $\theta_n = g(\overline{X})$; which is sometimes called an analogue estimator. By Jensen's inequality, we know that θ_n will be biased for θ if g is a nonlinear function: $\mathbb{E}\theta_n = \mathbb{E}g(\overline{X}) \neq g(\mathbb{E}\overline{X}) = g(\mu) = \theta_0$. Therefore, θ_n is a biased estimator of θ_0.

Some people only focus on minimizing the bias of an estimator. However, from a decision theory perspective, the aim is to minimize the mean squared error of an estimator. As we know, the mean squared error is the sum of the bias squared and the variance of the estimator. Reducing the bias with increasing the variance may therefore not be beneficial for the mean squared error of the estimator. We will use the bias reduction argument here not for the sake of reducing the bias, but it is one of the two means to reduce the mean squared error. To see how the bootstrap can reduce the bias of θ_n assume that g is four times continuously differentiable and that the X has finite fourth absolute moments. A Taylor expansion of $\theta_n = g(\overline{X})$ gives

$$\begin{aligned} \theta_n - \theta_0 &= g(\overline{X}) - g(\mu) && (20)\\ &= g'(\mu)(\overline{X} - \mu) + \frac{1}{2}g''(\mu)(\overline{X} - \mu)^2 + R_n; \end{aligned}$$

where R_n is a remainder term that satisfies $\mathbb{E}(R_n) = o(n^{-2})$. Taking expectations on both sides of (20) results in

$$\mathbb{E}(\theta_n - \theta_0) = \frac{1}{2}\mathbb{E}g''(\mu)(\overline{X} - \mu)^2 + o\left(n^{-2}\right) \tag{21}$$

The first term on the right hand side of (21) has size $o(n^{-1})$. Therefore, through $o(n^{-1})$ the bias B_n of θ_n is

$$B_n = \frac{1}{2}\mathbb{E}g''(\mu)(\overline{X} - \mu)^2 \tag{22}$$

Now let us consider the bootstrap. With the bootstrap we sample the EDF of the data. Let $\{X_1^*, X_2^*, \ldots, X_n^*\}$ be the bootstrap sample that is obtained in this way. Define $\overline{X}^* = \frac{1}{n}\sum_{i=1} X_i^*$. The bootstrap estimator of θ is $g(\overline{X}^*) = \theta_n^*$, while the bootstrap "truth" is $g(\overline{X}) = \theta_n$. The bootstrap analog of (20) is

$$\begin{aligned} \theta_n^* - \theta_n &= g(\overline{X}) - g(\mu) && (23)\\ &= g'(\overline{X})(\overline{X}^* - \overline{X}) + \frac{1}{2}g''(\overline{X})(\overline{X}^* - \overline{X})^2 + R_n^*; \end{aligned}$$

where R_n^* is the bootstrap remainder term. Let $\mathbb{E}^*$ denote the expectation with respect to the EDF of the estimation data. Taking $\mathbb{E}^*$ expectations of (23) gives

$$B_n^* = \frac{1}{2}\mathbb{E}^*g''(\mu)(\overline{X}^* - \overline{X})^2 + o\left(n^{-2}\right) \quad \text{a.s.} \tag{24}$$

By a Monte Carlo simulation we can compute B_n^* with arbitrary accuracy, since the distribution of the bootstrap samples is known. This Monte Carlo procedure is as follows

1. Use estimation data to compute θ_n.
2. Generate a bootstrap sample of size n by sampling data randomly with replacement. Compute $\theta_n^* = g\left(\overline{X}^*\right)$.
3. Compute $\mathbb{E}^*(\theta_n^*)$ by averaging results of many repetitions of step 2. Set $B_n^* = \mathbb{E}^*(\theta_n^*) - \theta_n$.

We have

$$\mathbb{E}(B_n^*) = B_n + o\left(n^{-2}\right)$$

This implies that if we use a modified estimator of θ_0; $\tilde{\theta}_n = \theta_n - B_n^*$ we can get a bias reduction from $o(n^{-1})$ to $o(n^{-2})$, since

$$\begin{aligned}
\mathbb{E}\left(\tilde{\theta}_n - \theta\right) &= \mathbb{E}(\theta_n - \theta) - \mathbb{E}(B_n^*) \\
&= B_n - B_n + o\left(n^{-2}\right) \\
&= o\left(n^{-2}\right).
\end{aligned}$$

Note that, we can calculate B_n^* as accurate since we control the number of repetitions of the Monte Carlo experiment. We illustrate the theory with an example in Section 6.1.

4.2.2 Distribution of statistics

Though we have shown that the bootstrap is consistent, we still haven't discussed the difference between $G_n(t; F_n)$ and $G_n(t; F_0)$. This can be done using Edgeworth expansions. For a rigorous treatment, this involves some tedious mathematics that will not be presented here. Instead we only provide a small portion of the mathematics that should provide the necessary insights. Consider

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1} X_i; \ \mathbb{E}X_i = 0; \ \mathrm{Var}X_i = 1 \tag{25}$$

Note that the restrictions on the mean and variance are mild, since by standardization we can transform most statistics in this way. Using the CLT, we know that $\lim_{n \to \infty} \mathbb{P}(S_n \le z) = \Phi(z)$; where $\Phi(\cdot)$ is the cumulative normal distribution function. What can we say about the difference $\mathbb{P}(S_n \le z) - \Phi(z)$? Under some regularity conditions on G_n we find that

$$\mathbb{P}(S_n \le z) = \Phi(z) + \frac{1}{\sqrt{n}} g_1(z) + \frac{1}{n} g_2(z) + \dots + n^{-j/2} g_j(z) + o\left(n^{-(j+1)/2}\right) \tag{26}$$

where $g_j(\cdot)$ is an even function of t of j is odd and an odd function of t when j is even[7]. After some calculations using Fourier inversions, we get

[7] A function is called even if $f(x) = f(-x)$ and is called odd if $f(x) = -f(-x)$.

$$p_n(z) = \phi(z)\left(1 + \frac{1}{6}\mathbb{E}\left(X^3\right)\frac{z^3 - 3z}{\sqrt{n}}\right). \tag{27}$$

For $n \to \infty$ we get the standard normal density function, since the second term between brackets goes to zero. For finite samples, however, we find an improvement taking third moments into account.

Assume the statistic of interest is a smooth function of the sample moments

$$T_n = \frac{\sqrt{n}(H(\overline{X}) - H(\mu))}{s} \tag{28}$$

where s^2 is a consistent estimator of the asymptotic variance of $\sqrt{n}(H(\overline{X}) - H(\mu))$. Generally, this will be unknown, if it is known (e.g. σ^2); we use σ^2 instead. Some examples of functions H that cannot be used are the least absolute deviation (not sufficiently smooth), and nonparametric kernel density and regression estimators (depends on sample size). Under some regularity conditions, we find

$$G_n(t;F) = G_\infty(t;F) + \frac{1}{\sqrt{n}}g_1(t;F) + \frac{1}{n}g_2(t;F) + \frac{1}{n\sqrt{n}}g_3(t;F) + o\left(n^{-2}\right) \tag{29}$$

where $g_j(t;F)$ is an even function of t if j is odd and an odd function if j is even. Also, g_j is a continuous functional of its second argument F_0. We can do the same expansion for the bootstrap estimator

$$G_n(t;F_n) = G_\infty(t;F_n) + \frac{1}{\sqrt{n}}g_1(t;F_n) + \frac{1}{n}g_2(t;F_n) + \frac{1}{n\sqrt{n}}g_3(t;F_n) + o\left(n^{-2}\right) \tag{30}$$

The error made by the bootstrap estimator is then

$$\begin{aligned} G_n(t;F_n) - G_n(t;F) &= G_\infty(t;F) - G_\infty(t;F_n) \\ &\quad + \frac{1}{\sqrt{n}}(g_1(t;F_n) - g_1(t;F)) + \frac{1}{n}(g_2(t;F_n) - g_2(t;F)) + o\left(n^{-3/2}\right). \end{aligned} \tag{31}$$

The leading term of the error made by the bootstrap is $G_\infty(t;F) - G_\infty(t;F_n)$. We know that $\rho(F_n;F) = o_p(n^{-\frac{1}{2}})$, therefore the bootstrap estimator and the asymptotic approximation make an equally large error. If the test statistic is, however, asymptotically pivotal, i.e. $G_\infty(t;F) = G_\infty(t)$, the leading term of (31) equals zero. Then the new leading term of the bootstrap error becomes $\frac{1}{\sqrt{n}}(g_1(t;F_n) - g_1(t;F))$ which is $o_p(n^{-1})$. Thus, in the case of asymptotically pivotal statistics the bootstrap is more accurate than a first-order asymptotic approximation. We call this an asymptotic refinement. It can be shown that the bootstrap becomes even better if the distribution is symmetric, since then the functions g_2 and g_4 drop out. The leading term is then of order $O\left(n^{-3/2}\right)$. We may therefore conclude that the bootstrap is more accurate that first order asymptotic approximation (e.g. asymptotic normal distribution) for estimating the distribution of an asymptotically pivotal statistic.

How do we implement the bootstrap for computing critical values? We have as statistic

$$T_n = \sqrt{n}\,(\theta_n - \theta_0)/s_n$$

where θ_n is the parameter estimator, θ_0 hypothetical value ($H_0 : \theta = \theta_0$), and s_n the standard error of asymptotic distribution of $\sqrt{n}\,(\theta_n - \theta_0)$.

The Monte Carlo procedure for computing bootstrap critical values.

1. Use estimation data to compute θ_n.

2. Generate a bootstrap sample of size n by sampling the distribution corresponding to F_n. (E.g. F_n is the EDF of the data, draw the data randomly with replacement.) Compute estimators θ_n^* and s_n^* from the bootstrap sample. Form the bootstrap test statistic $T_n^* = \sqrt{n}\,(\theta_n^* - \theta_n)/s_n^*$.

   Note: We do not center the bootstrap statistic around θ_0. This does not converge to zero, since the null hypothesis concerns the population, and not the data. (There is a zero probability of obtaining $\theta_n = \theta_0$, but we like to test whether they are not "too far apart".)

3. Use the results of many repetitions of step 2 to compute the empirical distribution of $|T_n^*|$ (assuming a symmetrical test here). Set z_α^* equal to the $(1 - \alpha)$-quantile of this distribution.

Example: Construct a nominal 95% confidence interval for $\theta = \exp\{\mu\}$, when $X \sim N(0; 6)$. The true value $\theta_0 = 1$. Now, $\theta_n = \exp\{\bar{x}\}$, and we take $n = 10$. The procedure is as follows:

1. Generate estimation data set of size $n = 10$ by sampling $N(0; 6)$.

2. Compute z_α^* by following steps 1–3 from the previous procedure.

3. Find the empirical coverage probabilities of the bootstrap and asymptotic confidence intervals by repeating the steps 1 and 2 of this procedure many times (e.g. 1000).

In Horowitz (1999), Horowitz finds with a nominal coverage probability of 95%, 94.3% for the bootstrap and 88.5% for the asymptotic approximation. In Section 6.2 we perform a simulation study for the classical regression model.

5 Selected topics

In this section we discuss some miscellaneous topics concerning the bootstrap. First, we discuss the topic of recentering in the context of GMM. Second, we look at the bootstrap in case of time series. Finally, we discuss some cases where the bootstrap flops.

5.1 Recentering

An often used econometric technique these days is the Generalized Method of Moments (GMM) estimation introduced by Hansen (1982). This happens to be a technique where the bootstrap does not necessarily provide an asymptotic refinement. To see why, let θ_0 be the true value of a parameter θ that is identified by the moment conditions $\mathbf{E}h(X;\theta) = 0$, where h is a known function. The dimension of h is $k \times 1$; and the dimension of θ is $q \times 1$; where $k \geq q$. We might want to estimate θ by solving $\frac{1}{n}\sum_{i=1} h(X_i;\theta) = 0$. In the case that $k = q$, this means that θ is exactly identified. There is no exact solution to these set of restriction when $k > q$; therefore we need to solve

$$\min_{\theta} \sum_{i=1} h(X_i;\theta)' \Omega_n \sum_{i=1} h(X_i;\theta). \tag{32}$$

where Ω_n is a weighting matrix. The solution to this problem is the GMM estimator, say θ_n. A natural way to proceed, would be to use the following bootstrap version of (32)

$$\min_{\theta} \sum_{i=1} h(X_i^*;\theta)' \Omega_n \sum_{i=1} h(X_i^*;\theta). \tag{33}$$

This would, however, be naive. The true value in the bootstrap world is namely, θ_n, not θ_0. This implicitly assumes that the moment conditions are exactly zero, but in the case of overidentifying restrictions the moment restrictions are not exactly zero.

$$\frac{1}{n}\sum_{i=1} h(X_i;\theta) \neq 0. \tag{34}$$

The problem can be resolved by recentering. Introduce

$$h^*(X_i;\theta) = h(X_i;\theta) - \frac{1}{n}\sum_{i=1} h(X_i;\theta) \tag{35}$$

We now have a new set of bootstrap moment restrictions

$$\mathbb{E}h^*(X_i;\theta) = 0. \tag{36}$$

This is the set of moment conditions we should use to get a bootstrap estimator θ_n^*. An insightful example is the classical regression without intercept.

Consider a linear model with fixed design (X deterministic) and no intercept, i.e.

$$Y_i = \beta X_i + \varepsilon_i$$

and $\mathbb{E}(\varepsilon_i) = 0$; $\mathrm{Var}(\varepsilon_i) = \sigma^2$. Now, estimate β by OLS to get estimator b and residuals $e_i = Y_i - bX_i$. We would like to use the bootstrap model $Y_i^* = \beta X_i^* + \varepsilon_i^*$ with $\mathbb{E}(\varepsilon_i^*) = 0$ and $\mathrm{Var}(\varepsilon_i^*) = \sigma^{*2}$. Since there is no constant term, the residuals do not necessarily sum to zero. Therefore, in the bootstrap model $\mathbb{E}(\varepsilon_i^*) \neq 0$. If we recenter the residuals, i.e. $\tilde{\varepsilon}_i^* = \varepsilon_i^* - \frac{1}{n}\sum_{i=1} e_i$ we get the model $Y_i^* = \beta X_i^* + \tilde{\varepsilon}_i^*$ with $\mathbb{E}(\tilde{\varepsilon}_i^*) = 0$ and $\mathrm{Var}(\tilde{\varepsilon}_i^*) = \sigma^{*2}$ which allows us to use the bootstrap.

5.2 Time series data

So far, we considered cases in which the data were i.i.d. Here we investigate the case where the data are dependent. Bootstrap sampling must then be done in a way that captures the dependence structure of the data. We consider two cases:

- If there is a parametric model available that reduces the DGP to independent random sampling, then i.i.d. results continue to hold. Consider, for example, an ARMA model

$$A(L;\theta_0)Y_t = B(L;\psi_0)\varepsilon_t \tag{37}$$

  where L denotes the lag operator and $\theta_0;\psi_0$ parameters. We estimate θ_0 by θ_n, ψ_0 by ψ_n and obtain the residuals $e_t = (B(L;\psi_n))^{-1}(A(L;\theta_n)Y_t$. As we saw in the previous section, we want residuals that are zero on average in the bootstrap world, therefore $e_t^* = e_i - \frac{1}{n}\sum_{i=1} e_i$ can be used. Now we can generate bootstrap series by simulating $Y_t^* = ((A(L;\theta_n))^{-1}B(L;\psi_n)e_t^*$.

- If there is no parametric model available, sampling in blocks is an alternative. This can be done by either non-overlapping blocks or overlapping blocks. Consider, for example, a time series with T observations $\{Y_1, \ldots, Y_T\}$. We can take blocks of length l, in the case of non-overlapping blocks block 1 is $\{Y_j, Y_{j+1}\}$; block 2 is $\{Y_{j+2}, Y_{j+3}\}$; etc. With overlapping blocks block 1 is $\{Y_j, Y_{j+1}\}$; block 2 is $\{Y_{j+1}, Y_{j+2}\}$; etc. The bootstrap sample is obtained by drawing blocks with replacement and laying them end-to-end in the order sampled. A more advanced method is the stationary bootstrap by Politis and Romano (1994b). In this case we use overlapping blocks with lengths drawn from the geometric distribution.

  Though the block bootstrap provides some asymptotic refinements, it is less attractive that in the i.i.d. case.

  – With the block bootstrap and a symmetrical confidence interval test we obtain $\mathbb{P}\left(|T_n| > z^*_{n\alpha/2}\right) = \alpha + o\left(n^{-\frac{4}{3}}\right)$. The bias is thus of order $o\left(n^{-\frac{4}{3}}\right)$. If we compare this with the i.i.d. case where we had $o(n^{-2})$, we see that we lose quite some efficiency gain, since the asymptotic case already yields $o(n^{-1})$.

  – We must use special forms of the test statistics, we need correction factors to account for the distortion caused by blocking. The reason for this is that time series relations are deterministic when blocks are formed.

5.3 The bootstrap can be a flop

Though the bootstrap often provides us with improved results, it should not be used blindly. As we saw, the bootstrap only leads to asymptotic refinements if a test statistic is asymptotically pivotal, if not, first-order asymptotic distribution theory approximations can be used just as well. It must be said that the bootstrap is still useful in this case if the first order asymptotic distribution theory approximation is difficult to compute. There are, however, also situation when the bootstrap should not be used:

1. If it is inconsistent, we saw examples of this in Section 4.1.

2. Instrumental variables estimation with weak instruments. Consider the model, $Y = X\beta + \varepsilon$; instrument Z; this results in IV estimator

$$b_n^{IV} = \frac{Z'Y}{Z'X} \tag{38}$$

If we have weak instruments, $\mathbb{E}(Z'X) \approx 0.$ Therefore, in small samples, $Z'X$ may bounce around zero and consequently b_n^{IV} may be a very large positive or negative number.

6 Applications

In this section, we implemented some of the bootstrap theory presented above. First, we apply the bootstrap for bias reduction of an estimator as discussed in Section 4.2.1. Second, we look at the classical regression model we discussed in Section 2[8].

6.1 Bias reduction

In Section 4.2.1 we discussed the issue of bias reduction. Here we present an example to illustrate this. The presented example is an extended version of Example 3.2 in Horowitz (1999). We have $X \sim N(0, 6)$; $n = 10$; $\theta = g(\mu) = \exp\{\mu\}$; $\theta_0 = 1$; and $\theta_n = \exp\left(\overline{X}\right).$ We perform the following procedure:

1. Generate estimation data set of size $n = 10$ by sampling the normal distribution and use this data to compute $\theta_n.$
2. Compute B_n^* and form $\tilde{\theta}_n = \theta_n - B_n^*.$
3. Estimate $\mathbb{E}(\theta_n - \theta_0)$ and $\mathbb{E}(\theta_n - B_n^* - \theta_0)$ by averaging results of many replications of steps 1 and 2.

In Tables 1 and 2 we see that the bootstrap provides high improvements on the MSE when the sample size is small. If we increase the sample size to 100, the results are less shocking, but still the bootstrap outperforms the analogue estimator. In Table 2 we see that increasing the number of bootstrap replications from 1000 to 5000 does not really improve our results. We have also checked the robustness of the bootstrap method for other values of σ^2, and found that the bootstrap provides similar improvements here. One is encouraged to perform these experiments for oneself using the program provided.

[8] Ox programs to perform the experiments in this section can be found on the author's web page: http://cwis.kub.nl/~few5/center/phd_stud/kerkhof/

	Horowitz			Gauss			Ox		
	Bias	MSE	Var	Bias	MSE	Var	Bias	MSE	Var
θ_n	0.356	1.994	1.867	0.274	1.149	1.074	0.3280	1.5670	1.6746
$\theta_n - B_n^*$	-0.063	1.246	1.242	-0.126	0.690	0.674	-0.0986	0.8424	0.8521

Table 1: Simulation results for the estimation of $\theta = \exp(\mu)$; with $\theta_n = \exp\left(\overline{X}\right)$. Furthermore, $X \sim N(0; 6)$; and $n = 10$. The first column restates the Table from Horowitz (1999). In the second column, the results from the Gauss program by Horowitz are displayed. The third column shows the results from an Ox program written by the author. In the first row, the analogue estimator is displayed, and in the second row the bootstrap estimator. Number of simulations used is 10,000 and the number of bootstrap replications equals 5,000.

			bootstrap = 5000		bootstrap = 1000	
n	σ^2	rep	Bias	MSE	Bias	MSE
10	6	10; 000	0.3280	1.6746	0.3206	1.5060
			-0.0986	0.8521	-0.0994	0.7717
100	6	10; 000	0.0256	0.0657	0.0254	0.0661
			-0.0053	0.0613	-0.0056	0.0617

Table 2: Results to check the influence of the sample size on the results.

6.2 Confidence intervals

We now return to the example of classical regression that we gave at the start of this paper. We want to test whether certain parameter estimates differ significantly from zero. Usually this is done using a (1-α)% confidence interval and checking whether 0 is in this interval. We now set up a Monte Carlo experiment to check the performance of the bootstrap for the classical regression example. Recall the model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i. \tag{39}$$

with an i.i.d. sequence $\{\varepsilon_i\}$ such that

$$\begin{aligned} \mathbb{E}(\varepsilon_i) &= 0 \\ \mathbb{E}(X_i \varepsilon_i) &= 0 \\ \mathbb{E}\left(\varepsilon_i^2\right) &= \sigma^2. \end{aligned} \tag{40}$$

We simulated random samples of this model with $\beta_0 = \beta_1 = 0$; and 1) $\mathcal{L}(\varepsilon_i) = N(0;1)$ and 2) $\mathcal{L}(\varepsilon_i) = U[-1;1]$. In both cases $\mathcal{L}(X_i) = N(0;1)$. We construct 4 confidence intervals based on

1. Asymptotically critical value with normal standard errors.
2. Asymptotically critical value with bootstrap standard errors.
3. Bootstrap critical value with normal standard errors.
4. Bootstrap critical value with bootstrap standard errors.

In the case where $\mathcal{L}(\varepsilon_i) = N(0;1)$; we know the exact confidence interval, since the critical value is known (from the $t-$distribution). The procedure is as follows:

a. Generate a random sample $\{(Y_1;X_1);(Y_2;X_2);\ldots;(Y_n;X_n)\}$.

b. Construct L bootstrap samples: $\{\{(Y_1^*;X_1^*);(Y_2^*;X_2^*);\ldots;(Y_n^*;X_n^*)\}_j;\ j=1;\ldots;L\}$ by drawing pairs $(Y_i;X_i)$ with replacement from the data.

c. Calculate the OLS estimators $\left\{b_j^*;\ j=1;\ldots;L\right\}$ for the bootstrap samples.

d. Compute bootstrap standard errors $\left\{\sigma_j^* = \sqrt{\operatorname{diag}\left(\operatorname{Var}\left(b_j^*\right)\right)};\ j=1;\ldots;L\right\}$.

e. Compute bootstrap critical values $\left\{z_{n\alpha;j}^*;\ j=1;\ldots;L\right\}$.

	A	B	A^*	B^*
intercept	0.901	0.960	0.917	0.964
slope	0.903	0.940	0.937	0.951

Table 3: Coverage ratios based on 4 types of confidence intervals for the case with $L(\varepsilon_i) = N(0;1)$; A : asymptotic critical value with normal standard errors. A^* : asymptotic critical value with bootstrapped critical value. B : bootstrap critical value with normal standard errors. B^* : bootstrap critical value with bootstrapped critical value.

	A	B	A^*	B^*
intercept	0.896	0.959	0.909	0.962
slope	0.901	0.928	0.930	0.940

Table 4: Coverage ratios based on 4 types of confidence intervals for the case with $L(\varepsilon_i) = U[-1;1]$; A : asymptotic critical value with normal standard errors. A^* : asymptotic critical value with bootstrapped critical value. B : bootstrap critical value with normal standard errors. B^* : bootstrap critical value with bootstrapped critical value.

f. Construct confidence intervals and check whether the original parameter estimates are within the confidence interval:

We see in Tables that based on the asymptotic critical value (both using normal standard errors as using bootstrapped standard errors), we reject a null hypothesis too often. The coverage ratio in the case of bootstrap critical values is much closer to the wanted 95%. The results are approximately the same for the both case 1) $L(\varepsilon_i) = N(0;1)$ and 2) $L(\varepsilon_i) = U[-1;1]$. We may therefore conclude that as indicated by the theory, the bootstrap provides us with a better coverage ratio if the we base our decision on the confidence interval using bootstrap critical values. We also see that bootstrap standard errors do not provide us with a better coverage ratio, though it doesn't underperform either.

7 Conclusions

In this paper, we suggest the bootstrap as an alternative to the commonly used first-order asymptotic theory. Though the idea of the bootstrap has been around since Efron (1979), it only became popular the last couple of years. This is not surprising, since for a practical implementation the bootstrap relies on heavy computer power.

In this paper, we started by motivating the use of the bootstrap in probably the most used econometric model, the classical regression. We showed in Section 6.2 that the bootstrap improves upon the asymptotic con¯dence interval for the t-statistics. We continued to treat the general theory behind the bootstrap and its main idea "treating the data as if it were the population". We discussed its properties and illustrated these with examples. To show that the bootstrap also has its de¯ciencies, some examples of bootstrap inconsistencies were provided.

Finally, we performed some simulation studies to investigate the performance of the bootstrap versus that of asymptotic theory. We showed that the bootstrap can provide us with bias reductions of estimators and asymptotic re¯nements of distributions of statistics.Considering the theoretical advantages combined with increasing computer power we can only expect the bootstrap to become a standard tool in econometrics.

References

Beran, R., and G. Ducharme (1991): Asymptotic Theory for Bootstrap Methods in Statistics. Centre de recherches mathematique, Universite de Montreal, Montreal, Canada.

Efron, B. (1979): \Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7, 1{26.

Hansen, L. P. (1982): \Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50, 1029{1054.

Horowitz, J. L. (1999): \The Bootstrap," Handbook of Econometrics, vol. 5, pp. 1{98.

Mammen, E. (1992): When Does the Bootstrap Work? Springer-Verlag, New York.

Politis, D., and J. Romano (1994a): \Large Sample Con¯dence Regions Based on Subsamples under Minimal Assumptions," Annals of Statistics, 22, 2031{2050.

Politis, D., and J. Romano (1994b): \The Stationary Bootstrap," Journal of the American Statistical Association, 89, 1303{1313.