Bootstrap Methods and Applications in Econometrics - a Brief Survey*

Pål Bergström[y]

January 1999

Abstract

This paper provides a brief survey of the bootstrap and its use in econometrics. As an introduction, the paper gives a description of the basics of the method, with a special empasis on bootstrap testing. A fairly large amount of space is devoted to discussing why bootstrap tests provide refinements compared to to equivalent asymptotic tests. A series of different recent applications in the econometrics litterature is then surveyed in order to give a picture of this rapidly evolving research field.

Keywords: Bootstrap; Sample Reuse Methods; Simulation Methods
JEL Classification: C4

*This survey has benefitted from comments from Eva Johansson and Anders Klevmarken, neither of whom can be held responsible for remaining obscurities and errors.

[y]Department of Economics, Uppsala University, Box 513, SE 751 20 Uppsala, Sweden. e-mail: Pal.Bergstrom@nek.uu.se

1 Introduction

This paper gives a brief and maybe a bit subjective survey of the bootstrap and its use in econometrics. Since research in this field has been very active, especially in the last five years or so, I will give a brief introduction to the "basics" of the bootstrap but mainly try to focus on the "frontier" of this rapidly evolving field. I will do this by first saying a few words on the basic ideas and then try to explain how the bootstrap can be applied in the three major contexts, i.e:

- Estimating variance (standard errors)
- Correcting for bias
- Forming tests

Having done that I will dwell upon the subject on as to why the bootstrap actually provides asymptotic refinements a bit longer. I believe that the understanding of these issues really helps telling us if it is a good idea to apply the bootstrap in a specific context. Having gone through these somewhat messy arguments, I will finally describe a selection of applications and also hint at some questions open for future research.

1.1 The Bootstrap

The bootstrap as a computational device was invented and introduced by Efron (1979) as a quite intuitive and (perhaps deceptively) simple way of finding approximations of quantities that are very hard, or even impossible to compute analytically. The basic idea is to take the sample that we are interested in and think of it as if it was a population and then by resampling create a new sample, a bootstrap sample, which we use to compute some quantity that we are interested in. If we repeat this several times, obtaining lots of bootstrap samples, we can use the mean of the computed quantities as an estimate of the expected value of this bootstrapped quantity.

Let us just consider a simple example: Suppose that we, just for the sake of the argument, would want to compute the bootstrap mean for a sample of, say, heights of 30 economists. To calculate the bootstrap mean, we consider the 30 heights we have as the entire population. We then, using some unbiased pseudo random number generator, perform 30 random drawings from our "fake" population, but all the time drawing with replacement, i.e., we put the heights drawn back into the population all the time. Then, almost certainly, some heights will be drawn several times and some not at all. The bootstrap sample that we get will thus di¤er somewhat from the fake population, i.e. the original sample. We now compute the quantity of interest, which in this case was the mean, using the bootstrap sample and obtain one realisation of the bootstrap estimator for the mean. We then repeat this several times, say a thousand, and thus get thousand bootstrap samples and thousand realisations of the bootstrap mean. Computing the mean of these thousand realisations will give us our estimate of the expected value of the bootstrap mean.

Now, why would anyone bother doing this? Well, doing precisely what we just did, we would be suprised indeed if the bootstrap mean deviated from the mean of the original sample. The above example is hence not a particularily interesting one. In many cases however, this simple principle can be used to approxiamte quantities that are very hard to compute analytically. In addition to this, bootstrap quantities can, under certain circumstances, be shown to converge to the true values more rapidly than asymptotic approximations, and also be used to correct for bias. We will return to this shortly.

That is basically it. However, as we shall see, the devil is in the details, and even giving this very basic explanation on what the bootstrap is all about, two important critisisms could be raised against the description above.

1. The description too sloppy. If we want to make a more rigorous description following e.g. Shao & Tu (1995) the bootstrap is really a combination of two techniques: the substitution principle and a numerical approximation. Without getting into details, the idea is that there exists a bootstrap distribution, which is the distribution we get when we do the resampling conditional on our present sample. The substitution that we make is to replace this unknown distribution of interest by an empirical distribution, e.g. the empirical distribution of our sample. This gives us a theoretical bootstrap distribution, which may have several interesting characteristics, and is the subject of study in theoretical work on bootstrap estimators. These estimators however seldom have closed form solutions[1], and we hence need to approximate them numerically, which is what we do when we actually carry out the repeated resampling and average over the bootstrap samples to get the expected value.

2. The definition is to narrow. The general tendency in the literature is to say that the bootstrap is a bit more than what we just have discussed, and therfore to label the above procedure "a non-parametric bootstrap". It is then thought of as non-parametric as opposed to an estimator where we would use family of distributions rather than an empirical one doing the bootstrap. In our simplistic experiment above, it would correspond to saying that $N(\bar{y}; s)$ is the best approximation to the heights of economists, and then draw values from this distribution

[1] See for example Shao & Tu (1995) pp. 10f for an example where such a closed form solution actually exists. The boostrap estimator of the variance of a sample median is shown to be equivalent to the estimator of Maritz & Jarret (1978).

rather than doing the resampling, yielding an even more rediculus estimator. The more general definition of the bootstrap would then be any procedure drawing samples from a DGP, either given by a parametric family or by an empirical distribution, which uses these so called bootstrap samples to draw inference about quantities of interest. In the following we will almost exclusively be discussing the non-parametric bootstrap and will hence surpress the "non-parametric" epithet if there is no risk of confusion.

2 Basic use of the bootstrap

Having looked briefly at the general principle we will now look at the fields where the bootstrap has been used most commonly up until present.

2.1 Standard Errors

The first use made of the bootstrap was to estimate the standard errors for e.g. estimators in cases where there were no analytical asymptotic approximations available. (Once again, you would hardly ever be interested in bootstrapping the variance for a sample mean.) The procedure of calculating bootstrap standard errors for basically any estimator is straightforwardly described in Efron & Tibshirani (1993), and we will here just sketch the general idea and then say a few words of warning.

Let's however first fix some notation, which I will try to stick to throughout. We will use the superscript b to indicate any bootstrapped quantity; we will e.g. label the original samples y, the quantities of interest $\mu(y)$ and hence their bootstrap analogues as y_j^b and μ_j^b y_j^b where j indicates the j:th out of B bootstrap realisations. The expected value of the bootstrapped quantity which is obtained by calculating the mean of the bootstrap realisations is denoted by μ^b:

In terms of the introduced notation, the general idea is now to estimate the standard error of a parameter estimate of interest by computing the standard deviation of θ_j^b; that is

$$\widehat{se}(\theta(y)) = \left(\frac{1}{B-1} \sum_{j=1}^{B} \left(\theta_j^b - \theta^b \right)^2 \right)^{\frac{1}{2}} .$$

A typical example where this type of bootstrap estimator has been advocated is the variance of Manski's maximum score estimator (see Greene (1997) pp 902 f). Since there is no likelihood argument behind the Manski (1975) estimator, standard information matrix estimates are not available and the bootstrap might seem useful. There are however, as yet to my knowledge, no theoretical results established on this bootstrap estimator, which makes using it a somewhat risky business, since we cannot be assured that the estimates will converge to true values at all[2].

So, are bootstrap standard errors useful? Well, in his lecture notes on the bootstrap, Marc Nerlove gives a simple example of resampling OLS residuals using these to estimate the standard errors for the regression coefficients, finding these severely downward biased (Nerlove (1998)). The reason claimed, as described by Nerlove (1998), is that the fitted residuals which he resamples

$$e = y - X\theta$$

will have a covariance matrix of

$$E(ee') = \sigma^2 \left(I - X(X'X)^{-1}X' \right)$$

[2] For the Horowitz (1992) smoothed maximum score estimator bootstrap results do exist (Horowitz (1996)), which however is a direct corrolary of assymptotic approximations existing for the original estimator, making the bootstrap less necessary but still useful according to the results in the latter study.

which is certainly different from the error term he uses in his DGP, which is $\varepsilon \sim N(0; \sigma^2 I)$: This will not work very well, and the reason is the dependence of the bootstrapped quantity of interest, the variance of the OLS-coefficients on unknown parameters. For reasons that soon will become evident, we would however not expect the bootstrap standard errors to perform substantially worse than the asymptotic ones. The main problem in Prof. Nerloves aplication is probably that he uses unadjusted residuals. Since OLS residuals are generally smaller than the error terms of the regression model, these should be adjusted by scaling them by $\sqrt{n/(n-k)}$. If done properly, the bootstrap estimate variance should, with an increasing number of bootstrap iterations, converge exactly to the asymptotic OLS variance-covariance estimator.

There are seldom any efficiency gains from applying the bootstrap to estimate standard errors, for those reasons that, once again, will become evident when we look at the properties of bootstrap tests, shortly. The literature has over the last decade hence moved away from variance estimation by using the bootstrap, and instead focused on bootstrap tests, tests which in most of the cases are the reasons as to why we are interested in the variance estimates in the first place.

2.2 Bias correction

Before turning to the field of bootstrap tests, we will consider another common application of the bootstrap, i.e. correcting for bias. Even if we know that an estimator is consistent, it might suffer from bias in finite samples. By applying the bootstrap we can try to correct this bias using the following simple procedure:

1. Estimate the parameter of interest, θ; by e.g. OLS, IV or ML to obtain $\hat{\theta}$:

2. Construct B bootstrap samples and compute μ^b as

$$\mu^b = \frac{1}{B}\sum_{j=1}^{B} \mu_j^b$$

where μ_j^b is the same estimator that was used to obtained $\hat{\mu}$ applied to the bootstrap sample.

3. Estimate the bias as

$$b^b\left(\hat{\mu}\right) = \mu^b - \hat{\mu}$$

4. Calculate the bias-corrected estimate as

$$\tilde{\mu} = \hat{\mu} - b^b\left(\hat{\mu}\right) = 2\hat{\mu} - \mu^b$$

The idea behind the procedure above is hence that the difference between the estimate $\hat{\mu}$ and the true value μ; should be the same as the difference between μ^b and $\hat{\mu}$; or, loosely, that the the relation of the bootstrap sample to the original sample is the same as the relation between the original sample and the true population. This forms the basis for the simplest form of bootstrap bias reduction. Further descriptions on these procedures may be found in Efron & Tibshirani (1993).

There is an obvious problem with this approach, namely that we assume that the bias is constant and does not vary with the parameter value. There is normally no good reason to expect this to be the case. In MacKinnon & Smith Jr. (1998) the use of the bootstrap is explored in settings where the bias function is not assumed to be constant. Their results of generalising the bias function is encouraging, though there is a clear trade-off in terms of efficiency loss from using the bias corrected estimator, to the extent that using the corrections may increase the mean squared error of the estimator.

In Ferrari & Cribari-Neto (1998), the authors seek to unify the literature of bootstrap bias correction with the one of analytical ditto. In the paper, which is somewhat involved, the equivalence of analytical and bootstrap correction is demonstrated for ML estimeators of models with one parameter. For more general models the authors provide som Monte Carlo evidence expressing a weak preference for the analytical correction, confirming the results of MacKinnon & Smith Jr. (1998) in the respect that the bootstrap corrections may induce increased MSE.

2.3 Bootstrap tests

The main reason for using bootstrap tests rather than asymptotic tests is that the latter may in finite samples be biased, i.e. they have empirical sizes that differ from their nominal ones. A main feature of bootstrap tests is that, under certain condition which we will look into shortly, their empirical sizes will converge to the true sizes faster than asymptotic tests and at times converge considerably faster. Bootstrap tests with correct sizes can also often be shown to have basically the same power properties as their asymptotic counterparts[3]. Before discussing the issues of convergence, we will describe what a bootstrap test is all about.

To pinpoint the differences, let us first briefly consider traditional hypothesis testing. Suppose that we have a sample from which we have obtained an estimate $\hat{\mu}$ of an unknown parameter μ: To test a hypothesis on this single parameter, say $H_0 : \mu = 0$, we simply employ a t-test, which we know will have a certain distribution at least asymptotically, given that the null is true. Using this approximation we will assess whether the test statistic is likely to have been drawn from the distribution in question.

[3] The results mentioned as well as those given below in this section are proofed and discussed at further length in Davidson & MacKinnon (1996b), Davidson & MacKinnon (1996a) and Horowitz (1997)

When forming a bootstrap test in this context, we could instead use the fact that we know the sample value of μ in the sample, and then treat this sample as if it was a population for which the true value of the parameter of interest is in fact the estimated parameter, $\hat{\mu}$.[4] We then create a bootstrap sample by randomly resampling observations from our original sample with replacement, thereby obtaining a sample with the same size as the original one, but not with the same composition. The bootstrap sample is then used to obtain a bootstrap estimate, $\hat{\mu}^b$; which in repeated (re)sampling will be equal to $\hat{\mu}$ on average:

The resampling procedure is then as usual carried out a large number of times and for each bootstrap estimate a test statistic is formed based on the null that $\hat{\mu}^b = \hat{\mu}$: Doing that we will obtain a distribution of test statistics which is generated taking the characteristics of the data in the original sample into consideration, while explicitly imposing the restriction that the null is true. Calculating the 95th percentile for the absolute values of the t-statistics obtained from the bootstrap estimates, we get the bootstrap critical value for our t-test at the 5% significance level, with which we can compare the value of the t-test obtained from the original sample testing e.g. the hypothesis $\hat{\mu} = 0$: The principle behind the bootstrap test is hence to construct a true null, e.g. $\hat{\mu}^b = \hat{\mu}$; and then simulate the distribution of the test statistic using the data at hand.

Even though the intuition might appear straightforward, the rigorous argument as to why the bootstrap provides refinements compared to first order asymptotics is somewhat involved. The fundamental property we require of the test in order for the bootstrap to provide refinements compared to first order asymptotics, is the one of pivotalness, i.e. that the test distribution does not depend on any unknown parameters.

[4] It is certainly not necessary to use the set-up suggested here, since all bootstrap schemes that impose the true null would be valid. However, in cases such as this when μ delimits H_0, it seems natural and straightforward to use the present setting.

This is certainly true asymptotically for the t-test, since we know that its distribution will converge to a $N(0;1)$; which is evidently independent of any unknown parameters. For the t-test we may obtain a distribution independent of unknown parameters even in finite samples[5], which would suggest that a bootstrap t-test should work quite well in even when the sample size is comparatively small. For most tests, the property of pivotalness is fulfilled asymptotically, since their limiting distributions are quite often normal och chi-squared or whatever. If we however in the linear regression context are willing to assume normal errors (under the null), several tests such as tests for serial correlation, heteroscedasticity (including ARCH), skewness and kurtosis are exactly pivotal and the bootstrap test will then be exact even in finite samples. To see why bootstrap test actually provide refinements as compared to asymptotic tests, and why the pivotalness is important in this context, we need argue a bit more rigorously, which will be the scope of next section.

3 Why do bootstrap tests provide refinements?

My conjecture from studying the literature is that the field where the bootstrap has been most commonly applied is the one of bootstrap testing. One reason for this may be that there is a clear theoretical support for the bootstraps ability to provide refinements as compared to asymptotic approximations when bootstrapping pivotal tests. Since the importance of bootstrapping "pivotal quantities" is by now well understood by researchers active in the field, I think that it is useful to spell out the most widely spread proof on as to why this refinement occurs in somewhat greater detail. The description of the proof is somewhat "sketchy", but basically follows Horowitz (1997) and Hall (1992). The reader who desires more of rigour should look

[5] That is, if we are willing to assume normal residuals or fixed regressors.

up especially the latter reference.

Let us start out with a set of data which is a random sample $\{X_i : i = 1; \ldots; n\}$ from a distribution, the CDF of which we will denote by F. If it would be possible to describe the distribution by some finite set of parameters, θ; we would write the CDF as $F(x;\theta)$, which would be equivalent to adopting an entirely parametric method. The proof we are about to go through could the straightforwardly carried out for a parametric bootstrap, but for the case most relevant to this survey, the non-parametric bootstrap, we will have to make use of a more general Empirical Distribution Function (EDF) which we will denote by F_n; the generic estimator of which is

$$\hat{F}_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \le x):$$

This estimator, $\hat{F}_n(x)$; will under mild regularity conditions converge to F almost surely at the rate of $O\left(n^{-\frac{1}{2}}\right)$. (See e.g. Davidson (1994) p. 332)

Let us furthermore introduce a test for a certain H_0 about the distribution from which $\{X_i\}$ is drawn, and label this test $T_n(X_1; \ldots; X_n)$; the finite-sample distribution of which under the true H_0 is $G_n(z;F) \equiv P(T_n \le z)$.(z being the critical value for rejection). If we now take the case of a symmetric two-sided test, we reject H_0 at the α level if $|T_n| > z_{n\alpha}$ where the critical value $z_{n\alpha}$ solves

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

Since we do not know F, we cannot obtain $z_{n\alpha}$ right away. Depending on the circumstances, there are now at least three different ways to proceed:

1. Suppose that T_n is pivotal in finite samples. This means that G_n will not depend on F at all, and we will know the value of $z_{n\alpha}$ exactly. This is for instance the case with a t-test on a regression coefficient if we have normally distributed errors. What we simply do is to obtain $z_{n\alpha}$ from

a standard t-distribution.

2. Few tests are pivotal in finite samples. Most of the tests employed in econometrics are however asymptotically pivotal. In the case of our t-test, it will converge to a standard normal variable, which of course is independent of F and any other parameter, for that matter. What we then do is to use the standard normal distribution to obtain an approximation for $z_{n\alpha}$.

3. These asymptotic approximations can at times be quite poor. A third route is hence to aproximate F by F_n and hence form the test based on $G_n(z; F_n)$; and that is essentially what a bootstrap test is all about, i.e. in our two-tailed case solving

$$\alpha = 1 - \left(G_n\left(z^b_{n\alpha}; F_n\right) - G_n\left(-z^b_{n\alpha}; F_n\right)\right) \qquad (2)$$

where $z^b_{n\alpha}$ is the bootstrap critical value. Normally we cannot obtain an analytical expression for $G_n\left(z^b_{n\alpha}; F_n\right)$ and we must hence resort to numerical simulations through Monte Carlo resampling, which is the way the bootstrap tests are normally carried out[6].

The main benefit of the third route, the bootstrap, is that these tests do converge faster than asymptotic approximation. To prove this we need the higher order approximation known as the Edgeworth expansion which applied to our empirical distribution for the test takes the form of

$$G_n(z; F) = G(z; F) + n^{-\frac{1}{2}} g_1(z; F) + n^{-1} g_2(z; F) + o\left(n^{-1}\right); \qquad (3)$$

where $G(z; F)$ is the the asymptotic CDF of T_n; g_1 an even function of z

[6] Recall the discussion in the introduction, that the bootstrap consists of combining two principles. First, the substitution principle; we replace F by F_n. Second, numerical approximation, carried out by resampling.

for each F and g_2 an odd ditto. Furthermore $g_2(z;F_n)$ will converge almost surely to $g_2(z;F)$ uniformly over z[7]. Now using (1) and (3) we get that

$$\alpha = P(|T_n| > z) = 1 - [G(z;F) - G(-z;F)] - 2n^{-1}g_2(z;F) + o\left(n^{-1}\right).$$

Note that we have used the evenness of g_1 and that $o(n^{-1}) \cdot o(n^{-1}) = o(n^{-1})$.

Suppose that we now form a bootstrap test replacing F by F_n, to get

$$\alpha^b = P^b\left(\left|T_n^b\right| > z\right) = 1 - [G(z;F_n) - G(-z;F_n)] - 2n^{-1}g_2(z;F_n) + o\left(n^{-1}\right).$$

Subtracting now the true value of the test size from the bootstrap test size we get

$$\begin{aligned} P^b\left(\left|T_n^b\right| > z\right) - P(|T_n| > z) = \; & [G(z;F) - G(z;F_n)] \qquad (4) \\ & [-G(-z;F) - G(-z;F_n)] \\ & + 2n^{-1}[g_2(z;F) - g_2(z;F_n)] \\ & + o\left(n^{-1}\right) \end{aligned}$$

If $G(\cdot)$ is sufficiently smooth, what matters here will be rate of convergence of F_n to F, which as we have previously stated is of $O\left(n^{-\frac{1}{2}}\right)$, and is the leading order in the expression above. We hence see that the bootstrap test under quite general conditions have sizes that converge to their true ones at the rate of $O\left(n^{-\frac{1}{2}}\right)$. This is however the same rate of convergence as the standard asymptotic and we would gain nothing from taking the trouble to use bootstrap tests. But here comes the trick. If the test that we bootstrap

[7] Just stating the Edgeworth expansion (or rather its inversion) is a regrettably unintuitive way of presenting the proof and certainly a flaw of this exposition. These higher order expansions are however quite tricky stuff, and must admit that I find it hard to convey any firm intuition here.

is asymptotically pivotal, its distribution will not depend on any unknown parameters, which directly implies that $G(z;F_n) = G(z;F)$ for all z; and that (4) simplifies to

$$\begin{aligned} P^b\left(\left|T_n^b\right| > z\right) - P(|T_n| > z) &= +2n^{-1}[g_2(z;F) - g_2(z;F_n)] + o\left(n^{-1}\right) \\ &= 2O\left(n^{-1}\right)O\left(n^{-\frac{1}{2}}\right) + o\left(n^{-1}\right) = o\left(n^{-1}\right) \end{aligned}$$

having applied the mentioned convergence result for $g_2(z;F_n)$ to $g_2(z;F)$:

We hence finally obtain the desired result: If the test we bootstrap is asymptotically pivotal, which almost all tests used in econometrics are, the bootstrap test will converge faster by an order of (at least) $O\left(n^{-\frac{1}{2}}\right)$ compared to the asymptotic approximation.

4 A few applications of the bootstrap

We will now survey a few recent contributions to the bootstrap part of the econometrics literature. The bias towards tests and the usage of pivotal quantities in this exposition, is claimed to be a manifestation of the state of the literature, rather than my own preferences.

4.1 SUR-regressions

It is a well-known property of the Zellner SUR estimator, that the asymptotic standard-errors of the regression coefficients may be severely downward biased. Already more than ten years ago Marais (1986) and in a published piece of work a few years later Atkinson & Wilson (1992), attempt to address the problem by using bootstrap standard errors. The evidence was mixed but did indicate some improvement. It should be evident from the earlier discussions on pivotalness, that the results that the authors obtained most likely were parameter dependent, and that a bootstrap test would have more clear-cut results. In Rilstone & Veall (1996) the approach of using percentile-t confidence intervals was adopted. This approach involves bootstrapping pivotal quantities, and hence yielded results much more clear-cut and encouraging than in the previous studies: the bootstrap confidence intervals in their Monte Carlo study almost exactly covered the nominal ones.

Whereas this approach does not add much conceptually compared to the test procedures described earlier, it however allows assymetric confidence intervals which might be a potentially important feature, and we will therefore look at the procedure used briefly.

The description on as to how the bootstrap sample was created is somewhat vague in the paper, but was presumably done resampling the residuals of the second step, which should be independent by construction. The algorith then works briefly as follows:

1. Estimate θ by SUR obtaining $\hat{\theta}$ and $\hat{\sigma}\left(\hat{\theta}\right)$
2. Resample and apply SUR to the bootstrap dataset to get $\hat{\theta}^b$ and $\hat{\sigma}\left(\hat{\theta}^b\right)$
3. Create a bootstrap t-statistic as

$$\hat{t}^b = \frac{\hat{\theta}^b - \hat{\theta}}{\hat{\sigma}\left(\hat{\theta}^b\right)}$$

4. Repeat steps 2-3 a few hundred times or so.
5. Sort the distribution of $\hat{t}^b$ and extract $\hat{t}^b_{1-\alpha/2}$ and $\hat{t}^b_{\alpha/2}$:
6. Form a $(1-\alpha)*100\%$ bootstrap-t confidence interval as

$$=^b_t = \left[\hat{\theta} - \hat{\sigma}\left(\hat{\theta}\right)\hat{t}^b_{1-\alpha/2}; \hat{\theta} - \hat{\sigma}\left(\hat{\theta}\right)\hat{t}^b_{\alpha/2}\right]$$

4.2 Hausman tests

Wong (1996) provides a nice and straightforward implementation of a bootstrap test. The Hausman exogeneity test, which is asymptotically pivotal converging to a χ^2 distribution, is claimed to have a bootstrap equivalent which is not just converging of $O\left(n^{-\frac{1}{2}}\right)$ but $O(n^{-1})$ faster than the asymptotic approximation[8]. Data is generated as

$$y = \beta_0 + x\beta_1 + u$$

and under a true null of exogeneity there is no correlation between x and u. For a false null these are correlated, but an additional regressor z is generated which is uncorrelated with u but correlated with x. The alternative estimator to OLS is standard IV. The algorithm for the bootstrap test is as follows:

1. Estimate the equation by OLS and IV and compute the Hausman test $\hat{Q}$
2. Resample residuals estimated as $\hat{u}^b = y - \hat{\beta}_0^{OLS} + x\hat{\beta}_1^{OLS}$
3. Construct $y^b = \hat{\beta}_0^{OLS} + x\hat{\beta}_1^{OLS} + \hat{u}^b$ (Note that using the OLS estimates constructing the bootstrap data-set, the true null is explicitly imposed)
4. Estimate once again by OLS and IV and compute the Hausman test $\hat{Q}^b$
5. Repeat steps 2 - 4 lots of times.
6. Sort the distribution of $\hat{Q}^b$
7. Reject exogeneity at the $\alpha \times 100\%$ - level if $\hat{Q} > \hat{Q}^b_{1-\alpha}$

[8] This is generally true for tests converging to a χ^2: See Hall (1992).

Whereas the asymptotic test appears size distorted (undersized) in the Monte Carlo presented in the paper, the bootstrap works perfectly well in the experiments reported. It also appears that the advantage of using bootstrap increases as the correlation between the instrument and the regressors becoms low, which should have important implications for practical applications.

4.3 Time Series and Dynamic models

An important issue that we have not yet touched upon, is how to carry out the bootstrap in dynamic models. There is an extensive survey available in Li & Maddala (1996), where an entire issue of Econometric Reviews is devoted to the paper and five commenting notes by other leading researchers in the field. We will here just discuss the general principle according to which the bootstrap is carried out in dynamic models, and how time series models with non-IID errors could be handled.

Let us first consider a simple dynamic model of the type

$$y_t = \alpha y_{t-1} + \beta_0 + \beta_1 x_t + \varepsilon_t;\ \varepsilon_t \sim IID\left(0; \sigma^2_\varepsilon\right)$$

As long as the errors are IID, the implementation of the so called recursive bootstrap is straightforward:

1. Estimate α and β by some estimator of preference.
2. Obtain rescaled[9] residuals

$$\tilde{\varepsilon}_t = \left(\frac{n}{n-k}\right)^{\frac{1}{2}} \hat{\varepsilon}_t$$

where

$$\hat{\varepsilon}_t = y_t - \hat{\alpha} y_{t-1} + \hat{\beta}_0 + \hat{\beta}_1 x_t$$

[9] The rescaling is necessary due to the aforementioned fact that OLS residuals underestimate the true errors.

3. Generate a bootstrap sample by resampling from the $\hat{\varepsilon}_t$:s, and either using an actual y_0 or drawing it from its unconditional distribution, i.e.

$$y_0 \sim N\left(\frac{\hat{\beta}_0}{1-\hat{\alpha}}, \frac{\hat{\sigma}^2}{1-\hat{\alpha}^2}\right)$$

and then creating the bootstrap data recursively as

$$y_t^b = \hat{\alpha} y_{t-1} + \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\varepsilon}_t^b \qquad (5)$$

4. Apply the estimator to the bootstrap sample and calculate a pivotal statistic of interest. Then carry on as usual.

If there is an error structure of more complex form present, the so called Moving Blocks bootstrap could be used in static time series models. The general idea is to perserve this error structures by resampling (overlapping) blocks of residuals rather than resampling them one by one. There are several problems with this approach, not the least that the serial dependence will just prevail within the blocks, making the approximation of the residuals distribution a rather "rough" one. This and other problems as well as different ways of implementing the moving blocks bootstrap is thoroughly discussed in Li & Maddala (1996).

4.4 Nonstationarity and the Bootstrap

First and foremost, it needs to be stressed that the issue as to if and when the bootstrap can be applied in nonstationary contexts appears to be an open question. Several studies exist, giving somewhat mixed evidence. The standard results of improvements compared to asymptotic approximations do not immediately carry over when the quantities we bootstrap come from none-stationary time series.

This particular research field is quite active and some progress has been achieved. The earliest reference in the econometrics literature seems to be Harris (1992), where a bootstrap test for a unit root in a univariate context (the ordinary Dickey-Fuller test) is shown to have better properties than the asymptotic test, once the unit root is imposed. There are no theoretical results, at least not in the econometrics literature, telling us if we should expect improvements from bootstrap tests in these cases, and trying to generalise the univariate test to a bootstrap version of the Johansen (1988)- test, Harris & Judge (1998) find that the bootstrap does not work at all. Their conjecture is that it is the mix of stationary and non-stationary series that makes the bootstrap backfire. There is however, as mentioned, no theoretical explanations, as yet.

As for cointegrating regressions, where the residuals have been brought to stationarity, there however seems to be a case for the bootstrap. Li & Maddala (1997) gives evidence for improved inference using bootstrap tests on the coefficients of cointegrating vectors. An important implication from their studies is that when the series involved are $I(1)$, it is necessary to bootstrap the residuals (which under cointegration are $I(0)$). Further results on the usefulness of bootstrap tests on cointegrating vectors estimated by the Johansen (1988)-procedure is also given in Gredenhoff (1998).

4.5 GMM Bootstrap tests

Even if the bootstrap is quite able to improve inference for these estimators as well, forming bootstrap tests for GMM estimators provides some specific difficulties. Suppose that we have used the GMM to obtain a parameter vector of interest, $\hat{\mu}$; and have hence assumed that the vector satisfies the population moment conditions

$$E[g(x;z;\mu)] = 0;$$

where $g(x;z;\theta)$ is a vector of moment conditions, x regressors and z instruments. In order to restrict $\hat{\theta}^b$ to equal $\hat{\theta}$; the bootstrap sample moments must satisfy the same moment conditions as the original sample, which would not be the case if we bootstrap observations with probability $1/N$ in the GMM case. The reason is that the sample moments

$$\hat{g}_N\left(\hat{\theta}\right) = \frac{1}{N}\sum_{i=1} g\left(x;z;\hat{\theta}\right);$$

are generally not zero when the model is overidentified. If we want to restrict $\hat{\theta}^b$ to equal $\hat{\theta}$, we must therefore restrict the sample moments used accordingly. This problem has recently been noted and addressed in two papers, each suggesting a different approach to solve the problem.

4.5.1 Brown & Newey

Brown & Newey (1995) consider an approach based on using an alternative estimator of the distribution of the data. The above problem is solved by replacing the empirical distribution with a moment restricted estimator of the distributions. By doing this, the moment conditions are explicitly imposed.

More formally, Brown & Newey use a distribution function estimator that imposes the moment conditions. Instead of resampling with probabilities $1/N$, each observation is given an individual probability, p_i, of being drawn. These estimated probabilities reflect how well the moment restrictions are fulfilled in each case.

The probabilities are calculated using a so called empirical likelihood approach[10]. Let observation i in the original data be drawn with a probability

[10] See further Owen (1988) for a treatise on Empirical Likelihood

p_i, where p_i solves the following maximisation problem:

$$\max_{p_1 \ldots p_N} \sum_{i=1}^{N} \ln(p_i) \quad \text{s.t.} \; p_i > 0; \tag{6}$$
$$\sum_{i=1}^{N} p_i = 1;$$
$$\sum_{i=1}^{N} p_i g\left(z_i, \hat{\theta}\right) = 0.$$

Here, $g\left(z_i, \hat{\theta}\right)$ is obtained from an efficient GMM estimation on the original sample. Instead of solving the maximisation problem in (6), Brown & Newey present an easier way to calculate the probabilities above without having to solve an N-dimensional maximisation problem. Let $\hat{g}_i = g\left(z_i, \hat{\theta}\right)$ be a $J \times 1$ vector of moments, $i = 1, \ldots, N$. Furthermore, let $\hat{\lambda}$ be a $J \times 1$ vector given by

$$\max_{\lambda} \sum_{i=1}^{N} \ln(1 + \lambda' \hat{g}_i) \quad \text{s.t.} \; 1 + \lambda' \hat{g}_i > 0.$$

Then p_i is given by

$$\hat{p}_i = N^{-1}\left(1 + \hat{\lambda}' \hat{g}_i\right)^{-1}.$$

This empirical likelihood estimator is a member of a class of distribution estimators, which are moment restricted. Brown & Newey show that, if one only has information about the moment conditions, the proposed moment restricted estimator is the asymptotically most efficient estimator available.

4.5.2 Hall & Horowitz

Hall & Horowitz (1996) propose a different way of recentring the GMM bootstrap. They create bootstrap samples in the traditional way, that is, by drawing each observation from the empirical distribution with probability $1/N$. Instead of recentring the distribution as Brown & Newey do, they recentre the moments around their empirical values and use these recentred moments when estimating the model and when forming a bootstrap version of the overidentification test-statistic.

Let $g\left(z; \hat{\theta}\right)$ be given by a GMM estimation on the original data. The recentred moments are then

$$g_N^b(\theta) = N^{-1} \sum_{i=1}^{N} g\left(z_i^b; \theta\right) - \hat{g};$$

where $\hat{g}$ is given by

$$\hat{g} = N^{-1} \sum_{i=1}^{N} g\left(z_i; \hat{\theta}\right):$$

4.5.3 Applications

Monte Carlo evidence for the performance of the two different approaches in a dynamic panel data model is provided in Bergström, Dahlberg & Johansson (1997) and Bergström (1997), where it is demonstrated that even if neither of the approches can ba argued to be better than the other as a whole, they both provide nice improvements as compared to asymptotic approximations. Quite few applications of the GMM bootstrap have been undertaken. To our knowledge the only ones are Dahlberg & Johansson (1997) and Bergström & Lindberg (1998).

5 Final remarks

This has been a rough and ready exposure of some of the basics and some of the ongoing research in the bootstrap field. There are surely important parts of and paths through this field that I have left out, but hopefully this paper could serve as a short introduction into this fascinating field. The bootstrap is certainly a useful device, if applied properly. If the problem of interest is designed rigorously, the correct (pivotal) quantity is bootstrapped, and the result is evaluated carefully, using the bootstrap could probably prove quite rewarding in many contexts.

The principles of the bootstrap are easy and intuitive. Yet, in order to fully understand why the bootstrap works better than first-order asymptotics in a specific case, it is important that the theoretical properties of the boostrap estimators are investigated thoroughly, something that might prove very important for estimators applied to especially nonstationary data. This should be an important field for further theoretical resarch. When it comes to empirical applications, bootstrap procedures have become, if not yet standard, but at least increasingly important to researchers in the time series field. Applications for micro- and panel data are really much more scarce, but should become much more common, since there are several encouraging theoretical and Monte Carlo results speaking in favour of the bootstrap in these settings.

References

Atkinson, S. & Wilson, P. W. (1992). The bias of bootstrapped versus conventional standard errors in the general linear and SUR models, Econometric Theory 8: 258–275.

Bergström, P. (1997). On bootstrap standard errors in dynamic panel data models., Working paper 23, Department of Economics Uppsala University.

Bergström, P., Dahlberg, M. & Johansson, E. (1997). GMM bootstrapping and testing in dynamic panels, Working paper 10, Department of Economics Uppsala University.

Bergström, P. & Lindberg, S. (1998). Firms' financial policy and labour demand: Theory and evidence, Working paper ??, Department of Economics Uppsala University.

Brown, B. & Newey, W. (1995). Bootstrapping for GMM, Mimeo, MIT.

Dahlberg, M. & Johansson, E. (1997). An examination of the dynamic behavior of local governments using GMM bootstrapping methods, Chapter III in "Essays on Estimation Methods and Local Public Economics." Doctoral Disseration, Department of Economics, Uppsala University.

Davidson, J. (1994). Stochastic Limit Theory, Oxford University Press, Oxford.

Davidson, R. & MacKinnon, J. G. (1993). Estimation and Inference in Econometrics, Oxford University Press, Oxford.

Davidson, R. & MacKinnon, J. G. (1996a). The power of bootstrap tests, Mimeo, Queen's Institute for Economic Research.

Davidson, R. & MacKinnon, J. G. (1996b). The size distortion of bootstrap tests, Discussion Paper 937, Queen's Institute for Economic Research.

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

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

Ferrari, S. & Cribari-Neto, F. (1998). On bootstrap an analytical bias correction, Economic Letters 58: 7–15.

Gredenhoff, M. (1998). Bootstrap Inference in Time Series Econometrics, PhD thesis, Stockholm School of Economics.

Greene, W. H. (1997). Econometric Analysis, 3d Edition, Prentice Hall, Upper Saddle River, New Jersey.

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

Hall, P. & Horowitz, J. L. (1996). Bootstrap critical values for tests based on Generalized-Method-of-Moments estimators, Econometrica 64: 891–916.

Harris, R. (1992). Small sample testing for unit roots, Oxford Bulletin of Economics and Statistics 54(4): 615–625.

Harris, R. & Judge, G. (1998). Small sample testing for cointegration using the bootstrap approach, Economics Letters 58: 31–37.

Horowitz, J. (1997). Bootstrap methods in econometrics: Theory and numerical performance, in D. M. Kreps & K. F. Wallis (eds), Advances in Economics and Econometrics: Theory and Applications, Cambridge University Press, Cambridge, Great Britain, chapter 7, pp. 188–222.

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

Horowitz, J. L. (1996). Bootstrap critical values for tests based on the smoothed maximum score estimator, Mimeo, Dpt of Economics, University of Iowa.

Johansen, S. (1988). Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12: 231–254.

Li, H. & Maddala, G. (1996). Bootstrapping time series models, Econometric Reviews 15(2): 115–158.

Li, H. & Maddala, G. (1997). Bootstrapping cointegrating relations, Journal of Econometrics 80: 297–318.

MacKinnon, J. G. & Smith Jr., A. A. (1998). Approximate bias correction in econometrics, Journal of Econometrics 85: 205–230.

Manski, C. (1975). The maximum score estimator of the stochastic utility of choice, Journal of Econometrics 3: 205–228.

Marais, M. (1986). On the finite sample performance of estimated genarlized least squars in seemingly unrelated regression, Working paper, Graduate School of Business, University of Chicago.

Maritz, J. & Jarret, R. (1978). A note on estimating the variance of the sample median, 73: 194–196.

Nerlove, M. L. (1998). Bootstrapping - lecture notes, Mimeo, AREC, Univeristy of Maryland.

Owen, A. (1988). Empirical likelihood, Annals of Statistics.

Rilstone, P. & Veall, M. (1996). Using bootstrapped confidence intervals for improved inferences with seemingly unrelated regression equations, Econometric Theory 12 (3): 570–581.

Shao, J. & Tu, D. (1995). The Jackknife and Bootstrap, Springer Verlag, Berlin.

van Giersbergen, N. & Kiviet, J. F. (1994). How to implement bootstrap tests in static and dynamic regression models, Discussion Paper 7-94-130, Tinbergen Institute.

Wong, K.-F. (1996). Bootstrapping hauman's exogeneity test, Economic Letters 53: 139–143.