BOOTSTRAPPING

ABSTRACT. We first discuss empirical bootstrap for the sample mean, and then generalize the analysis to GMM. We then discuss algorithmic details.

1. THE BOOTSTRAP

The bootstrap is a simulation method for computing standard errors and distributions of statistics of interest, which employs an estimated dgp (data generating process) for generating artificial (bootstrap) samples and computing the (bootstrap) draws of the statistic. Empirical or nonparametric bootstrap relies on nonparametric estimates of the dgp, whereas parametric bootstrap relies on parametric estimates of the dgp.

In what follows we mostly focus on the empirical bootstrap (bs).

Almost everything about empirical bootstrap can be understood by studying the bootstrap for the sample mean statistic

$$\bar{X} = \mathbb{E}_n X_i.$$

We first show how one can explore the behavior of $\bar{X}$ using simulation. Consider a fixed law F_0 with the second moment bounded from above and variance bounded away from zero. This characterizes the dgp sufficiently for understanding the standard behaviour of its sample means. As usual we will work with an i.i.d. sample

$$X_1^n = \{X_i\}_{i=1}^n$$

drawn from F_0. In illustrations given below, we shall use standard exponential distribution as F_0 and the sample size of $n = 100$. However, there is nothing special about this distribution and we could have used other distributions with bounded second moments and non-zero variance to illustrate our points.

Since we know the true dgp in this running example, we can in principle compute the exact finite distribution of the sample mean. However, setting aside special cases suitable for textbook problems, the exact distribution of $\bar{X}$ is not analytically tractable. Instead we proceed by simulating out the finite sample distribution of $\bar{X}$. Our simulation will produce the exact distribution, modulo numerical error, which we take as negligible. In Figure 1 we see the resulting finite sample distribution as well as the standard deviation (standard error) for this distribution. The distribution is represented by a histogram computed over $B = 1,000$ simulated samples. The standard error here is computed over the simulated draws of $\bar{X}$, namely

$$\sqrt{\sum_{b=1}^{B}\left(\bar{X}_b - \sum_{b=1}^{B}\bar{X}_b/B\right)^2/B},$$

where $\bar{X}_b$ is the mean in the bth simulated sample. This standard error is a numerical approximation to the standard deviation of $\bar{X}$

$$\sqrt{\mathrm{Var}(\bar{X})} = \sqrt{\mathrm{E}(X - \mathrm{E}X)^2/n},$$

which in the case of the standard exponential and $n = 100$ is $\sqrt{1/100} = 0.1$. Although the exact distribution of $\bar{X}$ is not available without knowledge of F_0, we know that it is approximately normal by the central limit theorem. Thus, since $\mathrm{E}\bar{X} = \mathrm{E}X$, we have that

$$\bar{X} \overset{a}{\sim} N(\mathrm{E}X, \mathrm{E}(X - \mathrm{E}X)^2/n), \quad \text{or}$$
$$\sqrt{n}(\bar{X} - \mathrm{E}X) \overset{a}{\sim} N(0, \mathrm{E}(X - \mathrm{E}X)^2). \tag{1.1}$$

Next we consider the empirical bootstrap. Now we want to understand the behavior of the sample mean $\bar{X}$ from an *unknown* dgp F with characteristics as above. Since we don't know F, we cannot simulate from it. The main idea of bs is to replace the unknown true dgp F with a good estimate $\hat{F}$. Empirical bs uses the empirical distribution $\hat{F}$, which assigns point-masses of $1/n$ to each of the data points $\{X_1, \ldots, X_n\}$. In other words, $\hat{F}$ is a multinomial variable that takes on values $\{X_1, \ldots, X_n\}$ with equal probability $1/n$. We proceed as above by simulating i.i.d. samples (bs samples)

$$X_1^{*n} = \{X_i^*\}_{i=1}^n$$

from $\hat{F}$, which is equivalent to sampling from the original data randomly with replacement.[1] Each bootstrap sample gives us a bootstrap draw of the sample mean

$$\bar{X}^* = \mathbb{E}_n X_i^*.$$

We repeat this procedure many times to construct many bootstrap samples and hence many draws of this statistic.

[1]Note that they key phrase is "with replacement". This means that some observations can be redrawn multiple times to form a bootstrap sample, and some may not be drawn at all.

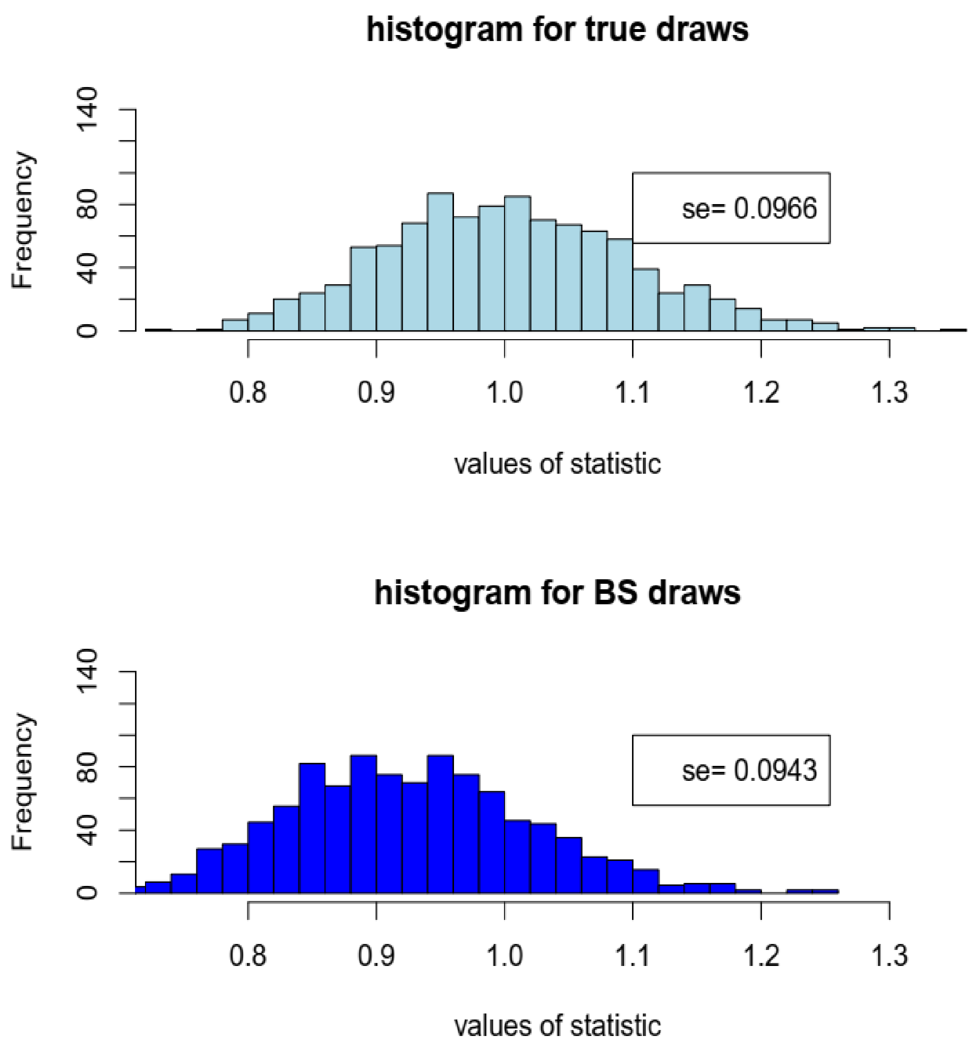


FIGURE 1. True and bootstrap distributions of the mean of a standard exponential random sample, with the sample size equal 100. Both distributions are approximately normal by the central limit theorem, but centered at different points: the true distribution is centered at the true mean and the second is centered at the empirical mean.

In Figure 1 we see the finite sample distribution of $\bar{X}^*$ as well as the standard deviation (standard error) for this distribution. An important point is that this bootstrap distribution is computed *conditional* on one draw of data X_1^n.[2]

[2] In our monte-carlo experiment we saved the first draw of X_1^n, which was then used as data in computation of the bootstrap distribution.

We note that not only the standard deviations of the bootstrap draws $\bar{X}^*$ and actual draws $\bar{X}$ look very similar, but also the overall distribution of bootstrap draws $\bar{X}^*$ and actual draws $\bar{X}$ look very similar. This is *not* a coincidence.

The mean of the bootstrap distribution of $\bar{X}^*$ is

$$\mathrm{E}(\bar{X}^* \mid X_1^n) = \sum_{i=1}^{n} X_i/n = \bar{X}.$$

Similarly, the standard deviation of the bootstrap distribution of $\bar{X}^*$ is

$$\sqrt{\mathrm{Var}(\bar{X}^* \mid X_1^n)} = \sqrt{\mathbb{E}_n(X_i - \bar{X})^2/n},$$

which is simply the root of the empirical variance scaled by n. By the law of large numbers and some simple calculations,

$$\mathbb{E}_n(X_i - \bar{X})^2 \to_P \mathrm{E}(X - \mathrm{E}X)^2,$$

we have that the ratio of bootstrap standard error and the actual standard error converges in probability to 1. Thus, the similarity of the computed standard errors was not a coincidence. Of course, we did not need the bootstrap to compute the standard errors of a sample mean, but we will need it very soon for less tractable cases.

We can approximate the exact distribution of $\bar{X}^*$, conditional on the data, by simulation. Moreover, it is also approximately normal in large samples. By the central limit theorem and the law of large numbers,

$$\bar{X}^* \mid X_1^n \overset{\mathrm{a}}{\sim} N(\bar{X}, \mathbb{E}_n(X_i - \bar{X})^2/n), \tag{1.2}$$

$$\overset{\mathrm{a}}{\sim} N(\bar{X}, \mathrm{E}(X - \mathrm{E}X)^2/n) \quad \text{or}$$

$$\sqrt{n}(\bar{X}^* - \bar{X}) \mid X_1^n \overset{\mathrm{a}}{\sim} N(0, \mathbb{E}_n(X - \bar{X})^2) \tag{1.3}$$

$$\overset{\mathrm{a}}{\sim} N(0, \mathrm{E}(X - \mathrm{E}X)^2).$$

Thus,

1) We see that the approximate distributions of $\sqrt{n}(\bar{X}^* - \bar{X}) \mid X_1^n$ and $\sqrt{n}(\bar{X} - \mathrm{E}X)$, namely $N(0, \mathbb{E}_n(X_i - \bar{X})^2)$ and $N(0, \mathrm{E}(X - \mathrm{E}X)^2)$, are indeed close.
2) This means that their finite-sample distributions must be close.

We summarize the discussion of empirical bootstrap of the sample mean diagrammatically:

world	dgp	sample	statistic	approximate distribution
real	F_0	X_1^n	$\bar{X}$	$\sqrt{n}(\bar{X} - \mathrm{E}X) \overset{\mathrm{a}}{\sim} N(0, \mathrm{E}(X - \mathrm{E}X)^2)$
bootstrap	$\hat{F}$	X_1^{*n}	$\bar{X}^*$	$\sqrt{n}(\bar{X}^* - \bar{X}) \mid X_1^n \overset{\mathrm{a}}{\sim} N(0, \mathbb{E}_n(X_i - \bar{X})^2)$.

Thus, what we see in Figure 1 is not a coincidence: we conclude that the empirical bootstrap "works" or "is valid" for the case of the sample mean. We formalize these statements further below.

It is clear that the reasoning about the approximate distributions (1.1) and (1.2) extends to vector-valued X_i's of fixed dimension. It is also clear that the central limit theorem and approximate normality play a crucial role in the above argument. The argument will generalize to a very large class of estimators that are approximately linear and normal.

The argument above also suggest that once we don't have approximate linearity and normality, the empirical bootstrap may fail. For example, empirical bootstrap does not work for weird statistics such as extreme quantiles, as shown in Figure 2. For such cases, subsampling or "m out of n" bootstrap methods that resample $m \ll n$ observations can often fix the problem (see, e.g. [10]).

The "m out of n" bootstrap methods are also useful for dealing with very large samples, when resampling a smaller number of observations brings about computational costs. Note that in the case of bootstrapping means, or GMM estimators more generally, modification from "n out of n" bootstrap is straightforward and its validity follows by the same argument as for "n out of n" bootstrap.

2. Bootstrapping GMM

2.1. **Some Elementary Ideas and Theory.** In L4 we established that the GMM estimator obeys:

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{\text{a}}{\sim} -(G'AG)^{-1}G'AN(0, \Omega) = N(0, V). \tag{2.1}$$

We want to construct a bootstrap draw $\hat{\theta}^*$ using the bootstrap method that would allow us to mimic the behavior of $\sqrt{n}(\hat{\theta} - \theta_0)$, namely

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \mid X_1^n \overset{\text{a}}{\sim} N(0, V). \tag{2.2}$$

Definition 1 (Definition of Validity of BS)**.** A bootstrap method producing $\hat{\theta}^*$ conditional on data X_1^n is valid if both (2.1) and (2.2) hold. Recall that the first statement means that

$$\sup_{A \in \mathcal{A}} \left| \mathrm{P}(\sqrt{n}(\hat{\theta} - \theta_0) \in A) - \mathrm{P}(N(0, V) \in A) \right| \to 0,$$

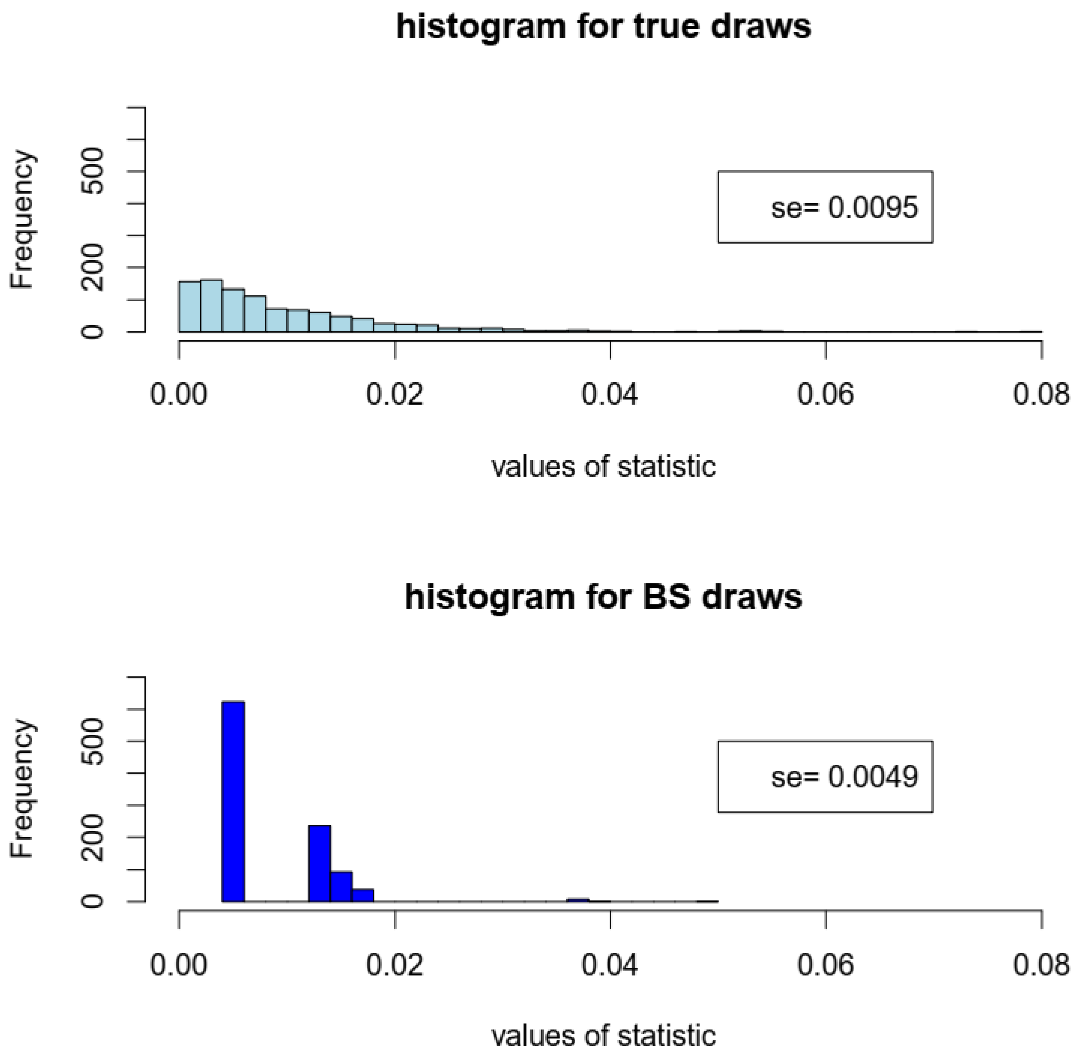


FIGURE 2. True and bootstrap distributions of the minimum of an exponential random sample, with the sample size equal 100. Here the true distribution is approximately exponential (since minimum of an exponential tends to an exponential distribution). The bootstrap distribution is highly discrete and fails to provide a consistent distributional approximation.

for $\mathcal{A}$ denoting the convex sets in $\mathbb{R}^p$, and the second statement means that:

$$\sup_{A \in \mathcal{A}} \left| \mathrm{P}(\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \in A \mid X_1^n) - \mathrm{P}(N(0, V) \in A) \right| \to_P 0.$$

Note that by the triangle inequality this definition implies the following "natural" definition:

$$\sup_{A\in\mathcal{A}} \mathrm{P}(\sqrt{n}(\hat\theta^* - \hat\theta) \in A \mid X_1^n) - \mathrm{P}(\sqrt{n}(\hat\theta - \theta_0) \in A) \ \to_P 0.$$

The previous definition however emphasizes the link with approximate normality, which is key to demonstrating that the empirical bootstrap works.

2.2. **A Quick BS Method for GMM.** A computationally quick way to bootstrap GMM is to bootstrap the average score appearing in the linear approximation to GMM. Recall that we obtained

$$\sqrt{n}(\hat\theta - \theta_0) = -(G'AG)^{-1}G'A \underbrace{\sqrt{n}\mathbb{E}_n Z_i}_{\text{vector mean}} + o_P(1),$$

where $Z_i = g(X_i, \theta_0)$. Thus GMM is approximately a sample mean over the scores Z_i times a fixed matrix $-(G'AG)^{-1}G'A$.

Thus we could simply bootstrap the scores Z_i's: Indeed, let Z^{*n}_1 denote the empirical bootstrap draws from the sample Z_1^n, and define the bootstrap draw $\hat\theta^*$ via the relation:

$$\sqrt{n}(\hat\theta^* - \hat\theta) = -(G'AG)^{-1}G'A\sqrt{n}\mathbb{E}_n(Z_i^* - \bar Z),$$

that is,

$$\hat\theta^* = \hat\theta - (G'AG)^{-1}G'A\mathbb{E}_n(Z_i^* - \bar Z).$$

By the central limit theorem, law of large numbers, and smoothness of the Gaussian law we have the following properties:

$$\begin{aligned} \sqrt{n}\mathbb{E}_n(Z_i^* - \bar Z) \mid Z_1^n &\overset{a}{\sim} N(0, \Omega_n) \\ &\overset{a}{\sim} N(0, \Omega), \end{aligned}$$

where $\Omega_n = \mathbb{E}_n(Z_i - \bar Z)(Z_i - \bar Z)'$ and $\Omega = \mathrm{E}(Z - \mathrm{E}Z)(Z - \mathrm{E}Z)'$.[3] This reasoning implies that our quick bs method is valid:

$$\sqrt{n}(\hat\theta^* - \hat\theta) \mid X_1^n \overset{a}{\sim} -(G'AG)^{-1}G'AN(0, \Omega) = N(0, V).$$

[3]Here the first and second approximations formally mean that:

$$\sup_{A\in\mathcal{A}} \left| \mathrm{P}(\sqrt{n}\mathbb{E}_n(Z_i^* - \bar Z^*) \in A \mid X_1^n) - \mathrm{P}(N(0, \Omega_n) \in A \mid X_1^n) \right| \to_P 0,$$

$$\sup_{A\in\mathcal{A}} \left| \mathrm{P}(N(0, \Omega_n) \in A \mid X_1^n) - \mathrm{P}(N(0, \Omega) \in A) \right| \to_P 0,$$

where $\mathcal{A}$ denotes the convex sets in $\mathbb{R}^p$.

In practice we need to replace G, A, and Z_i's with consistent estimators $\hat{G}$, $\hat{A}$, and $\hat{Z}_i = g(X_i, \hat{\theta})$ such that

$$\hat{G} - G \to_P 0, \quad \hat{A} - A \to_P 0, \quad \mathbb{E}_n \|g(X_i, \hat{\theta}) - g(X_i, \theta_0)\|^2 \to_P 0,$$

and then just define the bootstrap draws via the relation:

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) = -(\hat{G}'\hat{A}\hat{G})^{-1}\hat{G}'\hat{A}\sqrt{n}\mathbb{E}_n(\hat{Z}_i^* - \bar{\hat{Z}}).$$

Note that here we are bootstrapping the estimated scores. We can then arrive at the following conclusion.

Theorem 1 (Validity of Quick BS for GMM). *Under regularity conditions, the quick bootstrap method is valid. That is, the quick bootstrap method approximately implements the normal distributional approximation for the GMM estimator. Moreover, the bootstrap variance estimator* $\hat{V} = \mathrm{E}[(\hat{\theta}^* - \hat{\theta})(\hat{\theta}^* - \hat{\theta})' \mid X_1^n]$ *is consistent, namely that* $\hat{V} - V \to_P 0$.

2.3. **A Slow BS Method for GMM.** There is also a slow bs method for GMM. Here we just bootstrap the whole procedure.

Let $X_1^*, ..., X_n^*$ denote the bootstrap sample. Let $\hat{g}^*(\theta) = \mathbb{E}_n g(X_i^*, \theta) - \hat{g}(\hat{\theta})$, and

$$\hat{\theta}^* \in \arg\min_{\theta \in \Theta} \hat{g}^*(\theta)'\hat{A}^*\hat{g}^*(\theta).$$

Here $\hat{A}^*$ denote the estimator of A obtained using $X_1^*, ..., X_n^*$.

Then we might think that under regularity conditions we would have the linearization

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) = -(\hat{G}'\hat{A}\hat{G})^{-1}\hat{G}'\hat{A}\sqrt{n}\mathbb{E}_n(\hat{Z}_i^* - \bar{\hat{Z}}) + o_P(1),$$

and then this is first-order equivalent to the quick bs method for GMM. This reasoning suggests the following result.

Theorem 2 (Validity of Slow BS for GMM). *Under regularity conditions, e.g. those listed in [6], the slow bootstrap method above is valid. That is, the slow bootstrap method approximately implements the normal distributional approximation for the GMM estimator. Moreover, under appropriate strengthening of regularity conditions, see e.g. [9], bootstrap variance estimator* $\hat{V} = \mathrm{E}[(\hat{\theta}^* - \hat{\theta})(\hat{\theta}^* - \hat{\theta})' \mid X_1^n]$ *is consistent, namely that* $\hat{V} - V \to_P 0$.

Formal proofs of these results are beyond the scope of these lectures, but they can be found in the theoretical literature.

3. Algorithmic Details and Examples of Use

A basic use of bootstrap is for estimation of standard errors and construction of the confidence regions.

The following algorithm constructs estimates of the standard errors.

(1) Obtain many bootstrap draws $\hat{\theta}^{*(j)}$ of the estimator $\hat{\theta}$, where the index $j = 1, \ldots, B$ enumerates the bootstrap draws.
(2) Compute the bootstrap variance estimator

$$\hat{V}/n = B^{-1} \sum_{j=1}^{B} (\hat{\theta}^{*(j)} - \hat{\theta})(\hat{\theta}^{*(j)} - \hat{\theta})'.$$

Report $\hat{s}_k = (\hat{V}_{kk}/n)^{1/2}$ as standard errors for $\hat{\theta}_k$ for $k = 1, \ldots d$.
(3) An alternative is to report standard errors based on the interquartile ranges:

$$\hat{s}_k = [c_k(.75) - c_k(.25)]/(\Phi^{-1}(.75) - \Phi^{-1}(.25)), \quad k = 1, \ldots d,$$

where $c_k(a) = a$-quantile of $\{\hat{\theta}_k^{*(j)}, j = 1, ..., B\}$ and Φ^{-1} is the quantile function of the standard normal distribution Φ.

We illustrate the performance of bootstrap for GMM using the empirical example of L4. Here we focus on bootstrapping the 2-step GMM estimator for the first specification that we estimated. We can mechanically treat the data in that example as i.i.d., because the asymptotic distribution is the same as if we had i.i.d. sampling due to scores being an uncorrelated sequence. We show the histograms of the bootsrap draws $\hat{\theta}^* = (\hat{\beta}^*, \hat{\alpha}^*)$ for the estimator.

The following algorithm constructs a simultaneous confidence region (rectangle) for all components of θ_0.

(1) Obtain many bootstrap draws

$$\hat{\theta}^{*(j)}, \quad j = 1, \ldots, B$$

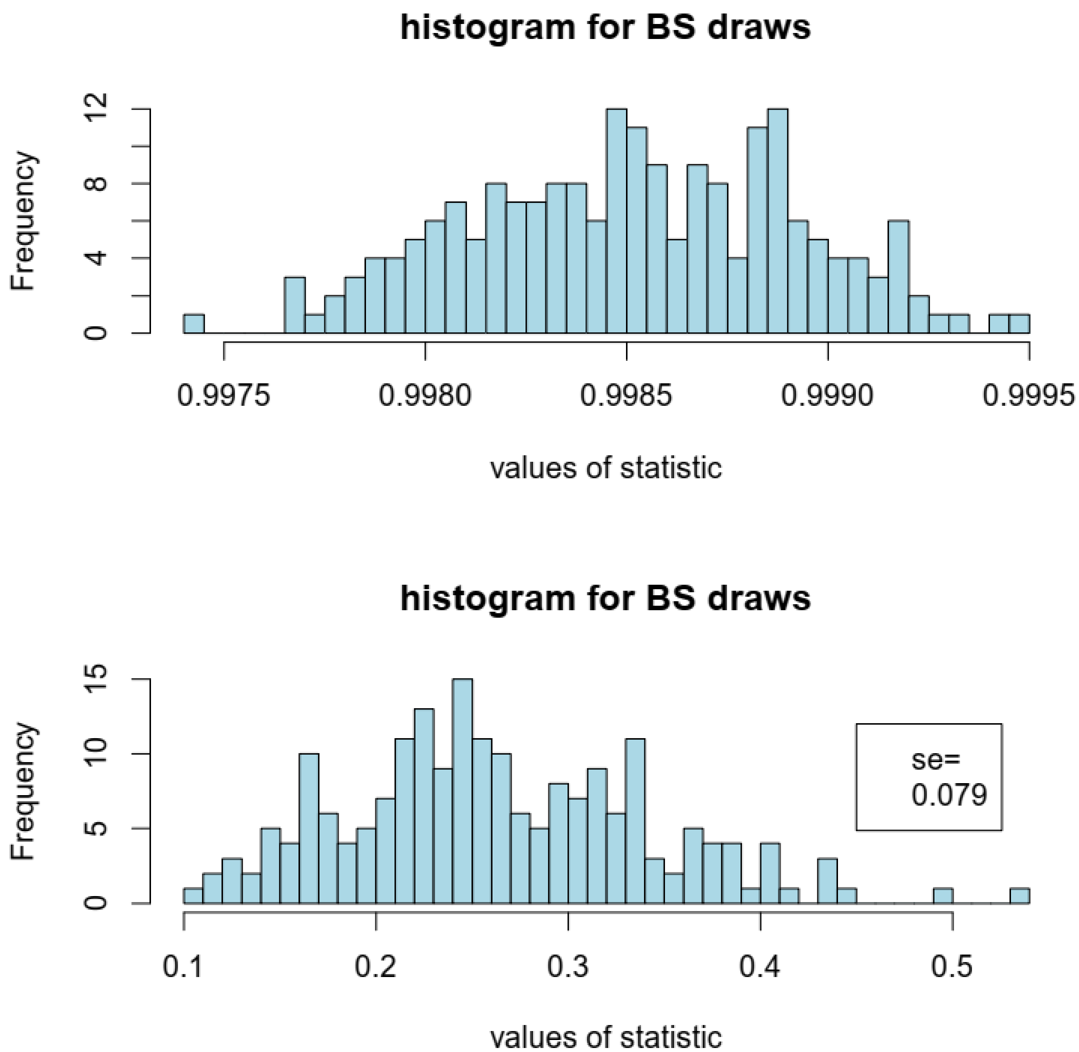


FIGURE 3. Bootstrap Results for GMM Estimators of the discount factor β and risk aversion parameter α in Hansen and Singleton's model for the baseline specification with instruments given by the first lags only. The bootstrap standard errors are quite a bit higher than the analytical standard error. This occurs because of the thick upper tail.

---

of the estimator $\hat{\theta}$, where index j enumerates the bootstrap draws.

(2) For each k in $1, \ldots, d$, compute the bootstrap variance estimate

$$\hat{s}^2(k) = B^{-1} \sum_{j=1}^{B} (\hat{\theta}_k^{*(j)} - \hat{\theta}_k)^2.$$

(Or use the estimate based on the interquartile range.)

(3) Compute the critical value

$$c(1-a) = (1-a)\text{-quantile of } \left\{ \max_{k\in\{1,\ldots,d\}} |\hat{\theta}_k^{*(j)} - \hat{\theta}_k|/\hat{s}(k), \quad j = 1,\ldots,B \right\}.$$

(4) Report the joint confidence region for θ_0 of level $1-a$ as

$$CR_{1-a} = \times_{k=1}^{d}[\hat{\theta}_k \pm c(1-a)\hat{s}(k)].$$

The justification of this confidence rectangle follows from the definition of the bootstrap validity, using arguments similar to those that we gave in L1 for joint confidence rectangles based on approximate joint normality. Indeed the result follows from the equality of the two events:

$$\{|\hat{\theta}_k^{*(j)} - \hat{\theta}_k|/s(k) \leq c(1-a), \text{ for each } k = 1,\ldots,d\} = \{\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \in A\}$$

where

$$A = \times_{k=1}^{d}[\pm c(1-a)s(k)\sqrt{n}]$$

is a rectangular region and applying definitions of validity of bs. The true standard error $s(k)$ can be replaced by the estimated standard error $\hat{s}(k)$, since estimation error has a vanishing impact due to the smothness property of the Gaussian distribution.

4. Extensions and Other Useful Things

4.1. **Delta Method for Bootstrap.** Here we are interested in some smooth nonlinear transformation $\beta_0 = f(\theta_0)$ of θ_0 that can be consistently estimated by a GMM estimator $\hat{\theta}$ such that

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{a}{\sim} N(0, \Omega).$$

We obtain a natural estimator of β_0 by using the "plug-in principle", namely we plug-in the estimator $\hat{\theta}$ instead of θ_0:

$$\hat{\beta} = f(\hat{\theta}).$$

Next we can think of bootstrapping the estimator $\hat{\beta}$. A natural way to define the bootstrap draw $\hat{\beta}^*$ is to apply the transformation f to the bootstrap draw of the GMM estimator $\hat{\theta}^*$, that is

$$\hat{\beta}^* = f(\hat{\theta}^*).$$

Bootstrapping smooth functionals of vector means provides a valid distributional approximation, namely that

$\sqrt{n}(\hat{\beta}-\beta_0) \overset{a}{\sim} N(0,V_f)$ and $\sqrt{n}(\hat{\beta}^*-\hat{\beta}) \mid X_1^n \overset{a}{\sim} N(0,V_f), \quad V_f = \nabla f(\theta_0)\Omega\nabla f(\theta_0)'.$

The approximate distribution of $\hat{\beta}$ is obtained by the delta-method:

$$\begin{aligned}
\sqrt{n}(\hat{\beta}-\beta_0) &= \nabla f(\bar{\theta})\sqrt{n}(\hat{\theta}-\theta_0) \\
&= [\nabla f(\theta_0)+o_P(1)]\sqrt{n}(\hat{\theta}-\theta_0) \\
&\overset{a}{\sim} \nabla f(\theta_0)N(0,\Omega) = N(0,V_f), \\
V_f &= \nabla f(\theta_0)\Omega\nabla f(\theta_0)',
\end{aligned}$$

where $\bar{\theta}$ stands for a point on the line connecting $\hat{\theta}$ and θ_0 and Ω is the variance of $\sqrt{n}(\hat{\theta}-\theta_0)$. Here we require that $\nabla f(\theta_0)$ to have singular values bounded away from zero, since otherwise the normal approximation here would be poor.

Then we can give a reasoning similar to that used above: conditional on the data X_1^n,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}^*-\hat{\beta}) &= \nabla f(\bar{\theta}^*)\sqrt{n}(\hat{\theta}^*-\hat{\theta}) \\
&= [\nabla f(\hat{\theta})+o_P(1)]\sqrt{n}(\hat{\theta}^*-\hat{\theta}) \\
&\overset{a}{\sim} \nabla f(\theta_0)N(0,\Omega) = N(0,V_f), \\
V_f &= \nabla f(\theta_0)\Omega\nabla f(\theta_0)',
\end{aligned}$$

where $\bar{\theta}^*$ is a point on the line connecting $\hat{\theta}^*$ and $\hat{\theta}$ and Ω is the variance of $\sqrt{n}(\hat{\theta}-\theta_0)$.

4.2. **Bootstrapping Dependent Data.** The idea is to divide data in blocks, where dependence is preserved within blocks, and then bootstrap the blocks, treating them as independent units of observations. Here we provide a brief description of the construction of the bootstrap samples. We refer to [7] for assumptions and theoretical results.

Let's assume that we want to draw a bootstrap sample from a stationary and strongly missing sequence X_1^n. For example, $X_i = -(\hat{G}'\hat{A}\hat{G})^{-1}\hat{G}'\hat{A}(\hat{Z}_i - \bar{\hat{Z}})$ in the quick bs for GMM, whereas X_1^n is the original sample in the slow bs for GMM. The blocks of data can be overlapping or non-overlapping. We focus on the non-overlapping case. Let $X_i^j = (X_i, X_{i+1}, \ldots, X_j)$ for $i < j$, and s be the block size. We assume for simplicity that $n = sb$ for some integer b. We construct the bootstrap sample $X_1^*, \ldots, X_n^*$ by stacking b blocks randomly drawn from $\{X_1^s, X_{s+1}^{s2}, \ldots, X_{s(b-1)+1}^{sb}\}$ with replacement. The block size s should be chosen such that $s \to \infty$ and $s = O(n^{1/3})$ as $n \to \infty$. For GMM problems, [4] and [8] recommend setting the block size equal to the trimming parameter in the estimation of Ω.[4]

Notes

The bootstrap method was introduced by Bradley Efron in [2]. Pioneer work in the development of asymptotic theory for the bootstrap includes [1] and [3]. Hall studied the higher order properties of the bootstrap in [5]. For applications to Econometrics, including GMM, see Horowitz's chapter in Handbook of Econometrics in [7].

5. Problems

Problem 1. Explain in exactly 1 page why empirical bootstrap works for the sample mean and why empirical bootstrap works for GMM (you can use "quick" bootstrap for this purpose).

Problem 2. Obtain bootstrap standard errors for the Hansen and Singleton example analyzed in the previous lecture. Obtain a confidence interval for the risk aversion parameter in that example.

Problem 3. Obtain joint confidence bands via empirical bootstrap for the four treatment effects in the Pennsylvania re-employment experiment in L1.

Problem 4. Modify algorithms in Section 3 for the case of "m out of n" bootstrap. Be careful with scaling. Present a brief reasoning similar to that in Section 2 arguing that "m out of n" bootstrap will also work for GMM.

References

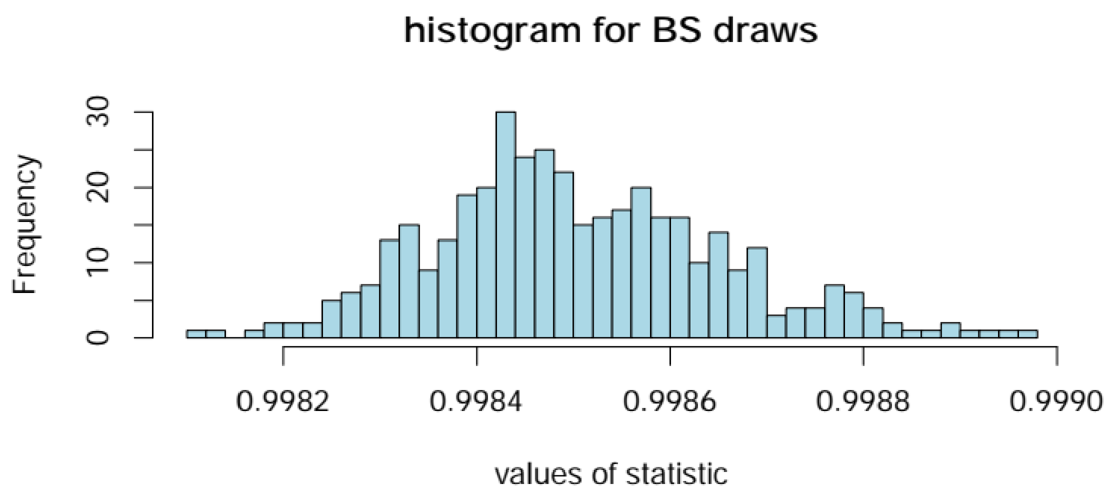
[1] Peter J. Bickel and David A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.
[2] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.
[3] Evarist Giné and Joel Zinn. Bootstrapping general empirical measures. *Ann. Probab.*, 18(2):851–869, 1990.

---

[4] Recall from L4 that the Newey-West estimator of Ω is

$$\hat{\Omega} = \hat{\Sigma}_0 + \sum_{\ell=1}^{L} \omega_{\ell L}(\hat{\Sigma}_\ell + \hat{\Sigma}_\ell'), \quad \hat{\Sigma}_\ell = \sum_{i=1}^{n-\ell} g(X_i, \tilde{\theta})g(X_{i+\ell}, \tilde{\theta})/n, \quad \omega_{\ell L} = 1 - \ell/(L+1),$$

where $\tilde{\theta}$ is a preliminary estimator of θ_0 and L is the trimming parameter.



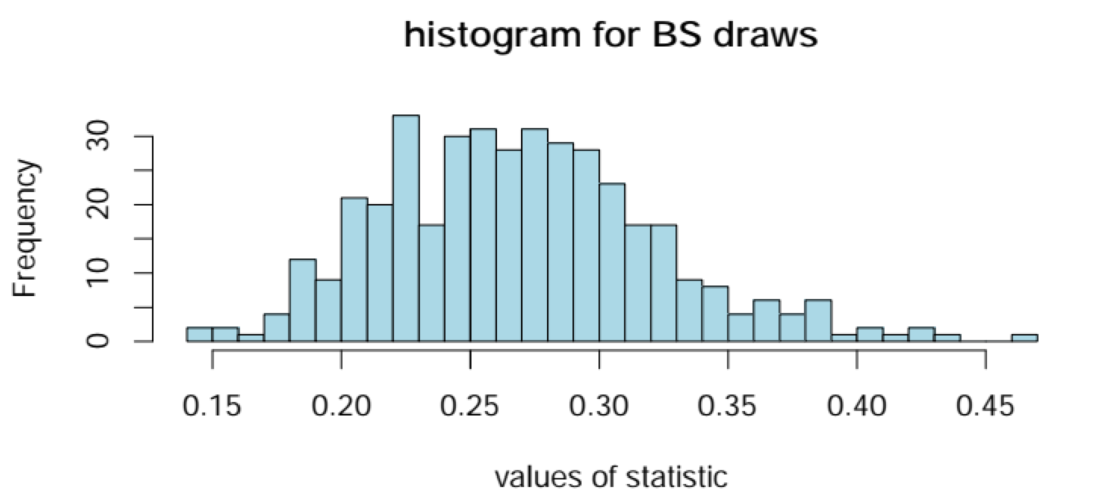


FIGURE 4. (Independent) Bootstrap Results for GMM Estimators of α and β in Hansen and Singleton's model for the baseline specification with instruments given by the first lags only.

[4] F. Götze and H. R. Künsch. Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.*, 24(5):1914–1933, 1996.

[5] Peter Hall. *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag, New York, 1992.

[6] Peter Hall and Joel L. Horowitz. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4):891–916, 1996.

[7] Joel L. Horowitz. The bootstrap. In James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics. Volume 5*. Elsevier: North-Holland, 2001.