

14.384 Time Series Analysis, Fall 2007  
 Professor Anna Mikusheva  
 Paul Schrimpf, scribe  
 September 2, 2007  
 revised October, 2012

## Lecture 9

# Bootstrap

The goal of this lecture is to cover the basics for bootstrap procedures. This lecture is NOT specific to Time series. For some unexplainable reasons bootstrap is missing in our econometric sequence. I have to teach you this topic since we'll be using some bootstrap-based procedures. A good reference is Chapter 52 in Handbook of Econometrics by Horowitz.

## Introduction

We have a sample  $z = \{z_i, i = 1, \dots, n\}$  from a distribution  $F_0$ . We have a statistic of interest,  $T(z)$  whose distribution we want to know because we want to test a hypothesis or to construct a confidence set based on this statistic. Let

$$G_n(t, F_0) = P(T(z) \leq t)$$

be the cdf of  $T(z)$ . In general,  $G_n$  is some complicated function of  $F_0$ . If we knew  $F_0$  we would be able to calculate  $G_n$ . However, in general  $F_0$  is unknown, and this poses the problem.

There are pretty much two ways to get around: asymptotics and simulation techniques. In both cases we want to approximate  $G_n$ .

## Asymptotics

One way is to create an asymptotic approximation (in which the dependence on  $F_0$  disappears) to the distribution  $G_n$  by letting the sample size  $n$  increase to infinity and using CLT and delta-method. If we can get  $G_n(t, F_0) \rightarrow G_\infty(t)$  as  $n \rightarrow \infty$ , then we can pretend that  $G_n(t, F_0) \approx G_\infty(t)$  and use quantiles of the latter to make inferences.

*Example 1.* An example of asymptotic approach is as follows:

Let  $z_i$  be iid with an unknown distribution  $F_0$ , which has an unknown mean  $Ez_i = \mu$ , and variance  $Ez_i^2 = \sigma^2$ . Suppose we want to test a hypothesis  $H_0 : \mu = \mu_0$  and to construct a confidence set for the mean by using t-statistic  $T(z) = \frac{\sqrt{n}}{\hat{\sigma}}(\frac{1}{n} \sum z_i - \mu_0)$ . In general  $T(z)$  has some complicated distribution depending on  $F_0$ . Say, if  $F_0$  is Gaussian, then  $T(z)$  is Student-t distributed, but in general, it's not known. If we knew  $F_0$ , we can simulate  $G_n$ . However, we know that

$$T(z) \Rightarrow N(0, 1)$$

so we can believe that  $G_n$  is very close to  $\Phi$  for large sample size  $n$ . This gives us a justification for using the normal distribution to compute p-values for hypothesis testing and to compute confidence intervals. For example, let  $z_\alpha$  be  $\alpha$ -quantile of the standard normal distribution ( $\Phi(z_\alpha) = \alpha$ ). Then we accept  $H_0$  : if  $Z_{\alpha/2} \leq T(z) \leq Z_{1-\alpha/2}$ . And a 95% confidence interval would be  $[\frac{1}{n} \sum z_i - \hat{\sigma} \sqrt{n} Z_{1-\alpha/2}, \frac{1}{n} \sum z_i + \hat{\sigma} \sqrt{n} Z_{1-\alpha/2}]$ .

## Bootstrap

The bootstrap is another approach to approximating  $G_n(t, F_0)$ . Instead of using the asymptotic distribution to approximate  $G_n(\cdot, F_0)$ , we use  $G_n(\cdot, F_0) \approx G_n(\cdot, \hat{F}_n)$  where  $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x \leq t)$  is the empirical cdf. The reasoning is based on a presumption that  $F_n$  is probably very close to  $F_0$ , and  $G_n$  depends on  $F_0$  in a continuous way.

It is usually numerically cumbersome to calculate how  $G_n$  depends on  $F_0$ . In practice, it is much easier to use simulations to compute  $G_n(\cdot, \hat{F}_n)$ . A general algorithm for the bootstrap with iid data is

1. Generate bootstrap sample  $z_b^* = \{z_{1b}^*, \dots, z_{nb}^*\}$  independently drawn from  $\hat{F}_n$  for  $b = 1..B$ . By this, we mean that  $z_{ib}^*$  are drawn independently with replacement from  $\{z_i\}_{i=1}^n$ .
2. Calculate  $T_b^* = T(z_b^*)$
3.  $G_n^B(t, \hat{F}_n) = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(T_b^* \leq t)$ .

If  $B$  is large, then  $G_n^B(t, \hat{F}_n)$  is very close to  $G_n(t, \hat{F}_n)$  (the simulation accuracy is in our hands). We will let  $Z_q$  be the  $q$ th quantile of  $G_n(t, \hat{F}_n)$ . In our procedure it is  $Z_q \approx T_{(\lfloor B\alpha \rfloor)}^*$ , here  $\lfloor \cdot \rfloor$  is a whole part, and  $T_{(i)}$  is  $i$ -th order statistics.

## Consistency

**Theorem 2.** *Conditions sufficient for bootstrap consistency are:*

1.  $\lim_{n \rightarrow \infty} P_n[\rho(F_0, \hat{F}_n) > \epsilon] = 0, \forall \epsilon > 0$ , where  $\rho(\cdot)$  is some metric (the exact metric depends on the application)
2.  $G_\infty(\tau, F)$  is continuous in  $\tau$
3.  $\forall \tau$  and  $H_n$  s.t.  $\rho(H_n, F) \rightarrow 0$ , we have  $G_n(\tau, H_n) \Rightarrow G_\infty(\tau, F)$

Under these conditions the bootstrap is weakly consistent, i.e.  $\sup_\tau |G_n(\tau, F_n) - G_n(\tau, F_0)| \xrightarrow{P} 0$ .

*Remark 3.* The bootstrap can also be consistent under weaker conditions.

*Remark 4.* We never said that  $G_\infty(\tau, F)$  should be normal, but in the majority of applications it is.

## Asymptotic Refinement

*Example 5.* Consider the same setup as the previous example. Consider calculating two different statistics from the bootstrap:

$$T_1(z) = \sqrt{n}(\bar{z} - \mu)$$

$$T_2(z) = \sqrt{n} \left( \frac{\bar{z} - \mu}{\sqrt{s^2(z)}} \right)$$

The bootstrap analogs of these are

$$T_{1,b}^*(z) = \sqrt{n}(\bar{z}^* - \bar{z})$$

$$T_{2,b}^*(z) = \sqrt{n} \left( \frac{\bar{z}^* - \bar{z}}{\sqrt{s^2(z^*)}} \right)$$

We can use these two statistics to compute two different confidence intervals.

Hall's interval is:  $[\bar{z} - Z_{1-\alpha/2}^1 \frac{1}{\sqrt{n}}, \bar{z} + Z_{\alpha/2}^1 \frac{1}{\sqrt{n}}]$

t-percentile:  $[\bar{z} - \mathcal{Z}_{1-\alpha/2}^2 \sqrt{\frac{s^2(z)}{n}}, \bar{z} + \mathcal{Z}_{\alpha/2}^2 \sqrt{\frac{s^2(z)}{n}}]$

where  $\mathcal{Z}_\alpha^i$  are quantiles of  $T_{i,b}^*(z)$ . Which confidence set may be better? The set based on the first statistic uses that:

$$\begin{aligned} T_1(z) &\Rightarrow N(0, \sigma^2) \quad \text{and} \quad T_1^*(z) \Rightarrow N(0, \sigma^2) \quad \text{as} \quad n \rightarrow \infty \\ T_2(z) &\Rightarrow N(0, 1) \quad \text{and} \quad T_2^*(z) \Rightarrow N(0, 1) \quad \text{as} \quad n \rightarrow \infty \end{aligned}$$

**Definition 6.** A statistic is (asymptotically) *pivotal* if its (asymptotic) distribution does not depend on any nuisance parameters.

The t-statistic (statistic  $T_2(z)$  above) is pivotal,  $T_1(z)$  is not.

The bootstrap usually provides an *asymptotic refinement* if used for a pivotal statistics. It means that the approximation by bootstrap is of higher order than the approximation achieved by asymptotic approach.

In our case, under some technical assumptions (moment and Cramer conditions) we have what's called Edgeworth expansion:

$$P\left(\frac{\bar{z} - \mu}{s(z)} \sqrt{n} \leq t\right) = \Phi(t) + \frac{1}{\sqrt{n}} h_1(t, F_0) + \frac{1}{n} h_2(t, F_0) + O\left(\frac{1}{n^{3/2}}\right)$$

Edgeworth expansion characterize the closeness of  $P\left(\frac{\bar{z} - \mu}{s(z)} \sqrt{n} \leq t\right)$  and  $\Phi(t)$ . We see that the difference between them is of order  $\frac{1}{\sqrt{n}}$ . It is also known that  $h_1(t, F_0)$  is continuous function in  $t$  and  $F_0$  and depend only on first 3 moments of  $F_0$ .

There is also Edgeworth expansion for the bootstrapped distribution:

$$P\left(\frac{\bar{z}^* - \bar{z}}{s(z^*)} \sqrt{n} \leq t\right) = \Phi(t) + \frac{1}{\sqrt{n}} h_1(t, \hat{F}_n) + \frac{1}{n} h_2(t, \hat{F}_n) + O\left(\frac{1}{n^{3/2}}\right)$$

Taking the difference between these two equations we have:

$$P\left(\frac{\bar{z} - \mu}{s(z)} \sqrt{n} \leq t\right) - P\left(\frac{\bar{z}^* - \bar{z}}{s(z^*)} \sqrt{n} \leq t\right) = \frac{1}{\sqrt{n}} \left( h_1(t, F_0) - h_1(t, \hat{F}_n) \right) + O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right)$$

The fact that  $h_1(\cdot)$  is uniformly continuous and  $\hat{F}_n - F_0 = O\left(\frac{1}{\sqrt{n}}\right)$  tells us that  $\frac{1}{\sqrt{n}} \left( h_1(t, F_0) - h_1(t, \hat{F}_n) \right) = O\left(\frac{1}{n}\right)$ . That is, when we bootstrap a pivotal statistic in our simple example, the accuracy of the approximation is  $O\left(\frac{1}{n}\right)$ , whereas the accuracy of the asymptotic approximation is  $O\left(\frac{1}{\sqrt{n}}\right)$ . This gain in accuracy is called asymptotic refinement.

Note that for this argument to work, we needed our statistic to be pivotal because otherwise, the first term in the Edgeworth expansion would not be the same for the true distribution and the bootstrap distribution. Consider:

$$\begin{aligned} P((\bar{z} - \mu) \sqrt{n} \leq t) &= \Phi(t/\sigma) + O\left(\frac{1}{\sqrt{n}}\right) \\ P((\bar{z}^* - \bar{z}) \sqrt{n} \leq t) &= \Phi(t/s(z)) + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

The difference is

$$\begin{aligned} P((\bar{z} - \mu) \sqrt{n} \leq t) - P((\bar{z}^* - \bar{z}) \sqrt{n} \leq t) &= \Phi(t/\sigma) - \Phi(t/s(z)) + O\left(\frac{1}{\sqrt{n}}\right) \\ &\approx \phi(x/\sigma) \frac{1}{\sigma^2} (\sigma - s(z)) + O\left(\frac{1}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

## Bias Correction

Another task for which bootstrap is used is bias-correction. Suppose,  $Ez = \mu$  and we're interested in a non-linear function of  $\mu$ , say  $\theta = g(\mu)$ . One approach would be to take an unbiased estimate of  $\mu$ , say  $\bar{z}$  and plug it into  $g(\cdot)$ ,  $\hat{\theta} = g(\bar{z})$ .  $\hat{\theta}$  is consistent, but it will not be unbiased unless  $g(\cdot)$  is linear. The bias is  $\text{Bias} = E\hat{\theta} - g(\mu)$ . We can estimate the bias using the bootstrap:

1. Generate bootstrap sample,  $z_b^* = \{z_{ib}^*\}$
2. Estimate  $\theta_b^* = g(z_b^*)$
3.  $\text{Bias}^* = \frac{1}{B} \sum_{b=1}^B \theta_b^* - \hat{\theta} \approx \text{Bias}$
4. Use  $\tilde{\theta} = \hat{\theta} - \text{Bias}^*$  as your estimate

Why it works? Let's denote  $G_1(\mu) = \frac{dg(\mu)}{d\mu}$  and  $G_2(\mu) = \frac{d^2g(\mu)}{d\mu^2}$ . Then

$$\hat{\theta} - \theta = G_1(\mu)(\bar{z} - \mu) + \frac{1}{2}G_2(\mu)(\bar{z} - \mu)^2 + o_p\left(\frac{1}{n}\right)$$

$$\text{Bias} = E(\hat{\theta} - \theta) = \frac{1}{2}G_2(\mu)E(\bar{z} - \mu)^2 = \frac{1}{2}G_2(\mu)\frac{\sigma^2}{n} + o\left(\frac{1}{n}\right)$$

similarly

$$\text{Bias}^* = \frac{1}{2}G_2(\bar{z})\frac{s^2}{n} + o_p\left(\frac{1}{n}\right)$$

As a result,

$$\text{Bias}^* - \text{Bias} = o_p\left(\frac{1}{n}\right)$$

*Remark 7.* This procedure required a consistent estimator to begin with.

## Bootstrap do's and don'ts

- If you have a pivotal statistic, bootstrap can give a refinement. So, if you have choice of statistics, bootstrap a pivotal one.
- Bootstrap may fix a finite-sample bias, but cannot help if you have inconsistent estimator.
- In general, if something does not work with traditional asymptotics, the bootstrap cannot fix your problem. For example, if we have an inconsistent estimate, the bootstrap bias correction does not fix anything.
- Bootstrap could not fix the following problems: weak instruments, parameter on a boundary, unit root, persistent regressors.
- Bootstrap requires re-centering (the null hypothesis to be true). The next section is about it.

## Re-centering

The constraints of our model should also be satisfied in our bootstrap replications of the model. For example, assume you are doing estimation using GMM for a population moment condition (that is, you are working under assumption that it is true),

$$Eh(z_i, \theta) = 0$$

When you are simulating bootstrap replications the analogous condition must hold:

$$E^* h(z_i^*, \hat{\theta}) = 0$$

which is equivalent to

$$\frac{1}{n} \sum h(z_i, \hat{\theta}) = 0$$

If the model is overidentified, this condition won't hold. To make it hold, we redefine

$$\tilde{h}(z_i^*, \theta) = h(z_i^*, \theta) - \frac{1}{n} \sum h(z_i, \hat{\theta})$$

and use  $\tilde{h}()$  to compute the bootstrap estimates,  $\theta_b^*$ .

More generally, we need to make sure that our null hypothesis holds in our bootstrap population.

## Bootstrap Variants and OLS

Model:

$$y_t = x_t \beta + e_t$$

we estimate it by OLS and get  $\hat{\beta}$  and  $\hat{e}_t$ . Now we generate bootstrapped samples such that

$$y_t^* = x_t^* \hat{\beta} + e_t^*$$

There are at least three ways to do sampling :

- Take  $z_i = \{(x_i, \hat{e}_i)\}$  ( $x_i$  is always drawn with  $\hat{e}_i$ ). This approach preserves any dependence there might be between  $x$  and  $e$ . For example, if we have heteroskedasticity.
- If  $x_t$  is independent of  $e_t$ , we can sample them independently. Draw  $x_t^*$  from  $\{x_t\}$  and independently draw  $e_t^*$  from  $\{\hat{e}_t\}$ . This is likely to be more accurate than the first approach when we really have independence.
- Parametric bootstrap: Draw  $x_t^*$  from  $\{x_t\}$ , and independently draw  $e_t^*$  from  $N(0, \hat{\sigma}^2)$

*Remark 8.* In time series, all of these approaches might be inappropriate. If  $\{x_t, e_t\}$  is auto-correlated, then these approaches would not preserve the time dependence among errors. One way to proceed is to use the block bootstrap, *i.e.* sample contiguous blocks of  $\{x_t, e_t\}$  together.

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.384 Time Series Analysis  
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.