14.384 Time Series Analysis
Fall 2008

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

Lecture 8

Bootstrap

A good reference is Horowitz (2004)

Introduction

We have a sample $z = \{z_i, i = 1, ..., n\} \sim F_0$ from a distribution F_0. We have a statistic of interest, $T(z)$ whose distribution we want to compute (perhaps because we want to do some hypothesis tests). Let

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

be the cdf of $T(z)$ from a sample of size n. G_n is some complicated function of F_0, which we do not know. We want to approximate G_n. One way to do this would be to use the asymptotic distribution of G_n. Let's say $G_n(t, F_0) \rightarrow G_\infty(t, F_0)$.

Example 1. z_i iid with $Ez_i = \mu$, $Ez_i^2 = \sigma^2$. Suppose we know σ^2 and we're estimating the mean by $T(z) = \frac{1}{n} \sum z_i$. We know

$$\sqrt{n}(T(z) - \mu) \Rightarrow N(0, \sigma^2)$$

so we can use the normal distribution to compute p-values for hypothesis testing and to compute confidence intervals. For example, a 95% confidence interval would be $[T(z) - \sqrt{n}\mathcal{Z}_{1-\alpha/2}, T(z) - \sqrt{n}\mathcal{Z}_{1-\alpha/2}]$, where $\mathcal{Z}_\alpha$ is the α quantile of the normal distribution $N(0, \sigma^2)$.

The bootstrap is another approach to approximating $G_\infty(t, F_0)$.

Bootstrap

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. In practice, we simulate to compute $G_n(\cdot, \hat{F}_n)$. A general algorithm for the bootstrap with iid data is

1. Generate bootstrap sample $z_b^* = \{z_{1b}^*, ..., z_{nb}^*\}$ 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. $G_n(t, \hat{F}_n) = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(T_b^* \leq t)$. We will let $\mathcal{Z}_q$ be the qth quantile of $G_n(t, \hat{F}_n)$

There are many modifications of the bootstrap. In particular, there are many different ways to sample from $\hat{F}_n$.

Consistency

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

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

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: $[\bar{z} - \mathcal{Z}^1_{1-\alpha/2}\frac{1}{\sqrt{n}}, \bar{z} + \mathcal{Z}^1_{\alpha/2}\frac{1}{\sqrt{n}}]$
t-percentile: $[\bar{z} - \mathcal{Z}^2_{1-\alpha/2}\sqrt{\frac{s^2(z)}{n}}, \bar{z} + \mathcal{Z}^2_{\alpha/2}\sqrt{\frac{s^2(z)}{n}}]$
where $\mathcal{Z}^i_\alpha$ are quantiles of $T^*_{i,b}(z)$

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

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

Asymptotic Refinement

The bootstrap usually provides an asymptotic refinement for the distribution of pivotal statistics.

One usually proves this using an Edgeworth expansion (or its analog). In our case, under some technical assumptions (moment and Cramer conditions) we have:

$$P(\frac{\bar{z} - \mu}{s(z)}\sqrt{n} \le t) = \Phi(t) + \frac{1}{\sqrt{n}}h_1(t, F_0) + \frac{1}{n}h_2(t, F_0) + O(\frac{1}{n^{3/2}})$$
$$P(\frac{\bar{z}^* - \bar{z}}{s(z^*)}\sqrt{n} \le t) = \Phi(t) + \frac{1}{\sqrt{n}}h_1(t, \hat{F}_n) + \frac{1}{n}h_2(t, \hat{F}_n) + O(\frac{1}{n^{3/2}})$$

Taking the difference between these two equations we have:

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

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

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

The difference is

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

Bias Correction

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

1. Generate bootstrap sample, $z_b^* = \{z_{ib}^*\}$
2. Estimate $\theta_b^* = g(\bar{z}_b^*)$
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

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

Remark 8. 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. Also, if we have weak instruments, so that our asymptotic distribution is a poor approximation, then the bootstrap also gives us a poor approximation.

Bootstrap and GMM

The constraints of our model should also be satisfied in our bootstrap replications of the model. For example, with GMM our population moment condition,

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

should hold in our bootstrap replications,

$$\begin{aligned} E^*h(z_i^*, \hat{\theta}) = 0 \\ \frac{1}{n}\sum h(z_i, \hat{\theta}) = 0 \end{aligned}$$

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, θ_b^*.

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

Bootstrap Variants and OLS

Model:

$$y_t = x_t \beta + e_t$$

estimate by OLS $\hat{\beta}$ and $\hat{e}_t$. Now we want to bootstrap. There are at least three ways to sample:

- 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 9. 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.