INTRODUCTION TO BOOTSTRAP TIME SERIES

Date: March 21, 2018

JEN-WEN LIN, PhD, CFA

INTRODUCTION

- “Bootstrap” generally refers to replicating an experiment, by resampling from observed data
- A parametric model may be specified
- Use to estimate bias, se, percentiles, or other properties of a statistic
- Compute bootstrap confidence intervals for quantities of interest
- Data and examples from Davison and Hinkley (1997)

GENERAL GUIDELINES FOR USING THE BOOTSTRAP APPROACH

The basic bootstrap approach consists of drawing repeated samples (with replacement)

Time series models

iid assumption

the simplest one that is valid for IID observations

IID assumption is **not** satisfied

so we need to solve

The method needs to be modified

BIGCITY DATA (BOOT PACKAGE)

- Population (in 1000's) of 49 U.S. cities 1920 (u) and 1930 (x). The 49 cities are a random sample taken from the 196 largest cities in 1920
- Parameter of interest is ratio of means $\theta = E(X)/E(U)$. Let $\hat{\theta} = \bar{X}/\bar{U}$.
- No obvious parametric model and we want to known uncertainty in $\hat{\theta}$.

```
library(boot)
summary(bigcity)
mean(bigcity[,2])/mean(bigcity[,1])
plot(bigcity,xlab="1920 pop.", ylab="1930 pop.", cex=.6)
grid()
colors <- densCols( bigcity)
smoothScatter( bigcity, col = colors, pch = 20 )
grid()
```

By resampling, we can get a lot of observations

HOW TO SAMPLE BIVARIATE OR MULTIVARIATE DATA

- Generate rows indices i using sample function
- Extract corresponding rows of data $dat[i,]$

seed?

```
## first 4 rows in data set
dat<-bigcity[1:4,]; dat
set.seed(1234)  ## control the outcome of random number generators
i<-sample(1:4,size=4,replace=TRUE)
i
dat[i,]
```

RATIO STATISTICS

- Parameter of interest is ratio of the means $\theta = E(X)/E(U)$
- Statistic: $\hat{\theta} = \bar{X}/\bar{U}$
- Function to compute $T = \hat{\theta}$

```
## function to compute theta.hat
 ratio<-function(X) mean(X[,2])/mean(X[,1])
```

BOOTSTRAP REPLICATES

- Sample size n, parameter θ, statistic $\hat{\theta}$
- Resample n observations with replacement and compute $\hat{\theta}$ for each sample

```
## Illustrated for bigcity data and ratio of means
 n<-NROW(bigcity)
 t0<-ratio(bigcity)
 set.seed(1234)
 t.replicates<-replicate(1000, expr={
   i<-sample(1:n, size=n, replace=TRUE)
   boots.obs<-bigcity[i,]
   ratio(boots.obs)
 })
 library(MASS); truehist(t.replicates); points(t0,0,pch=17,col=2)
```

如果不做*bootstrap*，需要*assume population normal*，才可以做*inference*

BOOTSTRAP STANDARD ERROR

$$\widehat{var}(\hat{\theta}) = \frac{1}{R-1}\sum_{j}^{R}\left(\hat{\theta}^{(j)} - \bar{\hat{\theta}}^{\star}\right)^2,$$

where $\bar{\hat{\theta}}^{\star}$ is the arithmetic mean of the replicates. (Compute sd of the replicates.)

```
## Bootstrap estimate of standard error
 mean(t.replicates)
 se.hat<-sd(t.replicates);
se.hat
```

BOOTSTRAP ESTIMATE OF BIAS

- The bias of $\hat{\theta}$ for θ is $E\left(\hat{\theta}\right) - \theta$
- Bootstrap estimate of bias is
- $\widehat{bias}\left(\hat{\theta}\right) = \bar{\hat{\theta}}^{\star} - \hat{\theta}$
- What are our estimates of $\hat{\theta}$, $\bar{\hat{\theta}}^{\star}$, $\widehat{se}(\hat{\theta})$, and $\widehat{bias}(\hat{\theta})$?

USING THE BOOT FUNCTION

- `boot` **function is in the boot package**
- **For ordinary bootstrap,** `boot` **has the basic syntax:**
- `boot` **(**`data`**,** `statistic`**,** `R`**)**
- `statistic` **is the name of a function to compute the statistic of interest**
- `R` **is the number of replicates to generate**
- `Boot` **does the job of generating the same indices and passes them to the** `statistic` **function**

SYNTAX OF THE STATISTIC FUNCTION

- First argument to `statistic` is the data (`matrix` or `dataframe`)
- Second argument to `statistic` is the indices of the sample
- Body of function extracts the sample from the data and compute the statistic of interest
- Modify our `ratio` function for use with `boot`:

```
ratio.boot<-function(X, i){
  y<-X[i,]
  mean(y[,2])/mean(y[,1])
}
```

BOOTSTRAP WITH BOOT

- Now we could revise our bootstrap code to use `ratio.boot`

```
out<-boot(bigcity, statistic=ratio.boot, R=999);  out
truehist(out$t, main="Replicate generated   by boot function")
points(t0,0,pch=17,col=2)
out$t0
mean(out$t)
```

- We can compute percentiles of the bootstrap replicates

```
## to compute percentiles of the bootstrap replicates
## the 2.5 and 97.5 percentiles determine a 95% bootstrap CI
  quantile(out$t,c(0.025,0.975))
  quantile(out$t,c(0.025,0.975),type=1)
## type=1 is the inverse ECDF method for computing percentiles
```

THE BOOT.CI FUNCTION

- `boot.ci` computes five types of bootstrap confidence intervals using output from `boot`
  - Normal CI (`norm`)
  - Percentile CI (`perc`)
  - Basic Bootstrap (`basic`)
  - Studentized "t" CI (`stud`)
  - BCa CI (`bca`)
- Percentile, Basic, and BCa are percentile intervals
- BCa is a "better" bootstrap CI

```
## boot.ci computes 5 types of bootstrap CI using output from boot
boot.ci(out,type=c("norm","perc","basic","bca"))
```

从而做 *autoregression*

BOOTSTRAPPING REGRESSION

- Consider the linear regression model

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \qquad E(u_t|\boldsymbol{X}_t) = 0, \qquad u_t \sim IID(0, \sigma^2),$$

- where there are *n* observations and *k* regressors. Regressors *may include lagged dependent variables*, but y_t is not explosive and does not have a unit root.

stationary

- There are many ways to bootstrap the above regression model. In general, making stronger assumptions results in better performance if those assumptions are satisfied, but it leads to asymptotically invalid inferences if they are not.
- The assumptions we make in bootstraping include
  1. Are the errors independent?
  2. Are the errors identically distributed?

RESIDUAL BOOTSTRAP

不强调 *error term*的*distribution*

- Required that the errors be independent of contemporaneous regressors and IID, but with minimal distributional assumptions.
- Steps of residual bootstrap include:
  1. Use OLS to obtain $\widehat{\boldsymbol{\beta}}$ and $\hat{u}_t$.
  2. (Optional) rescale residuals so that they have correct variance. For example, the simplest rescaled residual is

$$\ddot{u}_t \equiv \left({}^{n}\!/_{n-k}\right)^{1/2} \hat{u}_t.$$

  3. Generate a typical observation of the bootstrap sample as

empiralcal distribution

$$y_t^* = \boldsymbol{X}_t\widehat{\boldsymbol{\beta}} + u_t^*, \qquad u_t^* \sim EDF(\ddot{u}_t).$$

The u_t^* are often to said to be resampled from $\ddot{u}_t$.

PARAMETRIC BOOTSTRAP

- Assume that u_t follow a specific distribution, say normal distribution.
- Steps of parametric bootstrap include
  1. Use OLS to obtain $\widehat{\boldsymbol{\beta}}$ and $\hat{u}_t$.
  2. Generate a typical observation using

$$y_t^* = \boldsymbol{X}_t\widehat{\boldsymbol{\beta}} + u_t^*, \qquad u_t^* \sim NID(0, s^2),$$

  where s^2 denotes the sample variance of $\hat{u}_t$.

WILD BOOTSTRAP

- Specifically designed to handle heteroskedasticity in regression models.
- The wild bootstrap DGP is

$$y_t^* = \boldsymbol{X}_t\widehat{\boldsymbol{\beta}} + f(\hat{u}_t)v_t^*,$$

where $f(\hat{u}_t)$ is a transformation of the t-th residual $\hat{u}_t$, and v_t^* is a random variable with mean 0 and ~~variable~~ variance 1.

- A good choice for $f(\cdot)$ is $f(\hat{u}_t) = \hat{u}_t/\sqrt{1 - h_t}$, where h_t is the t-th diagonal of the hat matrix.
- There are various ways to specify the distribution of the v_t^*. The simplest is $v_t^* = \pm 1$, each with probability of 0.5.

PAIR BOOTSTRAP

- Proposed by Freedman (1981, 1984); see also Freedman and Peters (1984).
- Resample from the matrix with typical row $[y_t, \boldsymbol{X}_t]$. We no longer condition on the $\boldsymbol{X}_t$, since each bootstrap sample now has a different $\boldsymbol{X}$ matrix. A typical observation of the bootstrap sample is $[y_t^*, \boldsymbol{X}_t^*]$.
  1. The pairs bootstrap is valid even when the errors display heteroskedasticity of unknown form.
  2. It works even for dynamic models. If regressors include lagged dependent variables, we treat them like any other element of $\boldsymbol{X}_t$.
  3. Pairs bootstrap can be applied to an enormous range of models.
  4. In the case of multivariate models, we can combine the pairs and residual bootstraps. Organize residuals as a matrix and apply the pairs bootstrap to its rows. This preserves cross-equation correlations.

MORE PAIR BOOTSTRAP

- Unfortunately, the pairs bootstrap has two major deficiencies:
  - (1) If the null hypothesis imposes restrictions on $\boldsymbol{\beta}$, the bootstrap DGP does not impose them. We must therefore modify the bootstrap test statistic so that it is testing something which is true in the bootstrap DGP;
  - (2) Compared to residual bootstrap (when it is valid) and wild bootstrap, pairs bootstrap does not yield very accurate results.

BOOTSTRAP FOR DEPENDENT DATA

- All bootstrap discussed so far assume that the errors are independent
- Resampling breaks up any dependence and is therefore inappropriate for dependent data
- For dependent data, two of the popular approaches are **sieve bootstrap** and **block bootstrap**.

PARAMETRIC BOOTSTRAPPING

- Simulate unconditional ARMA (p,q) model (Mcleod and Hipel, 1978)
- Consider a stationary AR(1) model
$$X_t = \alpha + \phi X_{t-1} + a_t, \qquad a_t \sim NID(0, \sigma_a^2). \qquad (1)$$
- The unconditional distribution of X_t is given by
$$X_t \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_a^2}{1-\phi^2}\right), \qquad (2)$$
- The conditional distribution of X_t given X_{t-1} is given by
$$X_t | X_{t-1} \sim N(\phi X_{t-1}, \sigma_a^2). \qquad (3)$$
- The (unconditional) simulation procedure may be summarized as follow:
  1. Simulate X_0 by drawing a random number from eqn. (2);
  2. Simulate $X_1 = \alpha + \phi X_0 + a_t$, where X_0 is obtained from Step 1;
  3. Simulate $X_t = \alpha + \phi X_{t-1} + a_t, t = 1,2, \ldots$, recursively.

THE SIEVE BOOTSTRAP

- Suppose that the error term u_t in a regression model follow an unknown, stationary process with homoskedastic innovations.
- The sieve bootstrap approximates this process using an AR(p) process with p chosen by some sort of model selection criterion (like AIC or BIC), or by sequential testing.

STEPS OF SIEVE BOOTSTRAP

1. Estimate the model to obtain residuals $\hat{u}_t$;
2. Estimate AR(p) model

$$\hat{u}_t = \sum_{i=1}^{p} \phi_i \hat{u}_{t-i} + \varepsilon_t, \qquad (1)$$

for several values of p and choose best one. [need to ensure stationarity]

3. Generate bootstrap error terms

$$u_t^* = \sum_{i=1}^{p} \hat{\phi}_i u_{t-i}^* + \varepsilon_t^*, \qquad (2)$$

where the ε_t^* are resampled from the (rescaled) residuals from eqn. (1).

4. Generate the bootstrap data according to

$$y_t^* = \boldsymbol{X}_t \widehat{\boldsymbol{\beta}} + u_t^*.$$

Remark: Sieve bootsrap assume IID innovations, thus ruling out GARCH and other forms of heteroskedasticity.

GENERAL ERROR STRUCTURES – THE MBB

The Moving Block Bootstrap

Background

Application of the **residual based bootstrap** methods is straightforward if the error distribution is specified to be an ARMA(p,q) process with known p and q

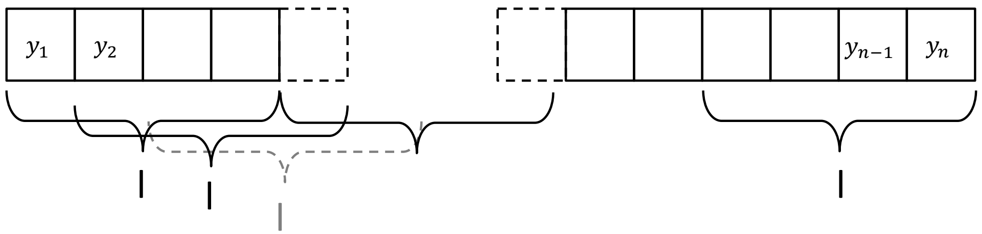
However, if the structure of serial correlation is not tractable or is misspecified, the **residual based methods will give inconsistent estimates**

Carlstein (1986) – first discussed the idea of **bootstrapping blocks of observations** rather than the individual observations. **The blocks are nonoverlapping**

Künsch (1989) and Singh (1992) – independently introduced a more general BS procedure, **the moving block BS** (MBB) which is applicable to **stationary time series data**. In this method the **blocks of observations are overlapping**.

The Moving Block Bootstrap

Divide the data of n observations into blocks of length l and select b of these blocks (with repeats allowed) by resampling with replacement all the possible blocks



For simplicity assume $n = bl$

In the *Carlstein* procedure: $\frac{n}{l} = b$ blocks

In the *Künsch* procedure: $n - l + 1$ blocks

The k^{th} block is $L_k = \{x_k, \dots, x_{k+l-1}\}$

$$k = 1,2, \dots, (n - l + 1)$$

For example with n = 6 and l = 3 suppose the data are: xt = {3,6,7,2,1,5).
The blocks according to Carlstein are {(3,6,7), (2,1,5)). The blocks according to Kiinsch are {(3,6,7), (6,7,2), (7,2,1), (2,1,5)).

The Moving Block Bootstrap

EXAMPLE

$$x_t = \{\ \boxed{3}\boxed{6}\boxed{7}\boxed{2}\boxed{1}\boxed{5}\ \}$$

The blocks in the *Carlstein* procedure are:

$\boxed{3}\boxed{6}\boxed{7}$ $\boxed{2}\boxed{1}\boxed{5}$

The blocks in the *Künsch* procedure are:

$\boxed{3}\boxed{6}\boxed{7}$ $\boxed{6}\boxed{7}\boxed{2}$ $\boxed{7}\boxed{2}\boxed{1}$ $\boxed{2}\boxed{1}\boxed{5}$

Draw a sample of **two blocks** with replacement in each case

Suppose, the first draw gave $\boxed{3}\boxed{6}\boxed{7}$ (WLOG)

Then, the probability of missing all of $\boxed{2}\boxed{1}\boxed{5}$ is:

Carlstein: 50%

Künsch: 25%

Higher probability of missing entire blocks in the *Carlstein* scheme (non overlapping blocks)

Carlstein scheme is not popular and not often used

Problems with MBB

There are some important problems worth noting about the MBB procedure

1. The pseudo time series generated by the moving block method is not stationary, even if the original series $\{x_t)$ is stationary

Politis and Romano (1994)

A stationary bootstrap method
The suggested method involves sampling blocks of random length, where the length of each block has a geometric distribution. They show that the pseudo time series generated by the stationary bootstrap method is indeed stationary.

Problems with MBB

There are some important problems worth noting about the MBB procedure

2. The mean $\bar{x}_n^*$ of the moving block bootstrap is biased in the sense that:

$$E(\bar{x}_n^* | x_1, x_2, \dots, x_n) - \bar{x}_n \neq 0$$

3. The MBB estimator of the variance of $\sqrt{n} \cdot \bar{x}_n$ is also biased

Davidson and Hall (1993)

This creates problems in using the percentile-t method with the MBB

the usual estimator:

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$$

Should be modified to:

$$\tilde{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \left\{ (x_i - \bar{x}_n)^2 + \sum_{k=1}^{i-1} \sum_{i=1}^{n-k} (x_i - \bar{x}_n)(x_{i+k} - \bar{x}_n) \right\}$$

With this modification the bootstrap-t can improve substantially on the normal approximation

Optimal Length of Blocks

Several rules that have been suggested are based on different criteria. However, the rules are useful as rough guides to selecting the optimal sized blocks

1. *Carlstein's* non-overlapping blocks < *Künsch's* moving blocks

2. *Politis and Romano's* stationary bootstrap method
   - The average length of a block is $\frac{1}{p}$, where p is the parameter of the geometric distribution
   - The application of **stationary bootstrap** is less sensitive to the choice of p than the application of **moving block bootstrap** is to the choice of l

BLOCK BOOTSTRAP METHODS

- Block bootstrap methods divide the quantities that are being resampled, which might be either rescaled residuals or $[y, \boldsymbol{X}]$ pairs, into blocks of b consecutive observations. We then resample the blocks.
- Blocks may be either overlapping or nonoverlapping; overlapping seems to be better.
- Block lengths may be fixed or variable; fixed seems to be better.
- For the moving-block bootstrap, there are $n - b + 1$ blocks. The first contains obs. 1 through b, the second contains obs. 2 through $b + 1$, and the last contains obs. $n - b + 1$ through n.
- Choice of b is critical. In theory, it must increase as n increases. Often proportional to $n^{1/3}$.

BLOCK BOOTSTRAP CONT'D

- If blocks are too short, bootstrap samples cannot mimic original sample. Dependence is broken whenever we start a new block. If blocks are too long, bootstrap samples are not random enough.
- The **block-of-blocks** bootstrap is the analog of the pairs bootstrap for dynamic models.
- Consider the dynamic regression model

$$y_t = \boldsymbol{X}_t \boldsymbol{\beta} + \gamma y_{t-1} + u_t, \qquad u_t \sim IID(0, \sigma^2).$$

- Let's define

$$\boldsymbol{Z}_t \equiv [y_t, y_{t-1}, \boldsymbol{X}_t].$$

BLOCK BOOTSTRAP CONT'D

- We can then construct $n - b + 1$ overlapping blocks as

$$\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_b$$

$$\boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_{b+1}$$

$$\ldots \ldots \ldots$$

$$\boldsymbol{Z}_{n-b+1}, \ldots, \boldsymbol{Z}_n$$

- The block-of-blocks bootstrap works with heteroskedasticity as well as serial correlation
- Although block bootstrap methods frequently offer higher-order accuracy than asymptotic methods, they generally do so to only a modest extent.
- Block bootstrap can yield more reliable standard errors than using HAC covariance matrices