BOOTSTRAP TIME SERIES AND BAGGING

Date: 08 August, 2018

JEN-WEN LIN, PhD, CFA

GENERAL GUIDELINES FOR USING BOOTSTRAP

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

the simplest one that is valid for IID observations

Time series models

IID assumption is **not** satisfied

The method needs to be modified

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.
- 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 bootstrapping include
  1. Are the errors independent?
  2. Are the errors identically distributed?

PARAMETRIC BOOTSTRAP

- Assume that u_t is independent and follows 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 $\hat{u}_t$.

RESIDUAL BOOTSTRAP

- 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
  $$y_t^* = \boldsymbol{X}_t \widehat{\boldsymbol{\beta}} + u_t^*, \qquad u_t^* \sim EDF(\ddot{u}_t).$$
  The u_t^* are often to said to be resampled from $\ddot{u}_t$.

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 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 (BLOCK) 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 ARMA MODEL

- 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

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

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

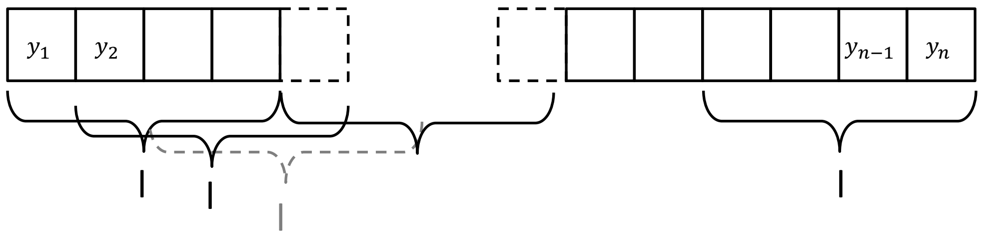
$$y_t^* = \boldsymbol{X}_t \hat{\boldsymbol{\beta}} + u_t^*.$$

THE MOVING BLOCK BOOTSTRAP

- 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 $L_k = \{x_k, \ldots, x_{k+l-1}\}$

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

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

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), "stationary bootstrap"*
2. The mean $\bar{x}_n^*$ of the moving block bootstrap is biased in the sense that $E(\bar{x}_n^* | x_1, x_2, \ldots, x_n) - \bar{x}_n \neq 0$
3. The MBB estimator of the variance of $\sqrt{n} \cdot \bar{x}_n$ is also biased

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 I

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

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

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

$$\dots \dots \dots$$

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

- 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

BAGGING (BREIMAN, 1996)

- Bagging is short for Bootstrap Aggregation and is designed for situations in which the number of predictors is moderately large relative to the sample size.
- In essence, bagging involves fitting the unrestricted model including all potential predictors to the original sample, generating a large number of bootstrap resamples from this approximation of data, applying the decision rule for each bootstrap sample.
- By averaging across resamples, bagging effectively removes the instability of decision rules. Hence, one would expect the variance of the bagged prediction model to be smaller than that of the model that would be selected based on the original data.
- In contrast, the forecast bias of the prediction model is likely to be of similar magnitude, with or without bagging.

BAGGING TIME SERIES MODELS
INOUE AND KILIAN(2004)

1. Arrange the set of tuples $\{(y_{t+h}, x_t')\}, t = 1, \ldots, T - h$, in the form of a matrix dimension $(T - h)\times(M + 1)$

$$\begin{matrix} y_{1+h} & x_1' \\ \vdots & \vdots \\ y_T & x_{T-h}' \end{matrix}$$

2. Construct bootstrap samples by drawing with replacement blocks of m rows of the above matrix, where the block size m is chosen to capture the dependence in the error term (e.g. Hall and Horowitz 1996, Goncalves and White, 2003).

3. For each bootstrap sample, compute the bootstrap pre-test predictor conditional on x_{T-h+1}

$$\hat{y}^{*PT}(x_{T-h+1}) = 0, \text{ if } |t_j^*| < 1.96\ \forall j \text{ and } \hat{y}^{*PT}(x_{T-h+1}) = \hat{\gamma}^{*\prime} S_T^* x_{T-h+1} \text{ otherwise}$$

where $\hat{\gamma}^*$ and S_T^* are the bootstrap analogues of $\hat{\gamma}$ and S_T (to define later).

BAGGING TIME SERIES MODELS
INOUE AND KILIAN(2004)

4. The bagged predictor is the expectation of the bootstrap pre-test predictor across bootstrap samples, conditional on x_{T-h+1}. In practice, it is calculated as

$$\hat{y}^{BAGGING}(x_{T-h+1}) = \frac{1}{B}\sum_{i=1}^{B} \hat{\gamma}^{*i\prime} S_T^{*i} x_{T-h+1}.$$

The authors suggest that in practice, $B = 100$ tends to provide a reason approximation.

- S_T is the stochastic selection matrix obtain from

$$\begin{bmatrix} I(|t_1| > 1.96) & 0 & \dots & 0 \\ 0 & I(|t_2| > 1.96) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I(|t_M| > 1.96) \end{bmatrix}$$

by deleting rows.