Applied Econometrics for Macro and Finance

Bootstrap

Shiu-Sheng Chen

Department of Economics
National Taiwan University

Spring 2023

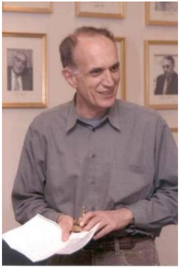
Readings

- Chapter 13 in Chen (2022)
- MacKinnon, James G. (2007) "Bootstrap Hypothesis Testing", working paper
- Davidson, Russell and MacKinnon, James G (2006) "The Power of Bootstrap and Asymptotic Tests", Journal of Econometrics
- MacKinnon, James G (2006) "Bootstrap Methods in Econometrics", Economic Record, Special Issue, v.82, pp. S2–18

The Bootstrap

- Bradley Efron (1979, 1981, 1982) and Efron and Tibshirani (1993)
- Bootstrap means that one available sample gives rise to many others by resampling.
    - In general, bootstrap is developed for inferential purposes: standard errors, confidence intervals, and hypothesis testing (Efron, 1981, 1982).

Bradley Efron



- Professor of Statistics and Biostatistics, Stanford University
- Best known for proposing the bootstrap resampling technique
- National Medal of Science, the highest scientific honor by the United States

The Bootstrap

- Confidence intervals, hypothesis testing, and standard errors are all based on the idea of the sampling distribution (or asymptotic distribution) of a statistic.
  - In many settings, we have no model for the population to construct exact sampling distribution.
  - Furthermore, we cannot obtain enough sample to ensure the large sample theory works (the small sample problem).
- The bootstrap may help us out of these troubles.
- Before discussing what bootstrap is, let's look at an example showing the poor performance of asymptotic approximation.

An Example of the Poor Performance of Asymptotic Approximation

- Consider the following regression model

$$\{y_i, x_{1i}, x_{2i}\}_{i=1}^{n}$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,$$

$$\begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right),$$

$$e_i \sim N(0, 9),$$

and $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 0.5$, n=300.

Poor Performance of Asymptotic Approximation

- The parameter of interest is

$$\theta = \frac{\beta_1}{\beta_2}.$$

- Hence, the true value of θ is

$$\theta_0 = \frac{1}{0.5} = 2.$$

Poor Performance of Asymptotic Approximation

- Estimate θ by $\hat{\theta} = \frac{\hat{\beta}_1}{\hat{\beta}_2}$.
- According to Delta Method,

$$t(\hat{\theta}) = \frac{\hat{\theta} - \theta}{S_n(\hat{\theta})} \xrightarrow{d} N(0,1)$$

where

$$S_n(\hat{\theta}) = \sqrt{(\hat{H}_\beta' \hat{V} \hat{H}_\beta)}, \quad \hat{H}_\beta = \begin{pmatrix} 0 \\ 1/\hat{\beta}_2 \\ -\hat{\beta}_1/\hat{\beta}_2^2 \end{pmatrix},$$

and

$$\hat{V} = \left[\frac{1}{n}\sum_i x_i x_i'\right]^{-1} \left[\frac{1}{n}\sum_i x_i x_i' \hat{e}_i^2\right] \left[\frac{1}{n}\sum_i x_i x_i'\right]^{-1}$$

is the estimated variance-covariance matrix.

Aside: The Delta Method

Suppose that $\mathbf{W_n}$ is a sequence of random k-vector such that

$$\sqrt{n}(\mathbf{W_n} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

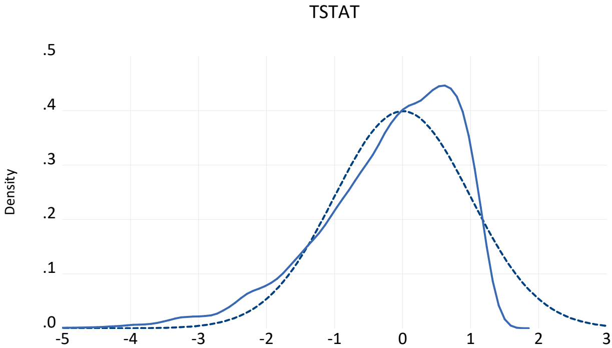
where Σ is positively definite. Let $g(\cdot) : \mathbb{R}^k \mapsto \mathbb{R}$ be a function such that

$$\nabla g(\mathbf{y}) = \frac{dg(\mathbf{y})}{d\mathbf{y}} = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \frac{\partial g}{\partial y_2} \\ \vdots \\ \frac{\partial g}{\partial y_k} \end{pmatrix},$$

and $\nabla g(\boldsymbol{\theta})$ is non-zero and continuous at $\boldsymbol{\theta}$. Then

$$\sqrt{n}(g(\mathbf{W_n}) - g(\boldsymbol{\theta})) \xrightarrow{d} N(0, \nabla g(\boldsymbol{\theta})' \Sigma \nabla g(\boldsymbol{\theta})).$$

Poor Performance of Asymptotic Approximation



- Distributions of $t(\hat{\theta})$: Exact Distribution (solid line) and Asymptotic Distribution (dashed line, $N(0,1)$)
- Size= $P(|t| > 1.96) = 6.53\% > 5\%$

Definition of the Bootstrap

- Assume data $\{X_i\}_{i=1}^n$ come from an unknown distribution function F.
- Let

$$T_n = T_n(X_1, \ldots, X_n, F)$$

  be a statistic of interest.
- Note that in most cases the statistic is written as

$$T_n = T_n(X_1, \ldots, X_n, \theta),$$

  where θ is an unknown parameter.

Definition of the Bootstrap

For example,

- Estimator

$$T_n = \hat{\theta}$$

- Bias

$$T_n = \hat{\theta} - \theta$$

- t-statistic

$$T_n = \frac{(\hat{\theta} - \theta)}{S(\hat{\theta})}$$

Definition of the Bootstrap

- Let $G_n(\tau, F) = P(T_n \le \tau | F)$ be the exact sampling distribution function of T_n.
- Ideally, inference would be based on exact sampling distribution, $G_n(\tau, F)$, but this is generally impossible when F is unknown.
- Asymptotic inference is based on approximating $G_n(\tau, F)$ with

$$G_\infty(\tau, F) = \lim_{n \to \infty} G_n(\tau, F)$$

- If $G_\infty(\tau, F) = G_\infty(\tau)$ does not depend on F, we say that T_n is asymptotically pivotal and use the distribution function $G_\infty(\tau)$ for inferential purposes.
    - For instance, $N(0, 1)$. CLT is asymtotically pivotal

Definition of the Bootstrap

The problems of Asymptotic Approximation are

- In most applications, asymptotic pivotal statistics are not available.
- Moreover, even if the asymptotic pivotal statistic is available, the asymptotic approximation may be very poor as shown in the above example.

Definition of the Bootstrap

- Efron (1979) proposed a different approximation: the bootstrap.
- The bootstrap method can be used even when T_n is complicated to compute and difficult to analyze.
- It is not necessary for T_n to have a known asymptotic distribution.

Definition of the Bootstrap

i.i.d. is essential to use Bootstrap

- The idea of bootstrap is simple: use the one sample we have as though it were the population, taking many resamples from it to construct the bootstrap distribution.
- Then in statistical inference, use the bootstrap distribution in place of the sampling distribution.

Definition of the Bootstrap

- Efron (1979) proposed that first estimate F by a consistent estimate F_n.

Fn is the sampling distribution of your sample, which is empirical distribution / cdf

- Then plug F_n into $G_n(\tau, F)$ to obtain

$$G_n^*(\tau) = G_n(\tau, F_n)$$

as an estimate of $G_n(\tau, F)$.

- $G_n^*(\tau)$ is called the bootstrap distribution.
- The bootstrap inference is based on $G_n^*(\tau)$.

Definition of the Bootstrap

Here I give you a rough idea about why Bootstrap works.

- Recall that $F(x) = P(X_i \leq x) = E(1(X_i \leq x))$.
- A natural choice of F_n is the empirical distribution function (EDF):

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x).$$

indicator function

- By the WLLN, $F_n(x) \xrightarrow{p} F(x)$, a consistent estimator.
- Hence, under some conditions, we have

$$\lim_{n\to\infty} G_n^*(\tau) = \lim_{n\to\infty} G_n(\tau, F_n) = G_n(\tau, F)$$

Definition of the Bootstrap

- Here, we have use a very sloppy notations and descriptions to give you some ideas about the consistency of the bootstrap. For rigorous treatments, see Horowitz (2001).
- Although some unusual conditions may cause inconsistency, Horowitz (2001) suggests that the bootstrap is consistent in most applications in econometric practice.

Nonparametric Bootstrap

- Efron (1979) proposed a Monte Carlo simulation to approximate G_n^*.
- The procedure is as follows.

Nonparametric Bootstrap: Step 1

- Draw a bootstrap sample $\{X_i^*\}_{i=1}^n$ from $F_n(x) = \frac{1}{n}\sum_{i=1}^n 1(X_i \leq x)$.
- Note that $F_n(x)$ puts mass $\frac{1}{n}$ at each data point $X_1, \ldots, X_n$. Therefore, drawing an observation from F_n is equivalent to drawing one point at random from the original data set $\{X_i\}_{i=1}^n$ with replacement.
- The bootstrap sample will necessarily have some ties and missings.

Nonparametric Bootstrap: Step 2

- The bootstrap statistic $T_n^* = T_n(X_1^*, \ldots, X_n^*, F_n)$ is calculated for each bootstrap sample.
- When the statistic T_n is a function of F, it is typically through dependence on a parameter, θ.
- Hence, we have the bootstrap statistic $T_n^* = T_n(X_1^*, \ldots, X_n^*, \hat{\theta})$.

Nonparametric Bootstrap: Step 3

- Repeat Steps 1 and 2 B times and yield B values of T^*_{nb}:

$$\{T^*_{n1}, \ldots, T^*_{nB}\}$$

empirical distribution function

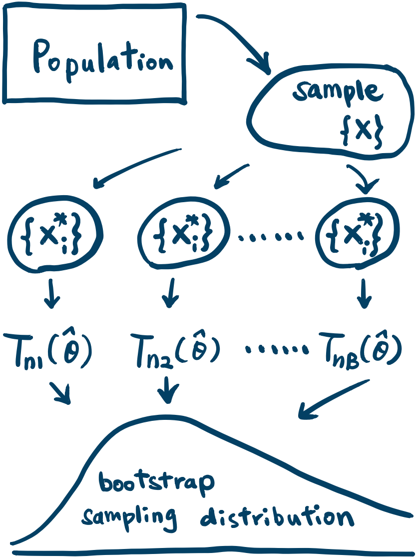
- Thus, the EDF of T^*_{nb} is

$$\hat{G}^*_n(\tau) = \frac{1}{B}\sum_{b=1}^{B} 1(T^*_{nb} \le \tau).$$

- It has been shown that as $B \to \infty$,

$$\hat{G}^*_n(\tau) \xrightarrow{p} G^*_n(\tau).$$

- It is desirable for B to be large ($B = 10000$).


Population
sample
{X}
{X*_i}
{X*_i}
{X*_i}
T_n1(θ̂)
T_n2(θ̂)
T_nB(θ̂)
bootstrap
sampling distribution

Tips for Nonparametric Bootstrap

the bootstrap sample should NOT exceed the sample size of your sample,
so nrow(x) = size is highly recommended.

- In R, a very simple command `sample(x,size,replace=True)` can be used. Moreover, many packages (such as `boot`) for bootstrap are available.
- In EViews, we can use the command `resample`.

Tips for Nonparametric Bootstrap

EViews Example

```
mode quiet
!N=10   ' the sample size
wfcreate(wf=boot_example) u 1 !N
rndseed 4321
genr X = nrnd
X.resample XB2
```

Bootstrap Confidence Interval

There are many different ways to construct bootstrap confidence interval. We are going to discuss the most popular one: Percentile Intervals.

Percentile Intervals

- Suppose that $G_n(\tau, F)$ is the distribution function of T_n.
- Let $q_n(\alpha, F)$ be its quantile function such that

$$\alpha = G_n(q_n(\alpha, F), F) = P(T_n \leq q_n(\alpha, F)).$$

- Let

$$q_n^*(\alpha) = q_n(\alpha, F_n)$$

denote the quantile function of the bootstrap distribution.

Percentile Intervals

- Given $T_n = \hat{\theta}$ be the estimate of a parameter of interest. In $100 \cdot (1-\alpha)\%$ of sample, $\hat{\theta}$ is covered by the region

$$\left[q_n\left(\frac{\alpha}{2}\right), \; q_n\left(1-\frac{\alpha}{2}\right) \right].$$

- This motivates a confidence interval for θ proposed by Efron

$$CI = \left[q_n^*\left(\frac{\alpha}{2}\right), \; q_n^*\left(1-\frac{\alpha}{2}\right) \right].$$

Percentile Intervals

- The simulation estimate of CI is

$$\widehat{CI} = \left[\hat{q}_n^* \left(\frac{\alpha}{2} \right), \;\; \hat{q}_n^* \left(1 - \frac{\alpha}{2} \right) \right],$$

  where $\hat{q}_n^*(\cdot)$ is the sample quantile of the bootstrap statistics $\{T_{n1}^*, \ldots, T_{nB}^*\}$.

- That is, we simulate $\{T_{n1}^*, \ldots, T_{nB}^*\}$, then sort them in ascending order. 由小排到大

- Finally, find the B α-th T_{nb}^* as the quantile $q_n^*(\alpha)$.

- For instance, with 1000 replications, a 95% interval is obtained by the 25th and 975th T_{nb}^*.

Bootstrap P-values (Hypothesis Testing)

- One-sided Tests
- Two-sided Tests

One-sided Tests

- Test $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ at significance level α.
- Let

$$T_n = \frac{\hat{\theta} - \theta}{S(\hat{\theta})}$$

be the test statistic.

- We first simulate the bootstrap distribution of

bootstrap 的精神是把sampling distribution 視為母體

$$T_n^* = \frac{\hat{\theta}^* - \hat{\theta}}{S(\hat{\theta}^*)},$$

where $S(\hat{\theta}^*)$ is the bootstrap standard error.

One-sided Tests

- We then find the bootstrap critical value $q_n^*(1-\alpha)$ such that

$$P(T_n^* > q_n^*(1-\alpha)) = \alpha,$$

and reject H_0 if $T_n(\theta_0) > q_n^*(1-\alpha)$.

- On the other hand, we may compute the bootstrap p-value:

$$p^* = \frac{1}{B}\sum_{b=1}^{B} 1(T_{nb}^* > T_n(\theta_0)).$$

indicator function: 數個數

Two-sided Tests

- Test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ at significance level α.
- Let
$$T_n = \frac{\hat{\theta} - \theta}{S(\hat{\theta})}$$
be the test statistic.
- Again, simulate the bootstrap distribution of
$$T_n^* = \frac{\hat{\theta}^* - \hat{\theta}}{S(\hat{\theta}^*)}.$$

Two-sided Tests

- Sort $|T^*_{nb}|$ and find $100 \cdot (1-\alpha)\%$ quantile, $q^*_n(1-\alpha)$.
- Reject H_0 if

$$|T_n(\theta_0)| > q^*_n(1-\alpha).$$

- The bootstrap p-value is:

$$p^* = \frac{1}{B}\sum_{b=1}^{B} 1\left(|T^*_{nb}| > |T_n(\theta_0)|\right).$$

Remarks on Bootstrap Tests

- Note that the bootstrap test statistic T_n^* is centered at the estimate $\hat{\theta}$, and the standard error, $S(\hat{\theta}^*)$ is calculated on the bootstrap sample.
- That is, $T_n^* = (\hat{\theta}^* - \hat{\theta})/S(\hat{\theta}^*)$ but NOT $(\hat{\theta}^* - \theta_o)/S(\hat{\theta}^*)$ or $(\hat{\theta}^* - \hat{\theta})/S(\hat{\theta})$.
- The guideline is proposed by Hall and Wilson (1991) and is often referred to as the Hall and Wilson rule.

Remarks on Bootstrap Tests

- As suggested by Hansen (2006), he states "[w]hen in doubt use $\hat{\theta}$". He also emphasizes that using θ_o rather than $\hat{\theta}$ is a "typical mistake made by practitioners".
- However, as indicated in Maddala and Kim (1998), the guideline has been violated in econometric practice (particularly in time-series econometrics) BUT with good reasons. We will talk about this later.

Bootstrap Methods for Regression Models

- Consider the following regression model:

$$y_t = \beta x_t + \varepsilon_t$$

$$\varepsilon_t \sim^{i.i.d.} (0, \sigma^2).$$

- Suppose that we are interested in testing $H_0 : \beta = \beta_0$.
- Use nonparametric bootstrap to sample (y, x) pairs from data.
- Or residual bootstrap.

Residual Bootstrap

- Step 1: Estimate the regression model and obtain estimator, $\hat{\beta}$ and $\hat{\sigma}$. Then get the residuals $\hat{\varepsilon} = \{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T\}$.
- Step 2: Get bootstrap residuals, ε^* from $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T\}$ by EITHER
    - **nonparametric method:** randomly sample from $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T\}$ with replacement, OR
    - **parametric method:** Generate bootstrap residuals ε^* form a parametric distribution, such as $\varepsilon_t^* \sim N(0, \hat{\sigma}^2)$.

Residual Bootstrap

- Step 3: The bootstrap sample of regressor, x_t^* can be generated by (1) nonparametric bootstrap, (2) parametric bootstrap or simply (3) $x_t^* = x_t$.
- Step 4: Consider two sampling schemes for the generation of the bootstrap samples

$$\mathbb{S}_1: \; y_t^* = \hat{\beta} x_t^* + \varepsilon_t^*$$
$$\mathbb{S}_2: \; y_t^* = \beta_0 x_t^* + \varepsilon_t^*$$

Residual Bootstrap

- Step 5: Consider two t-statistics

$$\mathbb{T}_1: \ T_n = \frac{\hat{\beta}^* - \hat{\beta}}{S(\hat{\theta}^*)}$$

$$\mathbb{T}_2: \ T_n = \frac{\hat{\beta}^* - \beta_0}{S(\hat{\theta}^*)}$$

- Four different combinations $[\mathbb{S}_1, \mathbb{S}_2] \times [\mathbb{T}_1, \mathbb{T}_2]$ can be applied.

Remarks

- $\mathbb{S}_1 \times \mathbb{T}_1$ is consistent with the Hall and Wilson rule.
- It is suggested (on the basis of a Monte Carlo study of an AR(1) model) that the use of $\mathbb{S}_2 \times \mathbb{T}_2$ is better than the use of $\mathbb{S}_1 \times \mathbb{T}_1$ in finite sample. (But no difference in large sample).
- Hence in most applications of Time-series econometrics, we do not follow the Hall and Wilson rule.

Bootstrapping Panel Data

- Consider the following panel regression model

$$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it}.$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

Bootstrapping Panel Data

- Let the residual be denoted by

$$
\begin{aligned}
\hat{\varepsilon}_1 &= (\hat{\varepsilon}_{11}, \hat{\varepsilon}_{21}, \ldots, \hat{\varepsilon}_{N1}) \\
\hat{\varepsilon}_2 &= (\hat{\varepsilon}_{12}, \hat{\varepsilon}_{22}, \ldots, \hat{\varepsilon}_{N2}) \\
\hat{\varepsilon}_3 &= (\hat{\varepsilon}_{13}, \hat{\varepsilon}_{23}, \ldots, \hat{\varepsilon}_{N3}) \\
&\vdots \\
\hat{\varepsilon}_T &= (\hat{\varepsilon}_{1T}, \hat{\varepsilon}_{2T}, \ldots, \hat{\varepsilon}_{NT})
\end{aligned}
$$

Bootstrapping Panel Data

- Then the first bootstrap sample of the residuals might be

$$\begin{aligned}
\hat{\varepsilon}_1^* &= (\hat{\varepsilon}_{17}, \hat{\varepsilon}_{27}, \ldots, \hat{\varepsilon}_{N7}) \\
\hat{\varepsilon}_2^* &= (\hat{\varepsilon}_{14}, \hat{\varepsilon}_{24}, \ldots, \hat{\varepsilon}_{N4}) \\
\hat{\varepsilon}_3^* &= (\hat{\varepsilon}_{19}, \hat{\varepsilon}_{29}, \ldots, \hat{\varepsilon}_{N9}) \\
&\vdots \\
\hat{\varepsilon}_T^* &= (\hat{\varepsilon}_{13}, \hat{\varepsilon}_{23}, \ldots, \hat{\varepsilon}_{N3})
\end{aligned}$$

Assume time (T) independent, but there exists dependency between individuals (N).

Limitations and Misuses

In most cases, bootstrap tests work better than the asymptotic tests. However, here are situations in which bootstrap tests perform badly.

1. Underlying residuals are serially correlated. violate i.i.d. in time dimension
2. Underlying residuals are heteroskedastic.
3. Simultaneous equation models.

However, the defects do not outweigh the merits due to the advantage of “simple-to-use”.

Limitations and Misuses

- Bootstrap is not a panacea for all statistical problem.
- Bootstrap is NOT a mthod to reduce or enlarge the data set.
- Random sampling is required (representative sample)!

random sampling 才能代表母體

Reasons to Bootstrap

- Using non-standard estimator
- Small sample
- Diagnostic check on traditional approach