The bootstrap

1. Basics on bootstrap
2. Bootstrap with asymptotic refinements
3. Bootstrap in the linear regression model
4. The number of Bootstraps
5. Final remarks

Basics on Bootstrap

- The bootstrap is a method introduced by Efron (1979) to estimate the distribution of an estimator, its variance or the distribution a test statistic by resampling the data.
- The term bootstrap is derived from the English idiom *"to pull oneself up by the bootstraps"* which means "to improve oneself by one's own efforts"
- This technique substitutes the theoretical statistical analysis by the force of computation.
- The advantages of the bootstrap are:
  - It does not require the derivation and computation of complicated formulae for, e.g., asymptotic variances.
  - In some situations, the bootstrap procedures have improved finite sample properties.
- Like conventional methods relies on asymptotic theory so only exact in infinitely large samples.
- A good reference on this topic is Horowitz (2001, Handbook of Econometrics) For a less technical approach see Cameron and Trivedi (2006, chapter 11).

- Suppose that we have a sample $(X_1, ..., X_n)$ and we would like to construct a confidence interval for a parameter θ. We estimate θ, using a mathematical formula that depends on the sample $\hat{\theta} = T(X_1, ..., X_n)$, $T(.)$ is a known function.
- **Example:** If $\mathrm{E}[X_t] = \mu$, $t = 1, ..., n$, then $\theta = \mu$ and $\hat{\theta} = \bar{X} = \sum_{t=1}^{n} X_i/n$.
- Let us assume for simplicity that $\hat{\theta}$ is root-n consistent and $\hat{\theta}$ is asymptotically normal. That is $\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \lambda^2)$
- The statistics of interest are:
  - the *estimator* $\hat{\theta}$
  - the *standard errors* $\hat{\theta}$: $s_{\hat{\theta}}$ (a consistent estimator of $\lambda/\sqrt{n}$)
  - The *t-statistic* $\hat{t} = \left(\hat{\theta} - \theta_0\right)/s_{\hat{\theta}}$ where $H_0 : \theta = \theta_0$.
  - the associated *critical value* or *p-value* for the test
  - *confidence interval*.

- The idea of bootstrap is the following:
  - Let $(X_1^*, ..., X_n^*)$ be a bootstrap sample generated from the original sample and $\hat{\theta}^* = T(X_1^*, ..., X_n^*)$. [As we will see later, bootstrap samples can be **generated in different ways**, depending on the problem in hand.]
  - Generate B different bootstrap samples and for each sample , $b = 1, ..., B$, compute $\hat{\theta}_b^* = T(X_{1,b}^*, ..., X_{n,b}^*)$.
- We can use the empirical distribution of $\hat{\theta}_b^*$ to approximate the (asymptotic) distribution of $\hat{\theta}$.
- For instance, the $100 \times (1-\alpha)$ % confidence interval for θ based on

$$\left(\hat{\theta}_{\alpha/2}^*, \hat{\theta}_{1-\alpha/2}^*\right),$$

where $\hat{\theta}_\alpha^*$ is the empirical percentile $\alpha \times 100$ of $\hat{\theta}_b^*$. This is called the *bootstrap percentile method.*

- In the *independent an identically distributed* (`i.i.d.`) *bootstrap* method proposed by Efron (1979), the bootstrap sample $(X_1^*, ..., X_n^*)$ is obtained by drawing samples with replacement from the original sample $(X_1, ..., X_n)$, where the probability of obtaining each $X_i, i = 1, ..., n$, is $1/n$. Denote $\bar{X}^*$ the average of this sample.
- Bickel and Freedman (1981) and Singh (1981) proved that if the elements of the original sample were `i.i.d.` the bootstrap distribution of $\sqrt{n}(\bar{X}^* - \bar{X})$ is close to the asymptotic distribution of $n^{1/2}(\bar{X} - \mu)$ for large n.
- This result justifies theoretically the validity of the bootstrap percentile method to construct a $100 \times (1 - \alpha)$ % confidence interval for the mean: In this case it is given

$$(\bar{X}^*_{\alpha/2}, \bar{X}^*_{1-\alpha/2})$$

where $\bar{X}^*_\alpha$ is the quantile $100 \times \alpha$ of $\bar{X}^*$.

Example: We compute a confidence interval for the mean using the bootstrap for the following sample:

Original sample					**Average**
196	-12	280	212	52	145.6

Generate $B = 100$ bootstrap samples and compute the average for each sample:

	Bootstrap Samples					**Average***
1	280	-12	196	196	52	142.4
2	-12	212	-12	-12	196	74.4
3	280	196	52	52	280	172
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	280	196	196	52	52	155.2
100	280	52	52	280	52	143.2

The 95% bootstrap confidence interval is $(48.48, 249.6)$ because the percentile 2.5 of **Average*** is 48.48 and the percentile 97.5 of **Average*** is 249.6.

- Most theoretical work on bootstrap has focussed in showing that the bootstrap distribution of $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ is close to the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta)$ for large n which justifies theoretically the validity of the *bootstrap percentile method*.
- We can use it to test hypothesis if the alternative is two sided: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ at $100 \times \alpha\%$ level. **Rejection Rule:** We reject H_0 in favor of H_1 if $\theta_0 \notin \left(\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2}\right)$.
- There is no guarantee that the bootstrap percentile method performs better than the one obtained using asymptotic theory (theoretically their performance should be similar)
- It has the advantage that it can be applied even in situations where it is difficult to obtain $s_{\hat{\theta}}$.
- There are other ways to obtain bootstrap confidence intervals from the sample $(\hat{\theta}^*_1, ..., \hat{\theta}^*_B)$.

The bootstrap estimator of θ is $\bar{\theta}^*_B = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*_b$ (usually it is close to $\hat{\theta}$). $s^2_{\hat{\theta}}$ can be approximated using

$$s^2_{boot} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^*_b - \bar{\theta}^*_B\right)^2. \tag{1}$$

Remarks:

- s^2_{boot} is very popular in empirical work because it is straightforward to compute, while in most situations it is difficult to obtain $s^2_{\hat{\theta}}$.
- However, there is not much work in the literature showing that (1) is consistent in general settings.
- Even in situations where it is consistent there is no guarantee that (1) performs better than $s^2_{\hat{\theta}}$.
- Hahn and Liao (2021) showed that $n \times s^2_{boot}$ is *never smaller* than λ^2 (as $n \to \infty$). Consequently if one uses the statistic $\left(\hat{\theta} - \theta\right)/s_{boot}$ to make inferences, it will produce large confidence intervals and tests with size smaller than α (the tests will not reject H_0 very often).

Bootstrap with asymptotic refinements

- Define α has the *nominal size* of a test.
- Consider *one sided tests*: $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$. In general, actual size of the test based on the statistic $\hat{t}$ and the critical values obtained from its asymptotic distribution is $\alpha + O(n^{-1/2})$.
- In *two sided tests*: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ the actual size of the test based on the statistic $\hat{t}$ and the critical values obtained from its asymptotic distribution is $\alpha + O(n^{-1})$.
- The term *"asymptotic refinements"* refers to the ability of the bootstrap to provide approximations to the distributions of statistics that are more accurate than the approximations of conventional asymptotic distribution theory. This implies that the actual size of the tests based on bootstrap converges faster to α than the actual size of the tests based on the conventional asymptotic distribution theory.
- For simplicity we consider here the known results for the *iid bootstrap method*.

- We need to assume that the statistic of interest is a *smooth function* of the data and that it is *asymptotically pivotal*.
- An *asymptotically pivotal* statistics means that the limit distribution does not depend on unknown parameters. Most test statistics are asymptotically pivotal, but most estimators are not.
- **Example:** Consider $\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{d} N(0,\lambda^2)$
  - Notice that under $H_0 : \theta = \theta_0$, $\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N(0,\lambda^2)$. This statistic is not asymptotically pivotal because the asymptotic distribution depends on .λ^2
  - Notice that under $H_0 : \theta = \theta_0$, $\hat{t} = \left(\hat{\theta}-\theta_0\right)/s_{\hat{\theta}} \xrightarrow{d} N(0,1)$. This statistic is asymptotically pivotal because the asymptotic distribution does not depends on .any parameters

Bootstrap with asymptotic refinements

- We bootstrap $\hat{t}$ since it is *asymptotically pivotal*.
- The idea of bootstrap is the following:
  - Generate B different bootstrap samples and for each sample , $b = 1, ..., B$, compute $\hat{t}_b = \left(\hat{\theta}_b^* - \hat{\theta}\right) / s_{\hat{\theta}_b}^*$.
- Note that we need to obtain $s_{\hat{\theta}_b}^*$ which is the standard error computed with the bootstrap sample.
- Consider $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$.
  - We define the bootstrap critical value $cv_{boot}\left(1-\alpha\right)$ as the percentile $100 \times (1-\alpha)$ of $\left(\hat{t}_1, ..., \hat{t}_B\right)$. **Rejection Rule:** We reject H_0 in favor of H_1 if $\hat{t}_{obs} > cv_{boot}\left(1-\alpha\right)$, where $\hat{t}_{obs}$ is the observed value of the statistic in the sample.
  - The actual size of this test is $\alpha + O(n^{-1})$ (which is better than $\alpha + O(n^{-1/2})$)

Bootstrap with asymptotic refinements

- Consider now the two sided tests: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$
- We can define the bootstrap critical values in two alternative ways:

1. *non-symmetric test:* Obtain the bootstrap critical values $cv_{boot}\left(\alpha/2\right)$ and $cv_{boot}\left(1-\alpha/2\right)$ as the percentiles $100 \times \alpha/2$ and $100 \times \left(1-\alpha/2\right)$ of $\left(\hat{t}_1, ..., \hat{t}_B\right)$ respectivelly: **Rejection Rule:** We reject H_0 in favor of H_1 if $\hat{t}_{obs} \notin \left(cv_{boot}\left(\alpha/2\right), cv_{boot}\left(1-\alpha/2\right)\right)$. The actual size of the test is $\alpha + O(n^{-1})$.
2. *symmetric test:* Obtain the bootstrap critical value $cv^a_{boot}\left(1-\alpha\right)$ as the percentile $100 \times \left(1-\alpha\right)$ of $\left(\left|\hat{t}_1\right|, ..., \left|\hat{t}_B\right|\right)$: **Rejection Rule:** We reject H_0 in favor of H_1 if $\left|\hat{t}_{obs}\right| > cv^a_{boot}\left(1-\alpha\right)$. The actual size of the test is $\alpha + O(n^{-3/2})$ (which is better than $\alpha + O(n^{-1})$).

Bootstrap with asymptotic refinements

- We can test hypothesis using *p-values*.
- Consider $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$.
    - The *p-values* is defined as

$$p - value = \frac{\sum_{b=1}^{B} 1(\hat{t}_b > \hat{t}_{obs})}{B}$$

- Consider now the two sided tests: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$
- The *p-values* is defined as

$$p - value = \frac{\sum_{b=1}^{B} 1(\left|\hat{t}_b\right| > \left|\hat{t}_{obs}\right|)}{B}$$

- In both situations we reject H_0 in favor for H_1 if $p - value < \alpha$.

Bootstrap with asymptotic refinements

We can also construct $100\left(1-\alpha\right)\%$ confidence intervals using the bootstrap t-statistic using one of the following formulae:

$$\left(\hat{\theta}-s_{\hat{\theta}}cv_{boot}\left(1-\alpha/2\right),\hat{\theta}-s_{\hat{\theta}}cv_{boot}\left(\alpha/2\right)\right)$$

or

$$\left(\hat{\theta}-s_{\hat{\theta}}cv_{boot}^{a}\left(1-\alpha\right),\hat{\theta}+s_{\hat{\theta}}cv_{boot}^{a}\left(1-\alpha\right)\right)$$

Remark: Other types of bootstrap t confidence intervals can also be constructed.

Bootstrap in the linear regression model

- How to generate *bootstrap samples*?
- For simplicity, consider the linear regression model $y_i = x_i'\beta + u_i$, $E(u_i|x_i) = 0$.
- Let $\hat{\beta}$ denote the OLS estimate of β and define the residuals as $\hat{u}_i = y_i - x_i'\hat{\beta}$.
- **Residual bootstrap**: denoting by u_i^* the bootstrap errors, generate the dependent variable as

$$y_i^* = x_i'\hat{\beta} + u_i^*, i = 1, ..., n$$

Different ways of generating the bootstrap errors are available:

  - **Parametric residual bootstrap**: if the distribution of the errors is known and independent of the errors, bootstrap errors can be generated from it.
  - **Semiparametric residual bootstrap**: if the errors are IID and independent of the regressors, bootstrap errors can be obtained as draws with replacement from the vector of estimation residuals $\hat{u}_i$.

Bootstrap in the linear regression model

- **Bootstrap:of the pairs** The bootstrap observations $(y_i^*, x_i^{*\prime})$ can be obtained as draws with replacement from the sample $\{(y_1, x_1'), ..., (y_n, x_n')\}$. We obtain n such observations.
- **Wild-bootstrap** (a variant of semiparametric residual bootstrap): in presence of heteroskedasticity, to preserve the link between the errors and regressors, the bootstrap error for observation i can be obtained as $u_i^* = f(\hat{u}_i)\,\varepsilon_i^b$, where
  - $f(\hat{u}_i)$ is an appropriate function of the residuals, e.g., the identity function $f(\hat{u}_i) = \hat{u}_i$;
  - ε_i^b is a random variable independent of u and x, such that $\mathrm{E}\left(\varepsilon_i^b\right) = 0$ and $Var\left(\varepsilon_i^b\right) = 1$;
  - a possibility is to use the Rademacher distribution for ε_i^b:

$$F_1 : \varepsilon_i^b = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}$$

- Moving **Blocks bootstrap**: in case the data are serially correlated, we draw (overlapping) blocks of ℓ consecutive observations from the sample $\{(y_1, x_1'), ..., (y_n, x_n')\}$ with replacement until we have a sample of size n.

Remarks:

- The bootstrap of the pairs and Wild bootstrap are valid under heteroskedasticity. Moving blocks bootstrap is valid under heteroskedasticity and serial correlation.
- There are other types of bootstrap methods in the literature.

The Number of Bootstraps

How to choose B?

There are different answers given in the literature.

- Note that for, iid bootstrap and a sample of size n there are n^n possible bootstrap samples of size n.
- For *standard error* computation Efron and Tibsharan (1993, p.52) say $B = 50$ is often enough and $B = 200$ is almost always enough.
- *Hypothesis tests* and *confidence intervals* at standard levels of statistical significance involve the tails of the distribution, so more replications are needed. See Davidson and MacKinnon (1999, 2000). For hypothesis testing at level choose B so that $(B+1)\alpha$ is an integer. e.g. at $\alpha = 0.05$ let $B = 399$ rather than 400. They recommend $B \geq 399$ if $\alpha = 0.05$ and $B = 1499$ if $\alpha = 0.01$. On the other hand, in the case of inference on means using the t-statistic and a iid sample, Babu and Singh (1983) suggest $B = n(\log n)^2$.
- Andrews and Buchinsky (2000) present an application-specific three-step numerical method to determine B for a given desired percentage deviation of the bootstrap quantity from that when $B = \infty$.

Final Remarks

- Although the bootstrap is often very accurate, it is *not always valid*. There are several examples where bootstrap leads to invalid inferences. **Examples:** distributions with heavy tails, inference about a parameter that is on the boundary of the parameter set and inference about the maximum or minimum of random variables (there are other examples).
- If the quantities involved are not asymptotically pivotal, the bootstrap procedures are no better than usual methods based on asymptotic theory but may be easier to implement.
- Do use the bootstrap to estimate the probability distribution of an asymptotically pivotal statistic or the critical value of a test based on an asymptotically pivotal statistic whenever such a statistic is available.