Bootstrap and Resampling Methods: TD 2

In this TD, we will use the bootstrap to estimate the bias and variance of an estimator/statistic.

Exercise 1

Consider the Data Generating Process (DGP) about the linear model we have seen in Session 2. The linear model was

$$Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \varepsilon \text{ with } \mathbb{E}\varepsilon X_j = 0 \text{ for } j = 1, 2$$

Let's recall that the t-statistic for the jth coefficient (with $j = 1, 2, 3$) is

$$\widehat{t}_j = \frac{\widehat{\beta} - \beta_j}{\sqrt{\widehat{\Sigma}_{jj}/n}}$$

where $\widehat{\Sigma}_{jj}$ is the jth element on the main diagonal of the matrix $\widehat{\Sigma}$ and $\widehat{\Sigma}$ the White-Huber estimator of the asymptotic covariance matrix of $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3)$. As we saw in class, a problem with the White-Huber estimator is that it does not estimate well the asymptotic variance in finite samples.

1. Can you propose a bootstrap estimator of the variance of $\widehat{\beta}$? (Hint: the variance of $\widehat{\beta}_j$ is $\mathbb{E}[\widehat{\beta}_j - \mathbb{E}\widehat{\beta}_j]^2$; starting from here, use the principle of substitution and the principle of similarity to build a bootstrap estimator of it; use the bootstrap resampling scheme we saw in class).

2. Build a function in R to compute the bootstrap estimator of the variance of $\widehat{\beta}_j$ (with $j = 1, 2, 3$) that you have constructed in Item 1.

3. In R, extract a sample from the usual DGP and use the bootstrap estimator of the variance to compute the 95 percent confidence interval for each coefficients of the linear model (so you have to construct 3 confidence intervals). (Hint: you should replace the bootstrap estimator of the variance obtained in the previous two items to $\sqrt{\widehat{\Sigma}_{jj}/n}$.)

4. In principle, we could also build confidence intervals for β_j based on the statistic $\sqrt{n}(\widehat{\beta}_j - \beta_j)$ and its bootstrap version. So, what is the advantage of bootstrapping the statistic $\widehat{t}_j = \frac{\widehat{\beta}-\beta_j}{\sqrt{\widehat{\Sigma}_{jj}/n}}$?

Exercise 2

As in TD 1, let $X \sim \mathcal{N}(0,6)$ (so that $Var(X) = 6$), and assume we observe an iid sample $\{X_i\}_{i=1}^n$ with $n = 30$. We are interested in estimating the parameter

$$\theta_0 := \exp(\mathbb{E}\{X\}) \,.$$

As we saw in TD1, a consistent estimator is $\widehat{\theta} = \exp(\overline{X})$.

1. Propose a bootstrap estimator of the bias of $\widehat{\theta}$. (Hint: Start from the definition of bias and use again the principle of similarity and the principle of substitution).

2. By using the estimator obtained in the previous item, propose a bias corrected estimator of θ_0. Call this $\widehat{\theta}_{boot}$ (bootstrap bias corrected estimator).

3. Construct a function in R that computes the bootstrap bias corrected estimator.

4. Run a Monte Carlo experiment in R to compute the bias and mean square errors of $\widehat{\theta}$ and $\widehat{\theta}_{boot}$, where $\widehat{\theta}_{boot}$ is the bootstrap bias corrected estimator.

5. Comment on the results of point 4.