OLS Bootstrapping

Pham Viet Hung

MŰEGYETEM 1782

Content

Resampling

Intro – Overview

- Extract all the info from the data (Rao, 1989)
- Earlier: a statistic is a RV having a probability distribution, that is, a sampling distribution of the statistic. Sometimes, no need to approximate the properties ("relative accuracy") of an estimator with some strong prior hypothesis. E.x.: OLS is a BLUE (Gauss-Markov Theorem).
- However, most of the time, those hypothesis do not apply. An estimation of a statistic is needed $\Longrightarrow$ Jackknife, Bootstrap

Intro – Traditional Approaches

Two kind of traditional approaches: asymptotic and exact.

Pros:

- Useful for understanding the underlying concepts and theories
- Working well when having a lot of data points

Cons:

- Assumptions of a lot of unrealistic and highly restrictive conditions. E.x.: ANOVA
- Asymptotic theories need large sample sizes.
- Difficult or sometimes not possible to derive a closed-form formula for a statistic. E.x.: Variance of the median.

Methods – Overview

Mainly Jackknife and Bootstrapping

- Widely used
- Estimating bias, a standard error of an estimate
- Constructing confidence interval(s)
- Conducting hypothesis testing
- Implicit assumption: The sample is somewhat representative of the population
- Cons: more computational power; theories are more challenging

Methods – Jackknife

A technique developed by Quenouille (1949) and popularized by Tukey (1958).

Main takeaways:

- Leave one out observation (the underlying concept similar to LOOCV in Machine Learning)
- Jackknifed distribution of a sample statistic from n leave-one-out estimators
- Much Less computationally expensive than bootstrapping, however not as powerful
- (Linear) approximation to the bootstrap, aka the more closer a statistic to linearity, the more accurate the approximation (Efron et. al, 1994)
- Many usages under the hood of some other methods/algorithms

Methods – Bootstrapping (1)

A technique developed by Bradley Efron in 1979.

- Sampling with replacement from a given sample size of n —> new samples are generated
- $P(\text{an observation in a bootstrap sample (at least once)}) = 1 - \left(1 - \frac{1}{n}\right)^n \longrightarrow 1 - e^{-1} \approx 0.632$
- Usually 10000 resampling number is (more than) enough

The result is a bootstrap sampling distribution of the statistic of interest.

Methods – Bootstrapping (2)

Advantages:

- Easy to derive statistics of our interest and confidence intervals.
- Not much assumption is needed
- Avoid the cost of repeating the experiment to obtain other groups of sample data
- Much powerful compared to jackknife

Disadvantages:

- The result strongly depends on the quality of a sample
- Used to be time-consuming and computationally expensive.

Algorithm 1: The Classic Bootstrap Algorithm For Estimating A Parameter

Input : $S = \{x_1, x_2, \ldots, x_n\}$ – a size n original samples
Output: $\hat{\theta}^* = (\hat{\theta_1^*, \ldots, \theta_n^*})$ – bootstrapped values
Init: B – num of bootstrap repetitions
$r := 1$
$L_{est} := []$ – an (empty) array of parameter estimate ;
T – A function of a statistic ;

while $r < B + 1$ **do**
    S_k := sampling a size n from S by *sampling with replacement*;
    $\hat{\theta}_r^* := T(S_k)$;
    $L_{est}[r] := \hat{\theta}_r^*$;
    r = r + 1
end

OLS

General Setup

$$Y_i = \beta_0 + \beta_1 \mathbf{X}_{i1} + \ldots + \beta_j \mathbf{X}_{ij} + \epsilon_i, \quad (Y = \mathbf{X}\beta + \epsilon) \tag{1}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots d$, that is, we have n observations and d features.

Notation:

- Y – An $n \times 1$ column-vector.
- $\mathbf{X}$ – An $n \times d$ matrix. Also called as a *Design Matrix* or *Regressor Matrix*
- $\mathbf{X}_j$ – An $n \times 1$ j^{th} regressor column-vector, where $j = 1, \ldots d$.
- $\mathbf{X}_{i.}$ – An $1 \times d$ i^{th} observation row-vector, where $i = 1, \ldots n$.
- β – A $d \times 1$ parameter/coefficient column-vector.
- ϵ – An $n \times 1$ residual column-vector.

The "classical" way – Model Assumptions (1)

There are 5 assumptions (Buteikis, 2018):

- 1.) Linear Model DGP
- 2.) Strict Exogeneity
- 3.) Condtional Homoskedasticity
- 4.) Conditionally Uncorrelated Errors
- 5.) Regressors are linearly independent
- 6.) (optional) The (conditional) residuals are normally distributed.

6.) could be useful for confidence interval estimation and hypothesis testing *when the sample size is relatively small*. The very same assumption also implies that Y is also normally distributed.

The "classical" way – Model Assumptions (2)

OLS Estimator

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X} Y \tag{2}$$

Gauss-Markov Theorem states that if the 1.) – 5.) conditions hold, the OLS estimators $\hat{\beta}$ are BLUE and consistent with the true parameter values β of the multivariate regression model.

Furthermore, if 1.) – 5.) **and** 6.) hold the conditional distribution of the OLS estimators is:

$$\hat{\beta}|\mathbf{X} \sim \mathcal{N}\left(\beta, \sigma \left(\mathbf{X}^T \mathbf{X}\right)^{-1}\right) \tag{3}$$

The "modern" way – In reality

In reality, the assumption 3.) (and 6.)) do not hold, that is, when applying OLS to real world data, the residuals are **NOT** homoskedastic. To paraphrase, they are heteroskedatic. Normality assumption on residuals is also not realistic.

There are several tests to determine the heteroskedasticity:

- Goldfeld-Quandt Test
- Breusch–Pagan Test
- White Test

The **White Test** is the mostly used test in this setting. For more info about the latter two, check out Bill Greene: Econometric Analysis (2017) textbook.

The "modern" way – Heteroskedasticity-Consistent Standard Errors (HCE)

Thanks to Halbert White (1980), new formula for adjusting the variance is available. So, we know that

$$\mathbb{V}ar\left(\epsilon|\mathbf{X}\right) = \mathbb{E}\left[\epsilon\epsilon^{T}|\mathbf{X}\right] = \Sigma = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right) = \mathbb{E}\left[\hat{\mathbf{\Sigma}}|\mathbf{X}\right], \tag{4}$$

where $\hat{\mathbf{\Sigma}} = diag\left(\hat{\epsilon}_1^2, \ldots, \hat{\epsilon}_n^2\right)$. Therefore, $\hat{\mathbf{\Sigma}}$ is a conditionally unbiased estimator for $\mathbf{\Sigma}$ Plugging into the covariance matrix of the OLS estimates, which is:

$$\mathbb{V}\left(\hat{\beta}\right) = \left(\mathbf{X}^{T}\,\mathbf{X}\right)^{-1}\,\mathbf{X}^{T}\,\mathbf{\Sigma}\,\mathbf{X}\,\left(\mathbf{X}^{T}\,\mathbf{X}\right)^{-1} \tag{5}$$

The "modern" way – Heteroskedasticity-Consistent Standard Errors (HCE)

Then, we get:

$$\mathbb{V}_{\mathsf{HCE}}\left(\hat{\beta}\right) = \left(\mathbf{X}^T\,\mathbf{X}\right)^{-1}\,\mathbf{X}^T\,\hat{\mathbf{\Sigma}}\,\mathbf{X}\,\left(\mathbf{X}^T\,\mathbf{X}\right)^{-1} \tag{6}$$

These kind of estimators are called **sandwich estimators**. Several flavours of $\hat{\mathbf{\Sigma}}$ exist.

The "modern" way – HCE versions (7 versions)

Notable ones:

$$\textbf{HC0} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \, diag\left(\hat{\epsilon}_1^2, \ldots, \hat{\epsilon}_n^2\right) \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \tag{7}$$

$$\textbf{HC3} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \, diag\left(\frac{\hat{\epsilon}_1^2}{(1-h_{11})^2}, \ldots, \frac{\hat{\epsilon}_n^2}{(1-h_{nn})^2}\right) \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1}, \tag{8}$$

where h_{ii}-s come from a weight matrix called $\mathbf{H}$ ("hat matrix") defined as:

$$\mathbf{H} = \mathbf{X} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \tag{9}$$

HC3 is called MacKinnon-White (1985) error. Numerically, HC3 is equivalent to the jackknife covariance matrix estimator.

The "modern" way – Clustered Sampling

There are several cases when the assumption of conditionally uncorrelated errors, a.k.a assumption 4.) is not justified.

Examples:

- Peer effect in a class.
- Supply chain dependence between two companies.

Solutions for these tasks are one-way and multi-way clustering in OLS, respectively.

OLS Bootstrapping

Techniques – 3 main ones

Applying bootstrap principles on OLS. These techniques are nonparametric bootstrapping.

3 ways to do that:

- Pairs Bootstrap
- Residual Bootstrap
- Wild

However, from the modelling perspective, Wild Bootstrap is the desirable one.

Techniques – Wild Bootstrap

Proposed by Liu (1988) and further developed by Mammen (1993).

The wild bootstrap DGP (Data Generating Process) can be defined as:

$$Y_i^* = \mathbf{X}_{i.}\,\hat{\beta} + \hat{\epsilon}_i\, v_i^*, \tag{10}$$

where v_i^* is a random variable usually with mean 0 and variance 1.

The typical v_i^* is drawn from a distribution called **Rademacher**:

$$v_i^* = \begin{cases} -1 & \text{w.p} \quad 1/2 \\ 1 & \text{w.p} \quad 1/2 \end{cases}$$

Techniques – Wild Bootstrap

- When conditional uncorrelatedness is assumed, for large sample size it's not worth using wild bootstrap to estimate the standard errors of the parameters. Simple OLS with HCE errors are surprisingly good estimator in this case.
- When conditional uncorrelatedness is **NOT** assumed, even for larger sample sizes it is worth using wild bootstrap. This is called **wild cluster bootstrap**.

OLS Bootstrapping – Simulation

https://github.com/pvh95/ols_bootstrap/blob/main/demonstration.ipynb

Thank you very much for your attention!

Reference 1

- Wasserman, L. (2006). *All of Nonparametric Statistics*. New York, NY: Springer.
- Buteikis, A. (2018). *Practical Econometrics and Data Science*. http://web.vu.lt/mif/a.buteikis/wp-content/uploads/PE_Book/
- Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Shao, J. & Tu, D. (1995). *The Jackknife and Bootstrap*. New York, NY, USA: Springer Verlag.
- Rao, C. R. (1989). *Statistics and Truth. Putting Chance to Work*. International Co-operative Publishing House, Burtonsville, Md.

Reference 2

- Hansen, B. (2022). *Econometrics*. Princeton University Press.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press.
- Davidson, R. & MacKinnon, J. G. (2021). *Econemtric Theory and Methods*. http://qed.econ.queensu.ca/ETM/ETM-davidson-mackinnon-2021.pdf
- Greene, W. (2017). *Econometric Analysis*. Pearson.
- White, H. (1980). *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*. Econometrica: journal of the Econometric Society, 817-838.

Reference 3

- MacKinnon, J.G. and White, H., 1985. *Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties*. Journal of econometrics, 29(3), pp.305-325.

- Liu, R.Y. (1988). *Bootstrap procedures under some non-I.I.D. models*. Annals of Statistics, 16, 1696–1708.

- Mammen, E. (1993). *Bootstrap and wild bootstrap for high dimensional linear models*. Annals of Statistics 21, 255–285.

- MacKinnon, J.G. (2013). *Thirty Years of Heteroskedasticity-Robust Inference*. https://www.econ.queensu.ca/sites/econ.queensu.ca/files/qed_wp_1268.pdf