Notes on Asymptotics of OLS Estimators

Chi-Chao Hung *

May 14, 2025

Contents

1 Asymptotics of OLS

The corner stone of the asymptotic theory of Ordinary Least Squares (OLS) estimators is the Central Limit Theorem (CLT).

Theorem 1 (Multivariate Central Limit Theorem). Let $\{Y_i\}_{i=1}^{n}$ be an Independent and Identically Distributed (IID) k-dimensional random vector with finite mean μ and variance Σ, where

$$\underset{(k\times 1)}{\mu} = \mathrm{E}\left(Y_i\right),$$
$$\underset{(k\times k)}{\Sigma} = \mathrm{Var}\left(Y_i\right) = \mathrm{E}\left(\left(Y_i-\mu\right)\left(Y_i-\mu\right)'\right).$$

Then,

$$\sqrt{n}\left(\bar{Y}_n-\mu\right) \xrightarrow{d} \mathcal{N}\left(0,\Sigma\right).$$

*All errors are mine. If you find any errors, please email me at r13323021@ntu.edu.tw.

Under some assunptions on the forth moment of the distribution of X_i abd Y_i, we can derive the **Asymptotic Normality of Least Squares Estimator**.

Theorem 2 (Asymptotic Normality of Least Squares Estimator)**.** Under assumption 7.2 in Hansen (2022), as $n \xrightarrow{p} \infty$,

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}\right),$$
$$\Omega = \mathrm{E}\left[(Xe)(Xe)'\right] = \mathrm{E}\left[XX'e^2\right].$$

The proof requires three steps. First, rewrite $\sqrt{n}(\hat{\beta}-\beta)$ as

$$\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i\right).$$

Second, applying the CLT to the second term, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i \xrightarrow{d} \mathcal{N}\left(0, \Omega\right).$$

Finally, we need to show that the first term converges in probability to $\mathbf{Q}_{XX}^{-1}$ (by WLLN and CMT) and the product of the two converges in distribution to $\mathcal{N}\left(0, \mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}\right)$ (by Slutsky Theorem).[1]

Remark. Some times people are interested in $\hat{\beta}$ alone and write it as

$$\hat{\beta} \overset{a}{\sim} \mathcal{N}\left(\beta, \frac{1}{n}\mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}\right).$$

Be careful that the $\overset{a}{\sim}$ is used becuase it is only an approximation.

Hansen (2022) use the notation $\mathbf{V}_\beta$ to denote the asymptotic variance-covariance matrix

$$\mathbf{V}_\beta = \mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}. \tag{1}$$

You can also use the notation

$$avar(\sqrt{n}(\hat{\beta}-\beta)), \tag{2}$$

[1] The formal proof has been introduced in the class and can be found in Hansen (2022) chapter 7.3.

which means the **variance of the asymptotic distribution** of $\sqrt{n}(\hat{\beta}-\beta)$. In class, professor refer to $\mathbf{V}_\beta$ as the **asymptotic variance** of $\hat{\beta}$. Despite the different names, they mean the same thing.

2 Varaince-Covariance Matrix Estimation

Since the asymptotic variance-covariance matrix $\mathbf{V}_\beta$ is not known, we need to estimate it.[2]

First define

$$\widehat{\mathbf{Q}_{XX}} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i'.$$

An natural estimator is to replace the population means with their sample counterparts.

$$\hat{\mathbf{V}}_\beta = \widehat{\mathbf{Q}_{XX}}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{e}_i \mathbf{X}_i X_i'\right)\widehat{\mathbf{Q}_{XX}}^{-1}. \tag{3}$$

where $\hat{e}_i$ is the residual from the regression of Y_i on X_i.

More often what we need is $\frac{1}{n}\mathbf{V}_\beta$, so we have

$$\begin{aligned} n^{-1}\hat{\mathbf{V}}_\beta &= n^{-1}\widehat{\mathbf{Q}_{XX}}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 X_i X_i'\right)\widehat{\mathbf{Q}_{XX}}^{-1}. && (4)\\ &= (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n} \hat{e}_i^2 X_i X_i'\right)(\mathbf{X}'\mathbf{X})^{-1}. && (5)\end{aligned}$$

You should be very famliar with this. It is the same as the **estimator of the finite sample variance-covariance matrix** of the OLS estimator, $\hat{V}_{\hat{\beta}}^{HC0}$.

HC stands for **Heteroskedasticity Consistent**. This estimator is often referred to as the **Eicker–Huber–White covariance matrix**, as it was first proposed by Eicker (1967) and Huber (1967) and later popularized by White (1980).

> **Remark.** In chapter 4, we introduce $\hat{\mathbf{V}}_{\hat{\beta}}^{HC0}$ as an estimator of
>
> $$\mathbf{V}_{\hat{\beta}} = \mathrm{Var}(\hat{\beta} \mid \mathbf{X}),$$

[2] In this note I will only focus on the case for heteroskedasticity. The case for homoskedasticity is much simpler.

the finite sample variance-covariance matrix conditional on $\mathbf{X}$. It appears again in chapter 7 but for different purpose. It is the estimator of $\mathbf{V}_\beta$ divided by n.

$$\hat{\mathbf{V}}_{\hat{\beta}}^{HC0} = \frac{1}{n}\hat{\mathbf{V}}_{\beta}^{HC0}$$

This is also the case for HC1 - HC3.

It's quite challenging to show the consistency of the estimator $\hat{\mathbf{V}}_\beta$ since WLLN does not apply (see HW8 Q7). The proof is already given in the lecure videos and I will not repeat it again.

3 Asymptotic Standard Errors

As described in Section 4.16 of Hansen (2022), the **standard error** of an estimator is a measure of the variability of the estimator. the definition of a standard error is as follows:

Definition 3.1. A standard error $s(\hat{\beta})$ for a real-valued estimator $\hat{\beta}$ is an estimator of the standard deviation of the distribution of $\hat{\beta}$.

In the following discussion of standard errors, we will focus on a particular element $\hat{\beta}_j$ in $\hat{\beta}$. Let $\mathbf{V}_{\beta_j} = \left[\mathbf{V}_\beta\right]_{jj}$, the j-th diagonal element of $\mathbf{V}_\beta$. Recall that the disribution of $\hat{\beta}_j$ is approximately

$$\hat{\beta}_j \overset{a}{\sim} \mathcal{N}\left(\beta_j, \frac{1}{n}\mathbf{V}_{\beta_j}\right). \tag{6}$$

The **asymptotic standard error** of $\hat{\beta}_j$ is defined as

$$s(\hat{\beta}_j) = \sqrt{\frac{1}{n}\hat{V}_{\beta_j}} = \sqrt{\hat{V}_{\hat{\beta}_j}}. \tag{7}$$

The second equality is the result of

$$\hat{\mathbf{V}}_{\hat{\beta}} = \frac{1}{n}\hat{\mathbf{V}}_{\beta}. \tag{8}$$

4 T-statistics

Standard error are useful for constructing t-statistics, which would be used in hypothesis testing. The t-statistics is defined as

$$T(\beta_j) = \frac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \tag{9}$$

where $s(\hat{\beta}_j)$ is the asymptotic standard error defined above. The requirement for a test statistics is that its has a known sampling distribution. In finite sample, we can only derive the distribution of $T(\beta_j)$ under the normal e_i assumption. In the asymptotic case, we can derive the distribution of $T(\beta_j)$ without assumptions on the distribution of e_i.

$$\begin{aligned} T(\beta_j) &= \frac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)} \\ &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\mathbf{V}}_{\hat{\beta}_j}}} \\ &= \frac{\sqrt{n}(\hat{\beta}_j - \beta_j)}{\sqrt{\hat{\mathbf{V}}_{\beta_j}}} \end{aligned}$$

By CLT and WLLN, the numerator converges in disribution to $\mathcal{N}\left(0, \mathbf{V}_\beta\right)$ and the denominator converges in probability to $\sqrt{\mathbf{V}_\beta}$. Slutsky's theorem implies that

$$T(\beta_j) \xrightarrow{d} \mathcal{N}(0,1).$$

5 What is reported in a regression table?

In a regression tables, we see the estimates of $\hat{\beta}$ and the standard errors of in the paratheses. For example, in professor Louh Ming-Ching's paper "Who are NTU students? (2001-2014) —the effect of Multi-Channel Admission Program", he analyzed the relation between admission channels and log of average household income. Table 18 from the paper is shown below.

Professor Louh used the statistical software Stata. Using the command `reg y x, r` in Stata, it report $\sqrt{\hat{\mathbf{V}}_{\hat{\beta}}^{HC1}}$. In R, the function `lm` report the

表 18: ln (居住地平均家庭所得) 與入學管道

	2001–2003	2004–2006	2007–2010	2011–2013
推薦甄選/學校推薦	−0.0240	−0.0114	0.0055	
	(0.025)	(0.013)	(0.009)	
申請入學/個人申請	0.0017	0.0196***	0.0428***	0.0132**
	(0.008)	(0.007)	(0.005)	(0.006)
繁星計畫/繁星推薦			−0.0862***	−0.1036***
			(0.013)	(0.009)
年份虛擬變數	√	√	√	√
常數項	6.8222***	6.8093***	6.7696***	6.8795***
	(0.005)	(0.005)	(0.004)	(0.005)
樣本數	10,368	10,347	14,518	10,454
R^2	0.015	0.008	0.025	0.016

註: 1. 對照組爲「考試分發」入學。

2. 括弧中爲 robust standard errors, $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$。

homoskedasticity version. For the heteroskedasticity-robust standard error, you can use the function `vcovHC` from the package `sandwich`. The package `fixest` also provides `vcov = "hetero"` as an option that report HC1.

References

Hansen, Bruce E., *Econometrics*, Princeton: Princeton University Press, 2022.

Acronyms

CLT Central Limit Theorem. 1, 2

IID Independent and Identically Distributed. 1

OLS Ordinary Least Squares. 1