2020 March 7

Scribed by Roh

hyunwoo@uchicago.edu

1 Review of Econometric Class

1.1 Asymptotic Normality of OLS β

Note that $Y, U \in \mathbb{R}$ and $X, \beta \in \mathbb{R}^{k+1}$. Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be an i.i.d sample from (Y, X) where $X = (1, X_1, \ldots, X_k)^T$. We write the model compactly as

$$Y = X^T \beta + U$$

where $E[XU] = 0$ and there is no perfect co-linearity in X. Suppose further that $E[Y^4] < \infty$ and $E[X_j^4] < \infty$ for all $1 \leq j \leq k$. Then, it is possible to show that

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma) \tag{1}$$

where

$$\Sigma = E[XX^T]^{-1} \operatorname{Var}[XU] E[XX^T]^{-1}$$

To establish the convergence (1) above, first recall that

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

We can then apply the multivariate CLT to see that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i U_i \xrightarrow{d} N(0, \operatorname{Var}[XU])$$

By WLLN, and CMT,

$$\left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \right)^{-1} \xrightarrow{p} E[XX^T]^{-1}$$

By Slutsky's Theorem, we see that

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} E[XX^T]^{-1} \times N\left(0, \operatorname{Var}[XU]\right)$$

In this setting, if we assume $\text{Var}[U|X] = \text{Var}[U]$ (Homoskedastic), then

$$\begin{aligned}
\text{Var}[XU] &= E\left[(XU - E[XU])(XU - E[XU])^T\right] \\
&= E\left[\left(XU - E[X \cdot \underbrace{E[U|X]}_{=0}]\right)\left(XU - E[X \cdot \underbrace{E[U|X]}_{=0}]\right)^T\right] \\
&= E\left[XU \cdot (XU)^T\right] \\
&= E\left[XUU^TX^T\right] \\
&= E\left[U^2XX^T\right] \qquad \because U \in \mathbb{R} \\
&= E\left[XX^TE\left[U^2 \mid X\right]\right] \\
&= E\left[XX^T\left(E\left[U^2 \mid X\right] - \underbrace{E\left[U|X\right]^2}_{=0}\right)\right] \\
&= E\left[XX^T\,\text{Var}[U|X]\right] \\
&= E\left[XX^T\,\text{Var}[U]\right] \\
&= E\left[XX^T\right]\text{Var}[U]
\end{aligned}$$

Hence, if U is **homoskedastic**, then

$$\begin{aligned}
\Sigma &= E[XX^T]^{-1}\,\text{Var}[XU]E[XX^T]^{-1} \\
&= E[XX^T]^{-1}\,\text{Var}[U]
\end{aligned}$$

If it is **heteroskedastic**, then

$$\begin{aligned}
\Sigma &= E[XX^T]^{-1}\,\text{Var}[XU]E[XX^T]^{-1} \\
&= E[XX^T]^{-1}E\left[XX^T\,\text{Var}[U|X]\right]E[XX^T]^{-1}
\end{aligned}$$

1.2 How to estimate this?

A natural estimator for each case is the following:

$$(Homoskedastic): \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i^T\right)^{-1} \cdot \hat{\sigma}_n^2$$

$$\hat{\sigma}_n^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(\hat{U}_i^2)$$

We divided by $(n-k-1)$ instead of n to get unbiased estimator.

$$(Heteroskedastic): \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i^T\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i^T\hat{U}_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i^T\right)^{-1}$$

This is one way of correcting heteroskedasticity. There are many other ways to do it. In R implementation, there are 5 options from HC0 to HC4. And the above formulation is equivalent to HC0.

Reference:

1. Azeem's Class Note

2 Maximum Entropy Methods

The original video is from Click For Original Video

It is also called Max Ent or Principle of Maximum Entropy.

We begin the prediction problem. Where does the max ent method excel? It works especially well in high dimensional data.

What is high dimensional Data?
A system is high dimensional if $N >> K$ where K is the amount of data and N is the number of possible configuration.

General method of enlarging the parsimonious methods. Let's think about the example of predicting the time waiting for the taxi.

How long do I have to wait?

We gather a data. Suppose that we have the following data that represents the time it took me to get a cab in minutes:

$$\{6, 3, 4, 6, 2, 3, 6, 4, 4\}$$

This is a set of observation. What should I believe about the waiting time for new york city cab?

$$P(x)?$$

Given the data we have, we are tempted to calculate the probability as follows:

$$P(6min) = \frac{3}{10}$$
$$P(3min) = \frac{2}{10}$$

and so forth. However, if we follow this naive model, the change of me waiting for the cap for other minutes that are not in the data set is all 0. This is another way to say our belief is over-fitted to the data.

2.1 Core of Max Ent Method

Continued.

Reference

1. https://www.youtube.com/watch?v=6YEn9QRy3ks&list=PLF0b3ThojznT3olRuplp5x41wUp_LZxHL