AREC422 Notes

Youpei Yan

April 23, 2019

Asymptotic

Basic asymptotic theory in this chapter: Consistency, Asymptotic Bias, Asymptotic Normality.
So far, we have visited finite data samples, this chapter, n goes to infinity.

1. Consistency

Def: W_n: an estimator of θ based on a sample with size n, if $\forall\ \epsilon > 0$, $P(|W_n - \theta| > \epsilon) \to 0$ as $n \to \infty$. We say, $plim(W_n) = \theta$, and W_n is a consistent estimator of θ. $plim$ means probability limit.

Theorem: OLS estimator $\hat{\beta}_j$ is consistent for β_j $\forall j$, i.e., $\forall \epsilon > 0$, $P(|\hat{\beta}_j - \beta_j|) > \epsilon \to 0$ as $n \to \infty$, $plim(\hat{\beta}_j) = \beta_j$.

The idea is that $\hat{\beta}_j$ is unbiased, the distribution of $\hat{\beta}_j$ has a mean value of β_j. When $n \to \infty$, $\hat{\beta}_j$'s distribution collapses into the single point β_j.

Another example of consistent estimator: sample mean.
Say, $y_1, y_2..., y_n$ are iid (independently, identically distributed) random variables with mean $/mu$, $plim(\bar{Y}_n) = \mu$. This is also known as the Law of Large Numbers (LLN).

With the single regression model: $y_i = \beta_0 + \beta_1 x_i + u_i$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i/n}{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}$$

With the LLN: $plim(\hat{\beta}_1) = \beta_1 + Cov(x,u)/Var(x) = \beta_1$, because $Cov(x,u) = 0$ (GMA4) and $Var(x) \neq 0$ (GMA3).

2. Asymptotic Bias (Inconsistency in $\hat{\beta}_j$)

If $E(u|x_1, ..., x_k) \neq 0$, OLS estimators are biased and inconsistent. Any bias persists as $n \to \infty$.

Inconsistency $= plim\hat{\beta}_1 - \beta_1 = Cov(x,u)/Var(x)$.

With the simple regression case, if the true model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v$, but we estimated as $y = \beta_0 + \beta_1 x_1 + u$, then $plim\tilde{\beta}_1 = \beta_1 + \beta_2 \delta_1$, where $\delta_1 = Cov(x_1, x_2)/Var(x_1)$.

Difference between bias and inconsistency is that the bias is from a sample, the inconsistency is caused by population variance and covariance. But it is the same in terms of determining the signs of bias/inconsistency.

3. Asymptotic Normality

We assume that OLS estimators are approximately normally distributed when $n \to \infty$.

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \overset{a}{\sim} Normal(0,1)$$

But remember that $sd(\hat{\beta}_j)$ depends on σ, and we use $\hat{\sigma}$ to estimate.

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \overset{a}{\sim} Normal(0,1)$$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \overset{a}{\sim} t_{n-k-1}$$

If we assume that $u|x_1, ...x_k \sim$ Normal distribution, then

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim Normal(0,1)$$

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

When $n \to \infty$, t-test is more reliable.

In fact, all kinds of tests (t or F) relies heavily on this “Normality Assumption”.