ECON 3740: INTRODUCTION TO ECONOMETRICS

Instructor: Chaoyi Chen
Department of Economics and Finance, University of Guelph

Lecture 12

Lecture outline

Last lecture, we studies the variance of the OLS estimators. Today, we will

- Estimate the error variance
- Efficiency of the OLS estimator - The Gauss- Markov Theorem
- Summary: MLR, estimation

- Statistical inference in regression model
- Normality assumptions
- t distribution for stabdardized estimators

MLR: Estimating the Error Variance

- The unbiased estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\widehat{\mu}_i^2 = \frac{SSR}{n-k-1},$$

where

n-k-1=n-(k+1)

=(number of observations)-(number of estimated parameters)

is called the degree of freedom (df).

- why $k+1$ degree of freedom? Cause we have $k+1$ FOCs. Given one knows $n-(k+1)$ of $\widehat{\mu}_i^2$, the rest of $k+1$ residuals can be computed based on the FOCs.
- Typically, in the SLR, $k=1$. That is why $\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{\mu}_i^2$.
- Theorem (Unbiased Estimation of $\widehat{\sigma}^2$): Under assumptions MLR.1-MLR.5,

$$E[\widehat{\sigma}^2] = \sigma^2$$

MLR: Estimation of the Sampling Variances of the OLS Estimators

- The true sampling variation of the estimated β_j is

$$sd(\widehat{\beta}_j) = \sqrt{Var(\widehat{\beta}_j)} = \sqrt{\frac{\sigma^2}{SST_j(1-R_j^2)}}$$

- The estimated sampling variation of the estimated β_j, or the standard error of β_j, is

$$se(\widehat{\beta}_j) = \sqrt{\widehat{Var}(\widehat{\beta}_j)} = \sqrt{\frac{\widehat{\sigma}^2}{SST_j(1-R_j^2)}}$$

- Note that these formulas are only valid under assumptions MLR.1-MLR.5 (in particular, there has to be homoskedasticity).

MLR: Efficiency of OLS

- Under assumptions MLR.1-MLR.5, OLS is unbiased.
- However, under these assumptions there may be many other estimators that are unbiased.
- Which one is the unbiased estimator with the smallest variance?
- In order to answer this question one usually limits oneself to linear estimators, i.e., estimators linear in the dependent variable:

$$\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i,$$

  where w_{ij} may be an arbitrary function of the sample values of all the explanatory variables.
- Typically, the OLS estimator can be shown to be of this form. For example, in the SLR,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n} w_{i1} y_i$$

  where

$$w_{i1} = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{x_i - \bar{x}}{SST_x}$$

  which is a function of $\{x_i : i = 1, ..., n\}$

MLR: The Gauss-Markov Theorem

- Theorem (The Gauss-Markov Theorem): Under assumptions MLR.1-MLR.5, the OLS estimators are the best linear unbiased estimators (BLUEs) of the regression coefficients, i.e.,

$$Var(\widehat{\beta}_j) \leq Var(\tilde{\beta}_j)$$

  for all $\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$ for which $E[\tilde{\beta}_j] = \beta_j j = 0, 1, ..., k$.
- OLS is only the best estimator if MLR.1-MLR.5 hold; if there is heteroskedasticity for example, there are better estimators (i.e., GLS, see Chapter 8).
- The key assumption for the Gauss-Markov theorem is Assumption MLR.5 (homoskedasticity).
- Due to the Gauss-Markov Theorem, assumptions MLR.1-MLR.5 are collectively known as the Gauss-Markov assumption.

MLR, estimation: summary

- In this topic, we have introduced the multiple linear regression (MLR) model. Also, we briefly introduced why we need the MLR, the motivations
- We have learned using OLS method to derive the estimators of MLR and interpreted the economic meaning with several examples.
- We learned a "Partialling Out" interpretation of multiple regression (two step estimation), and the FWL theorem
- We showed that under four (MLR.1-MLR.4) assumptions, the OLS estimators are unbiased. (The expected value of the OLS estimator equals to the true parameter)
- We dicussed two special cases, 1. *Including Irrelevant Variables in a Regression Model*. 2. *Omitted Variable Bias*. For the first case, we found the estimator is still unbiased. However, the OLS estimator of the second case will be biased.

MLR, estimation: summary continue

- By imposing one more assumption, "homoskedastic error", we showed the simpler expressions of the variance of the OLS estimators.
- We found that **1**. higher error variance increases the sampling variance. **2**. More sample variation leads to more precise estimates. **3**. Sampling variance of $\widehat{\beta}_j$ will be the higher the better explanatory variable x_j can be linearly explained by other independent variables.
- We showed an unbiased estimator of the σ^2
- Lastly, we discussed the estimator efficiency. We showed that under assumptions (MLR.1-MLR.5), the OLS estimator is the best linear unbiased estimator (The Gauss-Markov theorem).

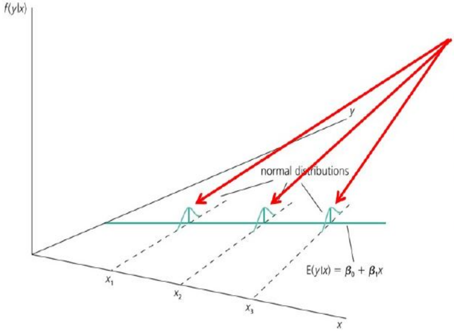
MLR: Statistical Inference in the Regression Model

- Hypothesis tests about population parameters.
- Construction of confidence intervals.
  - These two tasks are closely related.

- We need to study sampling distributions of the OLS estimators for statistical inference!
  - The OLS estimators are random variables.
  - We already know their expected values and their variances.
  - However, for hypothesis tests we need to know their distribution.
  - In order to derive their distribution we need additional assumptions.
  - Assumption about distribution of errors: normal distribution.

MLR: Statistical Inference in the Regression Model (continue)

- Assumption MLR.6 (Normality): μ_i is independent of $(x_{i1}, ..., x_{ik})$, and $\mu_i \sim N(0, \sigma^2)$.
  - It is stronger than MLR.4 (zero conditional mean) and MLR.5 (homoskedasticity).



It is assumed that the unobserved factors are normally distributed around the population regression function.

The form and the variance of the distribution does not depend on any of the explanatory variables.

It follows that:

$$y|\mathbf{x} \sim N(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2)$$

MLR: Discussion of the Normality Assumption

- The error term is the sum of "many" different unobserved factors.
- Sums of many independent and similarly distributed factors are normally distributed (central limit theorem or CLT).
- Problems:
  - How many different factors? Number large enough?
  - Possibly very heterogeneous distributions of individual factors.
  - How independent are the different factors?
  - Are they additive?
- The normality of the error term is an empirical question.
- In some cases, the error distribution should be "close" to normal, e.g., test scores.
- In many other cases, normality is questionable or impossible by definition.

MLR: Normal Sampling Distributions

- Under normality, OLS is the best (even nonlinear) unbiased estimator, i.e., it is the minimum variance unbiased estimator.
- Terminology:
  - MLR.1-MLR.5: Gauss-Markov assumptions
  - MLR.1-MLR.6: classical linear model (CLM) assumptions
- Theorem: Under assumptions MLR.1-MLR.6,

$$\widehat{\beta}_j \sim N(\beta_j,\ Var(\widehat{\beta}_j)).$$

Therefore,

$$\frac{\widehat{\beta}_j - \beta_j}{sd(\widehat{\beta}_j)} \sim N(0, 1).$$

- The estimators are normally distributed around the true parameters with the variance that was derived earlier.
  - Note that as before, we are conditioning on $\{\mathbf{x}_i, i = 1, .., n\}$.
- The standardized estimators follow a standard normal distribution.

MLR: t-Distribution for Standardized Estimators

- Theorem: Under assumptions MLR.1-MLR.6,

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1} = t_{df}.$$

- If the standardization is done using the estimated standard deviation (= standard error), the normal distribution is replaced by a t-distribution.
- The t-distribution is **close** to the standard normal distribution if $n - k - 1$ is large.
- The t-distribution is named after Gosset (1908), The probable error of a mean which Gosset published under the pseudonym Student. Consequently, this famous distribution is known as the students t rather than Gossets t! The name t was popularized by R.A. Fisher.

MLR: t-Distribution

- Recall from the lecture 2, we briefly reviewed the t distribution
- If Z is standard normal random variable and W be a chi-squared distributed random variable with k degree of freedom, then, The student t distribution with k degrees of freedom is the distribution with random variable $T = \frac{Z}{\sqrt{W/k}}$
- Now, we provide a rough idea why the standardized $\widehat{\beta}_j$ using its standard error following the t distribution.

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} = \frac{(\widehat{\beta}_j - \beta_j)/sd(\widehat{\beta}_j)}{se(\widehat{\beta}_j)/sd(\widehat{\beta}_j)}$$

Note that

$$se(\widehat{\beta}_j) = \sqrt{\frac{\widehat{\sigma}^2}{SST_j(1-R_j^2)}} = \sqrt{\frac{\sum_{i=1}^n \widehat{\mu}_i^2/(n-k-1)}{SST_j(1-R_j^2)}}$$

Therefore,

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} = \frac{(\widehat{\beta}_j - \beta_j)/sd(\widehat{\beta}_j)}{\sqrt{\frac{\sum_{i=1}^n \widehat{\mu}_i^2/(n-k-1)}{SST_j(1-R_j^2)} / Var(\widehat{\beta}_j)}} = \frac{(\widehat{\beta}_j - \beta_j)/sd(\widehat{\beta}_j)}{\sqrt{\frac{\sum_{i=1}^n \widehat{\mu}_i^2/(n-k-1)}{SST_j(1-R_j^2)} / \frac{\sigma^2}{SST_j(1-R_j^2)}}}$$

$$= \frac{(\widehat{\beta}_j - \beta_j)/sd(\widehat{\beta}_j)}{\sqrt{\sum_{i=1}^n (\frac{\widehat{\mu}_i}{\sigma})^2/(n-k-1)}} \sim \frac{N(0,1)}{\sqrt{\chi^2_{n-k-1}/(n-k-1)}} = t_{n-k-1}$$