EE425: Econometrics

Review for the Midterm Exam

Dr. Wanwiphang Manachotphong

Department of Economics, Thammasat University

26 Sep 2013

Stat Review

- Econometrics is
  - a subset of statistics (it is a regression analysis)
  - a tool which we can use to analyze the relationship between/among variables.
  - other tools can do this as well, e.g. path analysis, structural equation models, etc.
- OLS is
  - a subset of econometrics
  - we use it for linear regression. For non-linear regressions, we use other econometrics tools such as the logit model.
- Econometrics uses the "conditional distribution" concept most of the time.
  - Conditional expectation $\Rightarrow E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, E(u|x) = 0.$
  - Conditional variance $\Rightarrow Var(u|x) = \sigma^2$(homoskedasticity).

Joint vs. Conditional Distribution

Rice Export (kg.)	Country	Price	Type	Price_vietnam	Price_india
40,390	China	529	Jasmine	425	560
500,413	China	487	White	471	531
32,756	Turkey	527	Jasmine	403	532
21,427	Nigeria	580	Jasmine	417	563
803,791	Nigeria	450	White	398	552

- In this case, we can analyze
  $E(rice\ export|country,\ price,type,\ price\_vietnam,\ price\_india)$
  $= \beta_0 + \beta_1 China + \beta_2 Turkey + \beta_3 Price + \beta_4 Jasmine +$
  $\beta_5 Price_{vietnam} + \beta_6 Price_{india}$.

The Simple Regression Model : Objectives

$$y = \beta_0 + \beta_1 x_1 + u$$
$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1$$

- Understand the difference between "population regression function" and "sample regression function".
- Understand all the assumptions that make $\hat{\beta}_{OLS}$ unbiased (SLR1-4).
- Know that under some certain sets of assumptions (SLR1-5), $\hat{\beta}_{OLS}$ would be BLUE.
- Can prove that $\hat{\beta}_{OLS}$ is unbiased under the SLR1-4 assumptions.

Multiple regression Analysis: Objectives

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u.$$

- Know that we need to include more variables because otherwise SLR4 ($E(u_i|x_i) = 0$) (which leads to $Cov(x_i, u_i) = 0$) would be violated.
- Know that under MLR1-4 assumptions $\hat{\beta}_{OLS}$ would be unbiased.
  - And under MLR 1-5 assumptions, $\hat{\beta}_{OLS}$ would be the "BLUE".
- Know what factors can affect the size of $Var(\hat{\beta}_{OLS})$.
- Know how to proof and understand the "omitted variable bias".

The omitted variable bias

- If a relevant explanatory variable is omitted "and" it is correlated with an explanatory variable of interest, the $\hat{\beta}$ associated with that explanatory variable of interest would be biased.
- Suppose the true model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

- But we fail to include x_2. So, we estimate:

$$y = \beta_0 + \beta_1 x_1 + u.$$

- Then,

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{Var(x_1)}$$

We need to include relevant "x" to avoid the omitted variable bias

$$wage = \beta_0 + \beta_1 female + \beta_2 married + \beta_3 educ + \beta_4 exper + u.$$

- If "ability" and "ambition" explains "*wage*", but not included:
- MLR4 assumption will be violated $\Rightarrow Cov(married, u_i) \neq 0$ and $Cov(educ, u_i) \neq 0$
  - $\hat{\beta}_2$will account for the impact of *married* itself AND the part of "ability" and "ambition" that is correlated with *married*.
  - Likewise for $\hat{\beta}_3$ associated with *educ*.
  - $\hat{\beta}_2$and $\hat{\beta}_3$ will be biased!
- To solve the problem, we need to include "*ability*" and "*ambition*".

What if the relevant "x" is unobserved?

- Sometimes the relevant variable is "unobserved" – e.g. ability, ambition, social skills, etc.
- We can use the followings to fix the problem

1. Use proxy variables, e.g.
   1. For ability - IQ score, GPA, SAT score, etc.
   2. For social skills - whether the person used to take a lead role in school, whether the person joins any student associations, etc.
2. Employ the "instrumental variable" technique. (after midterm)
3. Employ the "Fixed effects" technique for panel data. (not covered in this class)

Control variables

- The relevant "x" that we have to include in the regression to avoid "omitted variable bias" is also called "Control Variables".

$$wage = \beta_0 + \beta_1 female + \beta_2 married + \beta_3 educ + \beta_4 exper + \beta_5 IQ + \beta_6 SAT + \beta_7 StudentAssoc + u.$$

- IQ, SAT, StudentAssoc are now included to make $\hat{\beta}_2$ and $\hat{\beta}_3$ become unbiased (or less biased).
- Control variables are not the variables of our main interest. We just include them to make $\hat{\beta}$ of other variables unbiased.

Comments on the $\widehat{Var\hat{\beta}}$

$$\widehat{Var(\hat{\beta}_j)} = \frac{\hat{\sigma}^2}{TSS_j(1-R_j^2)}$$

where

$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{(n-k-1)}$

$TSS_j = \sum_i (x_{ij} - \bar{x}_j)^2$

R_j^2 is the R^2 from $x_j = \alpha_0 + \alpha_1 x_1 + ... + \alpha_k x_k$ (not including x_j)

Comments on the $\widehat{Var\hat{\beta}}$

$$\widehat{Var(\hat{\beta}_j)} = \frac{\hat{\sigma}^2}{TSS_j(1-R_j^2)}$$

Suppose x_j= education

- The more variation in x_j(education), the lower $Var(\hat{\beta}_j)$
- The higher the linear correlation among the regressors, the LESS efficient the $\hat{\beta}_j$.
  - If other x are all about education, e.g. "years taking music lessons", "years going to tutoring schools", etc. (the correlation among all the x will be high => $\hat{\beta}_j$less efficient)
  - If other x are about different aspects of a person, e.g. "number of siblings", "times traveling outside the country", etc. (the correlation among all the x will be low => $\hat{\beta}_j$more efficient)
- If an irrelevant regressor is included, $\hat{\sigma}^2$will become unneccessarily big. (k is large)

Inference: Objectives

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u.$$

- Can test different types of hypotheses about the population parameter, i.e. about
  - an individual regression coefficient (e.g. $\beta_2 = 5$)
    - use t-test.
  - a single linear combination (e.g. $\beta_2 - \beta_3 = 0$)
    - modify the equation and use t-test.
  - multiple hypotheses ($\beta_2 = 0$ and $\beta_3 = 0$)
    - use F-test.

Further Issues: Objectives

- Learn the effects of data scaling.
- Understand how to interpret the results from different functional forms
  - Using Log (to estimate elasticity or % change or to linearize the function)
  - Using quadratics (to account for increasing or decreasing marginal effects)
- Be introduced to the use of adjusted-R^2.