2013605 Advanced Data Analysis (3-0-9)

Pre-requisite: one statistics course and some knowledge of matrix algebra

Instructor: Assoc. Prof. Pongsa Pornchaiwiseskul, Ph.D. (รศ.ดร. พงศา พรชัยวิเศษกุล)

E-Mail: Pongsa.P@chula.ac.th (general correspondence)
ppongsa@gmail.com (submit exam/project paper only)

Webpage: http://pioneer.chula.ac.th/~ppongsa/mycourses.htm
http://j.mp/pongsa

Course Objective:

The course is meant to introduce more quantitative tools that are considered appropriate for the course participants' research and dissertation in Logistics management. These tools have been extensively used to answer research questions in business and social science. They should work also in Logistics Management. It depends on the course participants' visions to use them. The tools to be covered are mainly statistical and econometric methods. Even though the course may be called "advanced", it is not so advanced in terms of Mathematics or quantitative techniques. The course will emphasize on the model formulation, estimation methods and interpretation rather than on theoretical proofs. Although it will not be emphasized on the mathematical proofs of mathematical statistics or econometric theories, the course participants are expected to have good understanding of basic calculus and matrix algebra. It is the course participants' responsibility to make sure that they have enough pre-requisite for the course. Only brief reviews of the pre-requisite statistics will be given.

Course Description:

The topics to be covered can be separated into four(4) parts. The first part will be the review of Least Square method in Regression Analysis. Assuming that the students will have good understanding of the single-equation linear model estimation and inference, the first part will cover a brief review of Uni-variate Statistics, Bi-variate Statistics and Multi-variate Statistics, the Least Square Method and its related statistical inference. The second part will aim at other Least Square estimation methods, such as, IV and GMM. The third part is to deal with the estimation of multiple-equation models. The last part is mainly Data Analysis for Categorical Dependent Variables.

Statistical Software:

There are quite a few statistical/econometric software packages available in the market (e.g., STATA, SAS, SPSS, RATS). **STATA** has been chosen for this course as it has more flexibility for the course participants to develop their own tools which are not provided as standards. Note that STATA User's Guide and Manual are parts of the reading assignment. STATA will be demonstrated in class when needed. However, it is responsibility of the course participants to learn how to use STATA as the exam questions will be based on the printouts of STATA runs. However, the course participants are free to use any software of their choice for the course project. The course participants can visit the following website for STATA instructions and manuals: http://www.ats.ucla.edu/stat/stata/

Reading Texts:

Basic Level

[1]   Gujarati, D.N., Basic Econometrics (4th edition), McGraw-Hill, 2003. The 5th edition is now available

[2]   Wooldridge, J., Introductory Econometrics: the Modern Approach(2nd edition), South-Western College, 2003 (equivalent to Gujarati). The 4th edition is now available.

[3]   Long, J.S. and J. Fresse, Regression Models for Categorical Dependent variables Using STATA (2nd edition), Stata Press, 2006

[4]   Rabe-Hesketh, S. and B. Everitt, A Handbook of Statistical Analyses using STATA (2nd edition), Chapman Hall/CRC, 2004

[5]   Myers, R.E. and R. Walpole,. Probability and Statistics for Engineers and Scientists, MacMilland, 1993

Advanced Level (optional)

[6]   Greene, W., Econometric Analysis, (6th edition), Prentice-Hall, 2008

[7]   Wooldridge, J., Econometric Analysis of Cross Section and Panel Data, MIT Press, 2001.

Grading Policy:

Mid-term Exam	30% (date to be announced)
Course project	40% (due on the final exam date)
Final exam	30% (date to be announced)

The mid-term and the final exams will be 24-hour take-home and personalized. Date and time will be announced later.

Course Project:

Each course participant must initiate a course project which involves primary or secondary data of his/her choice and uses course-related tools. Each will briefly present to the class his/her idea of course project by mid-course (the 8th week). Think about it now. There will be no presentation for the final project report. **The work reports (exams and final project) must be submitted in a single ZIP file which contains all the related workfiles and answer paper. Name the file with your Chula 10-digit ID number (e.g., 599xxxxx20<YourName><WhatYouSubmit>.zip). Submit your work report by the to-be-announced due date to ppongsa@gmail.com with the subject as "PHDLOG <WhatYouSubmit>". Please strictly follow the rule. Otherwise, your file may not reach me.**

Tentative Course Outline:

Week#	Topic#	Description	[text]chapter(s)
0	0.1	Matrix Algebra	[1]App.B, [2]App.D
1	0.2	Uni-variate Statistics	[1]App.A, [2]App.B-C
	0.3	Bi-variate Statistics	[1]chs.1-5, [2]ch.2
2	0.4	OLS Estimation of CLR Models	[1]ch.7, [2]chs.3
	1.1	Matrix Approach of OLS	[1]App.C, [2]App.E
3	1.2	Inference of OLS estimation	[1]ch.8, [2]chs.4-6
4	1.3	Dummy Independent Variables	[1]ch.9, [2]ch.7
	1.4	Two-way ANOVA techniques	[5]ch.12
5	1.4	Multicollinearity	[1]ch.10
	1.5	Heteroskedasticity	[1]ch.11, [2]ch.8
6	1.6	Model Misspecification and Selection Criteria	[1]ch.13, [2]ch.9
7	1.7	Basic Time series	[2]ch.10
	1.8	Serial Correlation	[1]ch.12,[2]ch.12
8	1.9	Panel Data Regression Models	[1]ch.16, [2]chs.13-14
9	2.1	IV and GMM Methods	[2]ch.15
10	3.1	Introduction to Simultaneous-Equation Models	[1]ch.18(717-734)
	3.2	Identification Problem	[1]ch.19(735-761)
	3.3	SEM Estimation: ILS, SURE, GMM	[1]ch.20(762-785)
11	4.1	Maximum Likelihood	[1]ch.4,7,8, [3]ch.3
12	4.2	Models for Binary Outcomes	[3]ch.4
13	4.3	Models for Ordinal Outcomes	[3]ch.5
14	4.4	Models for Nominal Outcomes	[3]ch.6
15	4.5	Models for Count Outcomes	[3]ch.7
16		************Reserved *************	