

# Introduction to Econometrics

## *Lecture 0: Introduction*

Zhaopeng Qu

Business School, Nanjing University

Sep. 4th, 2017



# Outlines

- 1 Introduction: Interesting and Important Questions
- 2 What is Econometrics?
- 3 Why and who should take the course?
- 4 The Structure of Economic Data
- 5 Logistics to the Course

# Question #1: Student's Performance and Class Size

- Economics suggests important relationships, often with policy implications, but virtually never suggests quantitative magnitudes of causal effects.
  - A Classical Issue in Economics of Education: *Student's Performance and Class Size*
  - Turn it into an empirical or policy question: What is the quantitative effect of reducing class size on student achievement?
  - Like by 5 student per class? or 10?



# Other Similar Questions

- Gender Discrimination in Labor Market
- Cigarette Taxes and Smoking
- Job Training program on earnings
- Quantitative easing(QE) on economic growth
- The effect of credit regulation on housing price
- The effect of coupon on products sales

# Quantitative Answers to Quantitative Questions

- Many decisions in economics, business and government hinge on understanding the relationship among variables in the world around us.
  - Economic theory provides clues about the direction of the answer.
  - but decisions require quantitative answers to quantitative questions.
- Therefore we have developed a framework and found a practical method that provide both a numerical answer to the question and a measure of how precise the answer is.
  - *Econometrics*

# Introduction:

- It could include
  - Testing economic theories.
  - Using historical data to give policy recommendations.
  - Estimating causal effects.
  - Using data to forecast future values of economic variables.
- Haavelmo (1944) states: “The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier.”

# Introduction:

- The term econometrics is attributed to Frisch; 1969 Nobel Prize co-winner (the first year for Economics)



# Introduction: Econometrics

- “Econometrics is by no means the same as *economic statistics*. Nor is it identical with what we call general *economic theory*, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous with the application of *mathematics to economics*. Experience has shown that each of these three view-points, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. **And it is this unification that constitutes econometrics.**”
- (Frisch, Econometrica, 1933, volume 1, pages. 1-2)

# Introduction:Econometrics

- “Ask a half dozen econometricians what econometrics is—you could get a half dozen different answers.
  - At a broad level, it is a **science and art** of using economic theory and statistical techniques to analyze economic data.”- Stock & Watson (2012)
- In general, Searching for Economic Logics from Data.
- More and more prevalence in
  - other social science such as political science, sociology, law and education studies etc
  - and business practice, such as Data Science.

# Introduction: Econometrics

- Social science(firstly start by Economics) is experiencing two methodological “revolutions” over the past few decades.
- On the one hand, there is the “credibility revolution”
  - A movement that emphasizes the goal of obtaining secure causal inferences(Angrist and Pischke, 2010)
- On the other hand, there is the “Big Data revolution”
  - A movement that emphasizes that how our increasing ability to produce, collect, store and analyze vast amounts of data is going to transform our understanding of the human affairs.
- Obviously, Econometrics are playing a key role in these two “revolution”.

# Why Econometrics is so important?

- Why should one study econometrics? How is studying econometrics helpful in understanding various economic problems? Can one excel in the field without learning econometrics?
- The answer to these questions is simple.
  - One hardly learn modern economics without the knowledge of econometrics.

# Why Econometrics is so important?

- Econometrics is one of three core courses in almost every department of economics.
- Econometrics is an important research field
  - Many Nobel Prize winners

# For Whom?

- those who want to pursue an academic career?
- Congratulations! You are in one of the most promising and internationalizing areas in China economic research.
  - Master econometrics knowledge will improve your research greatly.
  - Your research is judged on how convincing it is.
  - Econometrics helps ensure and formalize credibility.
  - Overwhelming majority of top journal articles are quantitative.

## For Whom

- those who want to enter professional market: Mastering Econometrics can get you a better job!
  - A curious phenomenon: Why did the Wall Street hire more mathematics/statistics /economics graduates than anytime in history?
  - overwhelming financial data to analyze
  - Credit cards, POS terminals and ERP systems widely used in supermarkets, banks and factories also revolutionize the business (or management) practices in marketing, accounting, management operation etc.
  - A lot of internet firms even hire economists to lead their special R&D department. Such as
    - Google
    - Microsoft
    - ebay

# For Whom

- Introduction to Econometrics could not be a boring and demanding variant of a mathematics course, but an interesting and have fun.
- You can just enjoy it by thinking in an empiricist's way in your daily life.
  - have novel ideas or new perspectives about our world.
  - Econometrics is kind of a bible or philosophy of economists.
- We will cover many very interesting topics .
  - Two Interesting Samples
    - Eg. Crime and Abortion in *Freakonomics* written by Steven Levitt.
    - Eg. What is the value to be the president's son? in *Economic Gangster* written by Raymond Fisman and Edward Miguel.

# Hard and Soft Skills

- Whatever you would like to be
  - Enjoy doing something seriously and cultivate a special quality for yourself!
- You SHOULD learn or improve several important skills during college.
- Hard Skills
  - Language
  - Computer
- Soft Skills
  - Critical Thinking
  - Presentation
  - Teamwork
- In this class, you will learn/practice all above skills.

# Conclusion

- In a word, *Introduction to Econometrics* is a *very very very* important and interesting course, you should work hard on it.

# Types of Data: Structure

- Data Structure
  - Cross-sectional data
  - Time series data
  - Pooled cross-sectional data
  - Panel data

# Cross-Sectional Data: (Main Focus)

- Units: individuals, households, firms, cities, states, countries, etc.
- Data on *multiple* agents at a *single* point in time

$$\{x_i, y_i \dots\}_{i=1}^N; N = \text{Sample Size}$$

- Usually obtained by random sampling from the underlying population. It means

$$\{x_i, y_i \perp x_j, y_j\}, i \neq j \in N$$

- Cross-sectional data are widely used in economics and other social sciences:
  - labor economics
  - public finance
  - industrial economics

# Cross-Sectional Data: (Main Focus)

• TABLE 1.1 Selected Observations on Test Scores and Other Variables for California School Districts in 1999

Observation (District) Number	District Average Test Score (fifth grade)	Student-Teacher Ratio	Expenditure per Pupil (\$)	Percentage of Students Learning English
1	690.8	17.89	\$6385	0.0%
2	661.2	21.52	5099	4.6
3	643.6	18.70	5502	30.0
4	647.7	17.36	7102	0.0
5	640.8	18.67	5236	13.9
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
418	645.0	21.89	4403	24.3
419	672.2	20.20	4776	3.0
420	655.8	19.04	5993	5.0

Note: The California test score data set is described in Appendix 4.1.

# Time Series Data:(Not Cover)

- Observations on a variable (or several variables) over time, thus data on a *single* agent at *multiple* points in time

$$\{x_t, y_t \dots\}_{t=1}^T; T = \text{Sample Size}$$

- Examples:
  - stock prices
  - money supply
  - consumer price index(CPI)
  - gross domestic product(GDP)
  - automobile sales
- Economic observations can rarely be assumed to be independent across time. So we have to account for the dependent nature of economic time series.
- Data frequency: minutes, hourly, daily, weekly, monthly, quarterly, annually

# Time Series Data:(Not Cover)

TABLE 1.2

Selected Observations on the Growth Rate of GDP and the Term Spread in the United States: Quarterly Data, 1960:Q1–2013:Q1

Observation Number	Date (year:quarter)	GDP Growth Rate (% at an annual rate)	Term Spread (% per year)
1	1960:Q1	8.8%	0.6%
2	1960:Q2	-1.5	1.3
3	1960:Q3	1.0	1.5
4	1960:Q4	-4.9	1.6
5	1961:Q1	2.7	1.4
.	.	.	.
.	.	.	.
.	.	.	.
211	2012:Q3	2.7	1.5
212	2012:Q4	0.1	1.6
213	2013:Q1	1.1	1.9

Note: The United States GDP and term spread data set is described in Appendix 14.1.

# Panel(or Longitudinal) Data

- Time series for each cross-sectional member in the data set, thus data on multiple agents at multiple points in time.
- The same cross-sectional units (individuals, firms, countries, etc.) are followed over a given time period.

$$\{x_{it}, y_{it} \dots\}_{i=1, t=1}^{NT}$$

- Advantages of panel data:
  - Controlling for (time-invariant) unobserved characteristics
  - Consideration of the effects of lag variables

# Panel(or Longitudinal) Data

**TABLE 1.3** Selected Observations on Cigarette Sales, Prices, and Taxes, by State and Year for U.S. States, 1985–1995

Observation Number	State	Year	Cigarette Sales (packs per capita)	Average Price per Pack (including taxes)	Total Taxes (cigarette excise tax + sales tax)
1	Alabama	1985	116.5	\$1.022	\$0.333
2	Arkansas	1985	128.5	1.015	0.370
3	Arizona	1985	104.5	1.086	0.362
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
47	West Virginia	1985	112.8	1.089	0.382
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
528	Wyoming	1995	112.2	1.585	0.360



# Pool(Repeat) Cross-Sectional Data(Pseudo-Panels)

- Pooled cross sections can be generated by combining two or more years cross-sectional Data.
- Cross-sectional data in each year is independent with other years.
  - It means that data does not track the respondent multiple times.
  - But the data come from a same population in different time.
- For it has both cross-sectional and time series features, so allows consideration of changes in key variables over time.
- Simple pooling may also be used when the number of observations of a single cross section is small.
- It is widely used in:
  - Before-after comparisons of a government policy
  - Cohort studies
  - Cross-sectional analyses

# Pool Cross-Sectional Data

TABLE 1.4 Pooled Cross Sections: Two Years of Housing Prices

obsno	year	hprice	proptax	sqrft	bdrms	bthrms
1	1993	85500	42	1600	3	2.0
2	1993	67300	36	1440	3	2.5
3	1993	134000	38	2000	4	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
250	1993	243600	41	2600	4	3.0
251	1995	65000	16	1250	2	1.0
252	1995	182400	20	2200	4	2.0
253	1995	97500	15	1540	3	2.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
520	1995	57200	16	1100	2	1.5

# Types of Data

- Traditional Collecting method
  - Census
  - Survey
  - Administrative data
- Digital times
  - Mass or Big data:
    - Online document
    - Social Media
    - Geolocations or Geographic data
  - High Frequency Data
    - Stock,futures or other financial transactional data

# Data in China

- Survey Data
  - China Household Income Project(CHIP)-Pool Cross Sectional
  - China Family Panel Survey(CFPS)-Panel
  - China Health and Retirement Longitudinal Study(CHARLS)-Panel

# Econometrics: sub-fields or sub-course

- Micro-Econometrics
  - Cross-Sectional
  - Pool Cross Sectional
  - Panel
- Macro-Econometrics
  - Times series

# Econometrics: sub-fields or sub

- Theoretical Econometrics

- It is concerned with methods, both their properties and developing new ones.
- It is closely related to mathematical statistics, and it states assumptions of a particular method, its properties etc.
- We could call theoretical econometricians as the *producer* of econometrics.

- Applied Econometrics

- More orientated to applied work, such as choice of technique and interpretation of research finding.
- **But it should be also based on a solid conceptual foundation and some practical experiences plus a little bit skills of computer.**
- We are the *consumers* of econometrics.

# About Me and our TA

- My name is *QU, Zhaopeng* (曲兆鹏)
  - Position and Affiliation: *Associate Professor, Institute of Population Studies, Business School.*
  - Research Fields: *Labor Economics and Applied Econometrics*
  - Office: room 2021, Anzhong Building
  - Office Hour: Make an appointment in advance
  - Tel: 83621232
  - Website: [frankqu.weebly.com](http://frankqu.weebly.com)
  - Email: qu@nju.edu.cn
- TA: *Xiaoguang Ling*(凌晓光 )
  - 3rd year graduate student
  - good at Stata and R
  - WeChat ID: hunadong
  - Email: hunadong@163.com

# Prerequisite

- I assume that You should be comfortable some basic concepts of probability theory and statistics, such as
  - expectation
  - variance
  - probability density function, p.d.f.
  - cumulative distribution function, c.d.f
  - covariance
- There'll be a brief review of basic statistics after we fix the big picture.

# Reference Textbooks

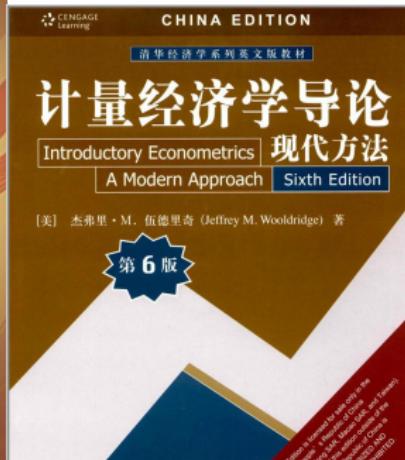
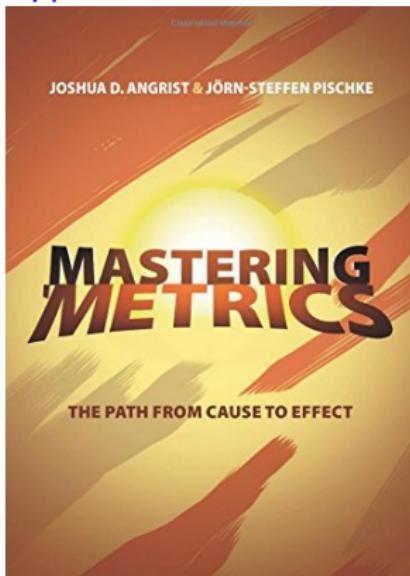
- Required Textbook: James H. Stock & Mark W. Watson, (2012). *Introduction to Econometrics*, 3rd Edition, Pearson Education.
  - 影印版：格致出版社 / 上海人民出版社。
  - 中文版：世纪出版集团 / 上海人民出版社。



# Reference Textbooks

- Supplementary textbook

- Joshua D. Angrist & Jörn-Steffen Pischke, (2014). *Mastering 'metrics: The Path from Cause to Effect*. Princeton University Press.
- Jeffrey M. Wooldridge,(2012). *Introductory Econometrics: A Modern Approach*, 5th or 6th Edition, South-Western College.



# Computing Tools

- The main computing tools used in the course are ***Stata*** and ***R***, optionally.
  - ***Stata*** is a complete, integrated statistical package that provides everything you need for data analysis, data management, and graphics.
  - ***R*** is a free software and useful for statistical study and econometrics.

# Reference Software Books

- Stata Instruction
  - 任何一本中文书或英文书。
- R Instruction
  - 任何一本中文或英文书。
  - too many resources online such as *Datacamp*

# Interesting Books

- Steven D. Levitt and Stephen J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, 2005. (中译本《魔鬼经济学》斯蒂夫·列维特、斯蒂芬·都伯纳著, 广东经济出版社, 2006年1月。)
- Steven D. Levitt and Stephen J. Dubner, *SuperFreakonomics: Global Cooling, Patriotic Prostitutes, and Why Suicide Bombers Should Buy Life Insurance*, 2009. (中译本, 《超爆魔鬼经济学》, 斯蒂夫·列维特、斯蒂芬·都伯纳著, 中信出版社, 2010年1月。)
- Ian Ayres, *Super Crunchers: Why Thinking-By-Numbers is the New Way To Be Smart*, 2007. (中译本《超级数字天才》, 伊恩·艾瑞斯著, 中国青年出版社, 2008年1月。)
- Raymond Fisman & Edward Miguel, *Economic Gangsters: Corruption, Violence, and the Poverty of Nations*, 2010. (中译本: 《经济黑帮: 腐败、暴力的黑帮经济学》, 中信出版社。)
- Abhijit V. Banerjee & Esther Duflo, *Poor Economics A Radical Rethinking of the Way to Fight Global Poverty*, 2011. (中译本: 《贫穷的本质: 我们为什么摆脱不了贫穷》, 中信出版社。)
- Angus Deaton, *The Great Escape:Health, Wealth, and the Origins of Inequality*, 2015. (中译本: 《逃离不平等: 健康、财富及不平等的起源》, 中

# What I promise to offer you

- Prepare lectures as well as possible.
- One to one interaction on topics covered in the course, especially for your own topics.
- Help you start to using Stata or R to analyze some popular data sets in China.
- A good score?
  - It depends on you.

# What I expect to you

- Class participation with a little bit aggressive attitude.
  - More questions, more scores!
  - Interrupt me as often as necessary!
  - Got a dumb question? Assume that you are the smartest person in class and you eventually will be!
- Read required materials and finish homework.

# Evaluation

- Class participation (10%)
- Homework(10%)
- Midterm Exam: (20%)
- Final Exam: (30%)
- Independent Project Report(30%)

# Welcome contact me and TA

- Any Questions?