Nature of Econometrics

Overview and Economic Data

Read Wooldrige (2013), Chapter 1

Outline

- I. What is Econometrics?
- II. Steps in Empirical Economic Analysis
- III. The Structure of Economic Data
- IV. Causality

I. What is Econometrics?

1) 2000 baht cheque handout
– a stimulus package to jump start the economy from the economic crisis of 2008

- We want to determine the effect of government expenditures on GDP or economic growth.

A Stock Analyst

- Effect of past stock returns on the stock price

- Effect of profit margins on stock values.
  – Economic fundamentals are the real “characteristics of the firm”.

Econometrics Defined

- Econometrics means "economic measurement"
- **Jeffery Wooldridge (2009)**
  Econometrics is based upon the development of statistical methods for
  - o estimating economic relationships
  - o testing economic theories, and
  - o evaluating and implementing government and business policy"

II. Steps in Empirical Economic Analysis

Empirical economic analysis – uses data to test a theory or to estimate a relationship.

- **1) Formulating a question of interest:**
- We should have
  - – a statement of theory or hypothesis
  - – specifications of the economic model and
    ***wine* = f(price, income)**
  - – econometric models.

Economic Model

- Example: Demand for imported wine (y) over 1994-1998.

Economic (Mathematical) Model

$wine = \beta_0 + \beta_1 price + \beta_2 income$

Econometric Model

$wine_t = \beta_0 + \beta_1 price_t + \beta_2 income_t + u_t$

u_t: unobserved Factors

- $wine_t = \beta_0 + \beta_1 price_t + \beta_2 income_t + u_t$
  u_t: unobserved factors. They are factors that affect wine other than price and income.
- "u_t" is called the **error** term or **disturbance** term.
- What are some "u_t" in the wine example?

2) Gathering Data

- Empirical economic analysis
- Wine Consumption between 1994:10 to 1998:09 (research done by Panipa Changpetch)
  - *wine* : quantity of imported *wine* demand.
    Excise Department (**thousands of liters**))
  - *price* : *price* of imported wine
    Ministry of Commerce (baht)
  - *income* : personal *income*
    National income data from NESDB (**millions of baht**)

Experimental vs. Nonexperimental

- Nonexperimental data
  - Examples:
    - wine prices (Commerce Ministry)
    - wine imports (Custom Department)
    - disposable income (NEDSB)
- Experimental data – from laboratory
  - values obtained from controlled experiments

3) Estimating the model

- Want to find the values of these parameters, β_0, β_1 and β_2 from the model.
  $wine = \beta_0 + \beta_1 price + \beta_2 income + u$
- From Estimation: **Regression Analysis**
- Estimates: $\hat{\beta}_0 = 1738$; $\hat{\beta}_1 = -9.1$; $\hat{\beta}_2 = .0027$

  $\widehat{wine} = 1738 - 9.1 price + 0.0027 income$

  - If price increases by one baht, the predicted value of the quantity demanded decreases by 9,100 liters, *ceteris paribus.*
  - If income increases by one million baht, the predicted value of wine consumption increases by .0027 thousand liters or 2.7 liters, *ceteris paribus.*

Eviews Results

Dependent Variable: WINE
Method: Least Squares
Sample: 1994:10 1998:09
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1738.059	584.4014	2.974085	0.0047
PRICE	-9.14414	0.743276	-12.30249	0
INCOME	0.00276	0.00181	1.524716	0.1343

R-squared	0.937366	Mean dependent var	530.3638
Adjusted R-squared	0.934582	S.D. dependent var	299.6162
S.E. of regression	76.63244	Akaike info criterion	11.57638
Sum squared resid	264263.9	Schwarz criterion	11.69333
Log likelihood	-274.833	F-statistic	336.7305
Durbin-Watson stat	1.945696	Prob(F-statistic)	0

4) Testing Hypothesis

- We want to test various hypotheses, which can state in terms of the unknown parameters.
  - Test whether price has any effect on wine demand

    $wine = \beta_0 + \beta_1 price + \beta_2 income + u$

    $H_0: \beta_1 = 0$

- Such confirmation and refutation of sample evidences is based on a branch of statistical theory **– statistical inference –** learning something about the population from a sample.
  - hypothesis testing

5) Interpreting the result

- Prediction and forecasting
- What is the predicted value of the quantity demanded of wine when *price=200 baht* and *income=250,000 million baht*?
- $\widehat{wine} = 1738 - 9.1 price + 0.0027\, income$

  $\widehat{wine} = 599.1760.$

III. The Structure of Economic Data

1. Cross section
2. Time series
3. Pooled cross section
4. Panel or longitudinal data

1) Cross-sectional Data

- **Cross-sectional data** refer to data collected from a population at a given point in time.
  - Example: a sample of individuals, households, firms, provinces, countries, or a variety of other units.
  - A Cross Section Data Set on Wages and Other Individual Characteristics [see Table]

A Cross Section Data Set on Wages and Other Individual Characteristics

obsno	wage	educ	exper	female	married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
..	..	..	..	..	..
..	..	..	..	..	..
..	..	..	..	..	..
525	11.56	16	5	0	1
526	3.50	14	5	1	0

Cross-sectional Data

Random sampling – assumption

- Example: Draw a random sample of 526 people from the working population.
- This requires that observations are **independently** drawn.

Cross-sectional Data

- Random sampling – definition

  If $Y_1, Y_2, \ldots, Y_n$ are independent random variables with a common pdf $f(y;\mu,\sigma^2)$, then $\{Y_1, Y_2, \ldots, Y_n\}$ is a random sample from the population represented by $f(y;\mu,\sigma^2)$

- *We also say that*
  - Y_i are ***i.i.d.*** (independent, identically distributed) random variables from a distribution.

2) Time series data

- **A time series** is data of a variable or several variables collected at different times.
  - Unlike cross section data, order is important.
  - Note that time series economic data can rarely be assumed to be drawn from a random sample.
  - Trends and seasonality will be important

A time series data set of foreign wine consumption in Thailand

obs	*year*	*Wine*	*Price*	*Income*
1	1994:10	131	241.41	228133.8
2	1994:11	408	213.82	230330.7
3	1994:12	404	218.18	235050.6
4	1995:01	200	237.35	235835.9
5	1995:02	264	238.24	236490.2
	..	..	..	..
46	1998:07	265.59	240.8445	229220.3
47	1998:08	151.41	234.0185	228601
48	1998:09	221.54	230.874	224751.3

3) Pooled Cross Sections

- The data from pooled cross sections have both cross-sectional and time series features.
- Example: The cross-sectional household survey (Socioeconomic Survey) in Thailand: one in 2006 and another in 2008
  - In each year, a random sample of households is surveyed for variables such as income, saving, family size, and so on.
  - Two years of housing prices (1993, 1995).

A Data Set of Pooled Cross Sections

obsno	*year*	*hprice*	*proptax*	*sprft*	*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

Pooled Cross Sections - continue

- Advantages:

1. An increase of sample size.
2. A study of a key relationship changing over time: a change in new policy.

4) Panel Data (Longitudinal data)

- A panel data set consists of *a time series for each cross sectional member* in the data set.
- We follow the same random individual observations over time – known as panel data or longitudinal data

- Example:
  - Investment and financial data of the same set of firms over a five-year time period.
  - A two-year panel data set on city crime statistics

A Panal Data Set on City Crime Statistics

obsno	city	year	murders	population	unem	police
1	1	1986	5	350000	8.7	440
2	1	1990	8	359200	7.2	471
3	2	1986	2	64300	5.4	75
4	2	1990	1	65100	5.5	75
..	..	..	..	..	..	..
..	..	..	..	..	..	..
..	..	..	..	..	..	..
297	149	1986	10	260700	9.6	286
298	149	1990	6	245000	9.8	334
299	150	1986	25	543000	4.3	520
300	150	1990	32	546200	5.2	493

Panel Data - Continue

- Advantages
  - 1) same unit allows us to control certain unobserved characteristics of individuals and other units.
  - 2) it allows us to study the importance of lags in behavior or the result of decision making

IV. Causality

- In most tests of economic theory and certainly for evaluating public policy, the goal is to infer that one or more variables have a causal effect on another variable.
  - Infer – to form an opinion that something is probably true because of information that you have.

- Simply establishing a relationship between variables is rarely sufficient

$$wine = \beta_0 + \beta_1 price + u$$

The Question of Causality

- If we've truly controlled for enough other variables, then the estimated ceteris paribus effect can often be considered to be causal.
- The notion of "ceteris-paribus" plays an important rule in economics.
  - "Ceteris paribus" refers to other things being equal.

Example of Causality

- Example : find the effect of price on wine consumption.

  $wine = \beta_0 + \beta_1 price + \beta_2 income + u$

  - Relationship between ***two*** variables.
  - Question: given an increase in price, by how much will the wine consumption increase?
- While the error term, *u*, includes other factors affecting wine consumption. We want to control for as much as possible
- Some things are still unobserved, which can be problematic.

Recap of Overview

- What is Econometrics?
- Steps in Empirical Economic Analysis
- The Structure of Economic Data
- Causality

GRADING SYSTEM

- 15 percent on homework assignments,
- 40 percent on the midterm examination and
- 45 percent on the final.
- Note that Students are strongly advised to attend lectures, for attendance will be checked randomly and be used partially to adjust grade distribution.

Grade distribution is as follows:

90-100	A;
80-90	B+;
65-80	B;
55-65	C+;
45-55	C;
35-45	D+;
25-35	D;
below 25	F.