ECO 344 - Applied Econometrics I
Chapter 1

Muhammad Salman Khalid

School of Economics & Social Sciences

August 27, 2026



Outline

1. What is Econometrics?
2. Data & its Types
3. Empirical Analysis
4. Causal Inference
5. Summary & Discussion

What is Econometrics?

Econometrics

"Econometrics is based upon the development of Statistical Methods for estimating economic relationships, testing economic theories, and evaluating different government, business and economic policies"

- Effect of Training Program on Worker's Hourly Wage?

Econometrics

"Econometrics is based upon the development of Statistical Methods for estimating economic relationships, testing economic theories, and evaluating different government, business and economic policies"

- Effect of Training Program on Worker's Hourly Wage?
- Do returns on different investments comply with the economic theories?

Econometrics

"Econometrics is based upon the development of Statistical Methods for estimating economic relationships, testing economic theories, and evaluating different government, business and economic policies"

- Effect of Training Program on Worker's Hourly Wage?
- Do returns on different investments comply with the economic theories?
- Another common application is the forecasting of important macroeconomic indicators such as interest rate, inflation rate, gross domestic product etc.

Econometrics

"Econometrics is based upon the development of Statistical Methods for estimating economic relationships, testing economic theories, and evaluating different government, business and economic policies"

- Effect of Training Program on Worker's Hourly Wage?
- Do returns on different investments comply with the economic theories?
- Another common application is the forecasting of important macroeconomic indicators such as interest rate, inflation rate, gross domestic product etc.
- Effect of political campaign expenditure on voting outcomes.

Econometrics

"Econometrics is based upon the development of Statistical Methods for estimating economic relationships, testing economic theories, and evaluating different government, business and economic policies"

- Effect of Training Program on Worker's Hourly Wage?
- Do returns on different investments comply with the economic theories?
- Another common application is the forecasting of important macroeconomic indicators such as interest rate, inflation rate, gross domestic product etc.
- Effect of political campaign expenditure on voting outcomes.
- Effect of school spending on student performance in the classroom.

Data & its Types

Data

Data is the set of information about particular subjects which is used to analyze something and make decisions.

- There are two types of classification of data.

Data

Data is the set of information about particular subjects which is used to analyze something and make decisions.

- There are two types of classification of data.
  - Classification based on data generating process.

Data

Data is the set of information about particular subjects which is used to analyze something and make decisions.

- There are two types of classification of data.
  - Classification based on data generating process.
  - Classification based on the structure of data.

Data Generating Process

Data generating process, as the name suggests, is how the data is generated.

- There are two types of data based on its generating process.

Data Generating Process

Data generating process, as the name suggests, is how the data is generated.

- There are two types of data based on its generating process.
  - Experimental Data - Collected mostly in lab environment - However, there are ways to collect experimental data in Social Sciences like Randomized Control Trial (RCT).

Data Generating Process

Data generating process, as the name suggests, is how the data is generated.

- There are two types of data based on its generating process.
  - Experimental Data - Collected mostly in lab environment - However, there are ways to collect experimental data in Social Sciences like Randomized Control Trial (RCT).
  - Non-Experimental Data - Observational Data like Census Data etc.

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.
3. Pooled Cross Section Data

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.
3. Pooled Cross Section Data
   - It is the combination of cross-sectional data and time series data.

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.
3. Pooled Cross Section Data
   - It is the combination of cross-sectional data and time series data.
   - Time series of cross-sectional data.

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.
3. Pooled Cross Section Data
   - It is the combination of cross-sectional data and time series data.
   - Time series of cross-sectional data.
4. Panel Data

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.
3. Pooled Cross Section Data
   - It is the combination of cross-sectional data and time series data.
   - Time series of cross-sectional data.
4. Panel Data
   - Panel data consists of a time series for each cross-sectional member in the data set.

Data Structure

There are four types of data based on its structure.

1. Cross-Sectional Data
   - Consists of a sample of individuals, households, firms, cities, states, countries or a variety of other units in a single time period.
2. Time Series Data
   - Time series data consists of observations on a variable or several variables over time like stock prices, GDP etc.
3. Pooled Cross Section Data
   - It is the combination of cross-sectional data and time series data.
   - Time series of cross-sectional data.
4. Panel Data
   - Panel data consists of a time series for each cross-sectional member in the data set.
   - Difference between Pooled Cross Sectional & Panel Data?

Empirical Analysis

Empirical Economic Analysis

An empirical economic analysis uses data to test a theory or estimate a relationship.

- The first step in any empirical analysis is to carefully formulate the question of interest.

In your final paper, you will follow all these five steps and hopefully have something worthy to pursue in the future.

Empirical Economic Analysis

An empirical economic analysis uses data to test a theory or estimate a relationship.

- The first step in any empirical analysis is to carefully formulate the question of interest.
- Second step is to construct a formal economic model.

In your final paper, you will follow all these five steps and hopefully have something worthy to pursue in the future.

Empirical Economic Analysis

An empirical economic analysis uses data to test a theory or estimate a relationship.

- The first step in any empirical analysis is to carefully formulate the question of interest.
- Second step is to construct a formal economic model.
- Third step is to find the appropriate data.

In your final paper, you will follow all these five steps and hopefully have something worthy to pursue in the future.

Empirical Economic Analysis

An empirical economic analysis uses data to test a theory or estimate a relationship.

- The first step in any empirical analysis is to carefully formulate the question of interest.
- Second step is to construct a formal economic model.
- Third step is to find the appropriate data.
- Fourth step is to estimate the economic model using appropriate data.

In your final paper, you will follow all these five steps and hopefully have something worthy to pursue in the future.

Empirical Economic Analysis

An empirical economic analysis uses data to test a theory or estimate a relationship.

- The first step in any empirical analysis is to carefully formulate the question of interest.
- Second step is to construct a formal economic model.
- Third step is to find the appropriate data.
- Fourth step is to estimate the economic model using appropriate data.
- Finally the last step is to report your findings in an organized format.

In your final paper, you will follow all these five steps and hopefully have something worthy to pursue in the future.

Randomness in Data

- Non-Random Data - Convenience Sampling

Randomness in Data

- Non-Random Data - Convenience Sampling
  - Selection Bias

Randomness in Data

- Non-Random Data - Convenience Sampling
  - Selection Bias
- Random Data - The data that is collected from a sample of random subjects.

Causal Inference

Causal Inference

- Economists are concerned with the causal relationships.

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.
  - The strong correlation between chocolate consumption and the number of Nobel Prizes - Causal relationship??

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.
  - The strong correlation between chocolate consumption and the number of Nobel Prizes - Causal relationship??
  - New agricultural policy affecting the crop yield this year? - Causal relationship?

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.
  - The strong correlation between chocolate consumption and the number of Nobel Prizes - Causal relationship??
  - New agricultural policy affecting the crop yield this year? - Causal relationship?
- Ceteris Paribus - Keeping all other factors constant.

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.
  - The strong correlation between chocolate consumption and the number of Nobel Prizes - Causal relationship??
  - New agricultural policy affecting the crop yield this year? - Causal relationship?
- Ceteris Paribus - Keeping all other factors constant.
- Potential Outcome/Counterfactual

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.
  - The strong correlation between chocolate consumption and the number of Nobel Prizes - Causal relationship??
  - New agricultural policy affecting the crop yield this year? - Causal relationship?
- Ceteris Paribus - Keeping all other factors constant.
- Potential Outcome/Counterfactual
  - The effect of Panadol on your body temperature?

Causal Inference

- Economists are concerned with the causal relationships.
- Finding an association between two variables might be suggestive & not causal.
  - The strong correlation between chocolate consumption and the number of Nobel Prizes - Causal relationship??
  - New agricultural policy affecting the crop yield this year? - Causal relationship?
- Ceteris Paribus - Keeping all other factors constant.
- Potential Outcome/Counterfactual
  - The effect of Panadol on your body temperature?
  - The effect of COVID on your school learning?

Summary & Discussion

Summary

- What is Econometrics?

Summary

- What is Econometrics?
- Types of data based on data generating process (Experimental vs Observational)

Summary

- What is Econometrics?
- Types of data based on data generating process (Experimental vs Observational)
- Types of data based on structure (Cross-Sectional vs Time Series)

Summary

- What is Econometrics?
- Types of data based on data generating process (Experimental vs Observational)
- Types of data based on structure (Cross-Sectional vs Time Series)
- Pooled Cross Section Data vs Panel Data

Summary

- What is Econometrics?
- Types of data based on data generating process (Experimental vs Observational)
- Types of data based on structure (Cross-Sectional vs Time Series)
- Pooled Cross Section Data vs Panel Data
- Five steps of Empirical Economic Analysis

Summary

- What is Econometrics?
- Types of data based on data generating process (Experimental vs Observational)
- Types of data based on structure (Cross-Sectional vs Time Series)
- Pooled Cross Section Data vs Panel Data
- Five steps of Empirical Economic Analysis
- Random Vs Non-Random Vs Quasi-Random Data

Summary

- What is Econometrics?
- Types of data based on data generating process (Experimental vs Observational)
- Types of data based on structure (Cross-Sectional vs Time Series)
- Pooled Cross Section Data vs Panel Data
- Five steps of Empirical Economic Analysis
- Random Vs Non-Random Vs Quasi-Random Data
- Causal Inference (Ceteris Paribus, Potential Outcomes, Counterfactual)

Discussion

Economic Model of Crime

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

where;

$$y = HoursSpentinCriminalActivity$$

$$x_1 = Wageforanhourspentincriminalactivity$$

$$x_2 = Hourlywageinlegalemployment$$

$$x_3 = Incomeotherthanfromcrime/employment$$

$$x_4 = Probabilityofgettingcaught$$

$$x_5 = Probabilityofbeingconvictedifcaught$$

$$x_6 = ExpectedSentenceifconvicted$$

$$x_7 = Age$$

Discussion

Job Training & Worker Productivity

$$wage = f(education, experience, training)$$

$$wage = \beta_0 + \beta_1 education + \beta_2 experience + \beta_3 training + \mu$$

Discussion

Data Types

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

Discussion

Data Types

obsno	country	gpcrgdp	govcons60	second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Belgium	2.56	13	69
4	Bolivia	1.24	18	12
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
61	Zimbabwe	2.30	17	6

Discussion

Data Types

obsno	year	avgmin	avgcov	prunemp	prgnp
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

Discussion

Data Types

obsno	year	hprice	proptax	sqrft	bdrms	bthrms
1	1993	85,500	42	1600	3	2.0
2	1993	67,300	36	1440	3	2.5
3	1993	134,000	38	2000	4	2.5
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
250	1993	243,600	41	2600	4	3.0
251	1995	65,000	16	1250	2	1.0
252	1995	182,400	20	2200	4	2.0
253	1995	97,500	15	1540	3	2.0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
520	1995	57,200	16	1100	2	1.5

Discussion

Data Types

obsno	city	year	murders	population	unem	police
1	1	1986	5	350,000	8.7	440
2	1	1990	8	359,200	7.2	471
3	2	1986	2	64,300	5.4	75
4	2	1990	1	65,100	5.5	75
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
297	149	1986	10	260,700	9.6	286
298	149	1990	6	245,000	9.8	334
299	150	1986	25	543,000	4.3	520
300	150	1990	32	546,200	5.2	493

Discussion

Question

1 Suppose that you are asked to conduct a study to determine whether smaller class sizes lead to improved student performance of fourth graders.

(i) If you could conduct any experiment you want, what would you do? Be specific.

(ii) More realistically, suppose you can collect observational data on several thousand fourth graders in a given state. You can obtain the size of their fourth-grade class and a standardized test score taken at the end of fourth grade. Why might you expect a negative correlation between class size and test score?

(iii) Would a negative correlation necessarily show that smaller class sizes cause better performance? Explain.