Lecture 5: Types of data; sources and quality of data

Econometrics ▪ Unit I: Introduction and generalized least squares

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

B.Sc. (Hons) Statistics, Semester V
Academic Session 2026–2027

This lecture

Previous lecture

Lecture 4: Meaning and nature of econometrics; exact and econometric relationships.

Today

- Types of data

1. Cross-section data

Econometric models are estimated from data, and the type of data decides which problems we must watch for. The main types are as follows.

1. Cross-section data. Observations on many units (households, firms, districts, states) at *one point of time*.

- For example, monthly consumption expenditure of 5,000 households collected in one survey round, or the output and employment of 200 factories in 2025.
- The order in which the units are listed has no meaning.
- A common problem with cross-section data is *heteroskedasticity* (Chapter 4), because large and small units usually show very different variability.

2. Time-series data. Observations on *one unit over successive periods of time*.

- For example, India's annual GDP from 2000 to 2025, monthly consumer price index, or daily closing prices of a share.
- The *frequency* of the data may be annual, quarterly, monthly, weekly or daily, and the order of observations is essential.
- Time series often show trends and *autocorrelation* (Chapter 5), because the value in one period is usually related to the values in nearby periods.

3. Pooled cross-section data. Two or more cross-sections taken at different times, with *different* units in each.

- For example, two household surveys conducted five years apart, each with a fresh sample of households.
- Pooling increases the sample size and allows comparison over time.

4. Panel (longitudinal) data. Observations on the *same* units over several periods.

- For example, the sales, wages and capital of the same 100 firms every year from 2015 to 2024.
- Panel data combine the features of cross-section and time series: we can study differences between units as well as changes within each unit over time.
- A panel with N units and T periods has NT observations.

5. Qualitative data and dummy variables.

- Some characteristics are not measured on a numerical scale, for example sex, region (rural or urban), religion, or whether a firm exports.
- They are recorded through *dummy (indicator) variables* that take the value 1 if the characteristic is present and 0 if it is absent.
- For example, $D_i = 1$ for an urban household and $D_i = 0$ for a rural household.
- The values 1 and 0 show presence or absence, not magnitude.

Recognising the type of data

Example 1.2 (Recognising the type of data)

Data set	Type
Wheat yield of 70 districts of Uttar Pradesh in 2024–25	Cross-section
Monthly rainfall at Meerut from 2000 to 2025	Time series
Household surveys of 2011–12 and 2022–23, each with a new sample	Pooled cross-section
Profits of the same 50 companies for each year 2015–2024	Panel
Whether each sampled household owns a two-wheeler (1 = yes, 0 = no)	Qualitative (dummy variable)

Sources of data

Sources of data.

- Data may be *primary* (collected by the investigator through a survey or experiment) or *secondary* (already collected and published by another agency).
- In India, important secondary sources are the surveys of the National Statistical Office (NSO), the publications of the Reserve Bank of India, and the Census.
- Economic data are mostly *non-experimental*: the investigator observes them but cannot control them as in a laboratory experiment. This is one reason why econometrics needs special methods.

Quality of data

Quality of data.

- Errors of measurement, omissions, rounding and changes of definition over time reduce the reliability of results.
- Variables must refer to the same units, the same period and the same definitions before they are used together in one model.

After this lecture

Practice

- Exercise 1.3 (Exercises and Solutions booklet).
- R practical P01: Economic data structures.

Next lecture

Lecture 6: Objectives of econometric models; stages of model building.