Introduction to Econometrics

Prepared by

Aditya Korekallu Srinivasa

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute
New Delhi

Contents

1 The Scope and Foundations of Econometrics

Econometrics is the discipline that unifies economic theory, mathematics, and statistical inference to empirically analyze economic phenomena. It is not merely the measurement of economic relationships, but the systematic application of statistical tools to test, estimate, and forecast economic models, thereby providing empirical content to economic theory.

1.1 Why Econometrics?

While economic theory postulates qualitative relationships (e.g., the law of demand), econometrics quantifies these relationships, answering questions such as: by how much does quantity demanded change for a given change in price? Mathematical economics provides the formal structure, but econometrics is essential for empirical verification and policy relevance.

1.2 Econometric Modelling: From Theory to Data

The econometrician's workflow is a sequence of steps:

1. **Statement of Theory or Hypothesis**: For example, Keynes' consumption theory postulates that consumption increases with income, but less than proportionally.

2. **Specification of the Mathematical Model**: The theory is expressed in equations, such as $Y = \beta_1 + \beta_2 X$, where Y is consumption, X is income, and β_2 (the marginal propensity to consume) is expected to be between 0 and 1.

3. **Specification of the Econometric Model**: Recognizing that real-world relationships are not exact, a disturbance term u is added: $Y = \beta_1 + \beta_2 X + u$.

4. **Obtaining Data**: Data may be time series, cross-sectional, or panel, and must be carefully sourced and validated.

5. **Estimation of Parameters**: Using regression analysis, parameters like β_1 and β_2 are estimated from data.

6. **Hypothesis Testing**: Statistical inference is used to test whether the estimated parameters align with theoretical expectations.

7. **Forecasting or Prediction**: The model is used to predict future values or simulate policy scenarios.

8. **Policy Analysis and Control**: Estimated models inform policy by quantifying the impact of interventions.

1.3 Types of Data in Econometrics

- **Time Series Data**: Observations on a variable over time (e.g., annual GDP).

- **Cross-Section Data**: Observations on multiple units at a single point in time (e.g., household income in a particular year).

- **Panel Data**: Combines both, observing multiple units over time.

1.4 Nature and Quality of Economic Data

Most economic data are nonexperimental, i.e., observational, which introduces challenges such as omitted variables, measurement error, and limited control over confounding factors. Data quality, aggregation level, and measurement scales (ratio, interval, ordinal, nominal) must be considered in empirical analysis.

2 Methodology of Econometrics

The methodology of econometrics is systematic, iterative, and grounded in both theory and empirical reality. Each step is essential for ensuring the validity and usefulness of empirical results.

2.1 Step 1: Statement of Theory or Hypothesis

Economic theory provides the conceptual framework. For instance, Keynesian theory suggests that consumption increases with income, but the rate of increase (MPC) is less than one.

2.2 Step 2: Specification of the Mathematical Model

The theory is formalized as a mathematical equation, such as $Y = \beta_1 + \beta_2 X$, where Y is the dependent variable, X is the independent variable, and β_1, β_2 are parameters.

2.3 Step 3: Specification of the Econometric Model

To account for real-world randomness and omitted factors, we introduce a stochastic disturbance term:

$$Y = \beta_1 + \beta_2 X + u$$

where u captures all unobserved influences.

2.4 Step 4: Data Collection and Description

Relevant data must be identified, collected, and described. Data may be sourced from government agencies, international organizations, or private databases. The choice of data type (time series, cross-section, panel) affects the modeling strategy.

2.5 Step 5: Estimation of Parameters

Parameters are estimated using appropriate statistical techniques, most commonly regression analysis. For example, the method of Ordinary Least Squares (OLS) provides estimates of β_1 and β_2.

2.6 Step 6: Hypothesis Testing

Hypothesis testing determines whether the estimated parameters are statistically significant and consistent with theoretical expectations. For example, is the estimated MPC significantly less than one?

2.7 Step 7: Forecasting or Prediction

The estimated model is used for forecasting future values of the dependent variable or simulating policy interventions.

2.8 Step 8: Policy Analysis and Model Revision

Models are used for policy analysis and, if necessary, revised in light of new data or diagnostic checks. Competing models are compared based on their empirical fit and theoretical coherence.

3 Regression Analysis: The Core of Econometric Methodology

Regression analysis is the principal tool of econometrics, allowing us to quantify the dependence of one variable on others.

3.1 The Simple Linear Regression Model

The basic regression model is:

$$Y = \beta_1 + \beta_2 X + u$$

where Y is the dependent variable, X is the independent variable, and u is the error term.

3.2 Interpretation

- β_2 measures the marginal effect of X on Y (e.g., the increase in consumption for a one-unit increase in income).
- The error term u captures all other influences on Y not explicitly modeled.

3.3 Estimation (Ordinary Least Squares)

OLS estimates the parameters by minimizing the sum of squared residuals:

$$S = \sum_{i=1}^{n} (Y_i - b_1 - b_2 X_i)^2$$

The estimators b_1 and b_2 are chosen to best fit the observed data.

3.4 Population vs. Sample Regression

- **Population regression function**: The true (but unobservable) relationship in the population.
- **Sample regression function**: The estimated relationship from observed data.

3.5 Causality and Model Specification

Regression quantifies association, not causation. Establishing causality requires theoretical justification and careful control for confounding factors. Model misspecification (e.g., omitted variables, incorrect functional form) can bias estimates.

4 Assumptions of the Classical Linear Regression Model

The reliability of OLS estimates depends on several key assumptions. Each is critical for valid inference.

4.1 Linearity in Parameters

Assumption: The model is linear in parameters (β_1, β_2). This allows for non-linear relationships in variables (e.g., quadratic or log-linear forms), provided the parameters enter linearly.

Example: Modeling crop yield as $Y = \beta_1 + \beta_2 X + \beta_3 X^2 + u$ is linear in parameters.

4.2 Random Sampling

Assumption: Data are a random sample from the population.

Example: If only high-income households are surveyed, the results may not generalize to the entire population.

4.3 No Perfect Multicollinearity

Assumption: No explanatory variable is an exact linear function of others.

Example: Including both "total expenditure" and "income" as regressors when they are always equal leads to perfect multicollinearity.

4.4 Zero Conditional Mean

Assumption: $E[u|X] = 0$; the error term has zero mean given the explanatory variables.

Example: If unobserved ability affects both education and wages, omitting ability violates this assumption.

4.5 Homoscedasticity

Assumption: The variance of the error term is constant for all values of X.

Example: In income regressions, error variance may increase with income, violating homoscedasticity.

4.6 No Autocorrelation

Assumption: Error terms are uncorrelated across observations.

Example: In time series data, shocks may persist over time, inducing autocorrelation.

4.7 Normality of Errors

Assumption: The error term is normally distributed.

Example: In large samples, normality is less critical due to the Central Limit Theorem, but in small samples, it is important for valid inference.

Glossary

- **Econometrics**: The application of statistical and mathematical methods to the analysis of economic data.
- **Regression Analysis**: A statistical technique for estimating the relationships among variables.
- **Ordinary Least Squares (OLS)**: A method for estimating the parameters of a linear regression model by minimizing the sum of squared residuals.
- **Panel Data**: Data containing observations on multiple entities over multiple time periods.
- **Homoscedasticity**: The property that the error variance is constant across all observations.
- **Autocorrelation**: Correlation of error terms across observations, often a problem in time series data.
- **Endogeneity**: When an explanatory variable is correlated with the error term, violating the zero conditional mean assumption.
- **Specification**: The process of translating economic theory into an empirically testable model.
- **Time Series Data**: Data observed over time for a single unit.
- **Cross-Section Data**: Data observed at a single point in time across multiple units.

Practice Questions

1. Explain the distinction between economic theory, mathematical economics, and econometrics.
2. Outline the steps in the methodology of econometric modeling and explain the importance of each.
3. Define regression analysis and discuss its role in empirical economics.
4. List and explain the assumptions of the classical linear regression model, providing a real-world example for each.
5. Discuss the consequences of violating the zero conditional mean assumption.

6. How would you address the presence of heteroscedasticity in an empirical study?
7. Provide an example of model misspecification and discuss its implications.
8. Explain the difference between time series, cross-sectional, and panel data. Why does the data structure matter for econometric analysis?
9. Critically evaluate the statement: "Regression analysis is the backbone of empirical economic research."
10. What are the main challenges in using nonexperimental (observational) data in econometrics?

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.