Chapter 1

The Nature of Applied Econometrics
Section 1.1

What Is Applied Econometrics?
Applied Econometrics

- Applied econometrics deals with the measurement of business, economic, or financial relationships.

“The study of the application of statistical methods to the analysis of economic phenomena.” (Tintner 1953)

- Business analysts often need to be in position to do the following:
  - interpret the economic landscape
  - identify and assess the influence of several *exogenous* or *predetermined* factors on one or more *endogenous* variables
  - provide ex-ante forecasts of one or more endogenous variables

How do you achieve these objectives?
Why Do Business Analysts Want to Achieve These Objectives?

To improve decision making!

Example: Investigate key determinants of demand for Prego spaghetti sauce:

- price of Prego
- price of competitors (Ragu, Classico, Hunt’s, Newman’s Own)
- in-store displays
- coupons
- price of pasta

Forecast sales of Prego spaghetti sauce one month, one quarter, or even one year into the future.
Section 1.2

Course of Action – Development of Formal Quantitative Models
Basic Approaches To Model Building

- Intuitive approach
- Bayesian models (to bridge the gap between intuitive models (prior beliefs) and more formal statistical models)
- Time-Series approach
- Structural (Econometric) approach
Section 1.3

Disciplines in Applied Econometrics
Disciplines in Applied Econometrics

- Economics, mathematics, and statistics are combined.
- Economic theory establishes the boundaries of the questions you can address.
- Empirical models are associated with some sort of mathematical representation.
- Derivatives are the cornerstones of the calculation of marginal effects.
- Statistics provide the vehicle for making inductive inferences from a sample to a population.
- Little opportunity exists for controlled experiments.
- Econometrics emphasizes the importance of Monte Carlo simulation as an alternative to controlled experiments.
Section 1.4

Empirical Models and Modeling Approaches
Empirical Models

- At the heart of applied econometrics lies the development of empirical models.
- Empirical models are nothing more than parsimonious representations of reality.
- Empirical models indigenous to applied econometrics are stochastic, not deterministic.
- The purpose is to model the behavior of agents such as consumers, firms, banks, and the stock market, all of which have in common the human element.
- Empirical models can be single-equation or multi-equation specifications.
- Empirical models can be linear or nonlinear in parameters.
In some instances, econometric models might encompass elements of time-series models (for example, ARCH and GARCH models, adjustments for serial correlation).
Model Specification

Model specification is driven by the following:

- Data (number of observations)
- Data (number of variables)
- Theoretical Considerations
- A theoretical Considerations
- Dynamics
- Experience
- Question(s) Addressed
Intuitive Approach

The basic premise is that any data series (typically a data series that varies over time) can be decomposed into a trend component, a seasonal component, a cyclical component, and a random component. This decomposition can either be additive or multiplicative. This approach only deals with a single data series and, as such, is univariate in nature.

With the intuitive approach, analysts focus on models designed to capture trends, seasonal movements, and cyclical patterns.

Time series models also capture these components but in a more sophisticated fashion.
Time-Series Approach

- ARIMA models – Univariate approach
- Vector autoregression models – Multivariate approach
- Causality tests
- Impulse response functions
- Cointegration
- Error correction models – Multivariate approach
- Other models (for example, ARCH and GARCH models)
Structural or Econometric Approach

- Single-equation specifications – Multivariate approach
- Multi-equation specifications – Multivariate approach
- Interpretation of coefficients
- Tests of hypotheses
Applied Econometrics

Simply put, applied econometrics entails the following:

- unification of economics, mathematics, and statistics
- measurement of business, economic, or financial relationships
- measurement of underlying relationships for the purpose of testing and developing economic theory (a form of data mining)
Section 1.5

Components of Applied Econometrics
Components of Applied Econometrics

Applied econometrics involves four phases of analysis:

- Specification – the model-building activity
- Estimation – fitting the model to data
- Verification – testing the model
- Prediction – producing ex-ante forecasts and conducting ex-post forecast evaluations
Components of Applied Econometrics

- Specification
- Estimation
- Verification
- Prediction
The Translation of Economic Models into Statistical Models

- A key issue in applied econometrics is the use of proxy variables. Proxy variables generally arise due to unavailability of appropriate data.

“Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many milliseconds to bake the mixture at exactly 474 degrees of temperature. But when the statistical cook turns to raw materials, he finds that hearts of cactus fruit are unavailable, so he substitutes chunks of cantaloupe; where the recipe calls for vermicelli he uses shredded wheat; and he substitutes green garment dye for curry, ping-pong balls for turtle’s eggs, and for a vintage bottle (1883) of champagne, a can of turpentine.” (Valavanis 1959, p. 83)
The Applied Econometric Approach

Theory

Model

Econometric Model

Data

Refined Data

Statistical Theory

Econometric Techniques

Estimation of the Econometric Model with the Refined Data Using Econometric Techniques

Structural Analysis (Marginal Effects)

Forecasting

Policy Evaluation
Section 1.6

Products of Applied Econometrics
1. Structural analysis (marginal effects)

2. Forecasting

3. Policy evaluation

Interrelated By-Products of Applied Econometrics
Structural Analysis and Forecasting

- Structural analysis – Use of estimated econometric models for the quantitative measurement of economic relationships
- Forecasting – Use of estimated econometric models to predict quantitative values of certain variables outside the sample of data actually observed
- Confidence or tolerance intervals of forecasts, taking into account standard errors
- Conditional versus unconditional forecasts
The Structure of Applied Research

- Problem Definition
- Model Specification
- Model Definition (variables)
- Data Assembly
- Descriptive Analysis
- Analytical Phase
- Evaluate and Interpret Results
- REPORT
Section 1.7

Getting Started
Applied Econometrics and Model Specification

Applied econometrics begins with model specification:

- Model specification entails the expression of theoretical constructs in mathematical terms.
- This phase of applied econometrics constitutes the model building activity.
- In essence, model specification is the translation of theoretical constructs into mathematical/statistical forms.
- Fundamental principles in model building:
  - The principle of parsimony (If other things are the same, simple models generally are preferable to complex models, especially in forecasting.)
  - The shrinkage principle (Imposing restrictions either on estimated parameters or on forecasts often improves model performance.)
  - The KISS principle (Zellner, 2001) “Keep it Sophistically Simple”
Example: Estimation of Demand Relationships

- Often in applied econometrics, analysts have an interest in estimating demand relationships, particularly for commodities.
- Analysts might want to estimate the demand for cosmetic products, automobiles, various food products, or various beverages.
Demand Curve

- The *Demand Curve* shows the theoretical relationship between price and quantity demanded, holding all other factors constant.
- Axes: Price is on the $y$-axis, and quantity on the $x$-axis.
- Example: Demand curve for Lipton tea:
  \[ Q = 2500 - 500P \]
- Key question: How are these numbers obtained?
Demand Curve for Lipton Tea

Demand Curve

\[ Q = 2,500 - 500P \]

\[ P = 5 - 0.002Q \]

Average price per package

Packages of Lipton tea
Own-Price Effects

- **Movement** along a given demand curve reflects a change in the price and quantity of the commodity in question.

- **Answer the following question:** What happens to the quantity demanded if the price changes (exogenously) but holds all other factors constant?

- **Law of demand:**
  - The demand curve is negatively sloped.
  - An increase (a decrease) in price leads to a decrease (an increase) in quantity demanded.
  - This is an important empirical finding in economics.
Other Factors Affecting Demand

- Prices of other products
- Income
- Advertising (positive information)
- Food recalls or food scares (negative information)
- Health and nutrition factors
- Lifestyle factors
- Tastes and preferences
Translations of the Theoretical Construct into a Statistical Model

1. \( Q = a - bP \)

2. \( Q = a_0 - a_1P + a_2l + a_3A + a_4PS \)

- The coefficients \( a_0, a_1, a_2, a_3, \) and \( a_4 \) are labeled the demand parameters.
- You expect certain signs and magnitudes of the demand parameters according to economic theory.
- Different versions of the econometric model for applied analysis are possible.
Linear Regression Model Development

- The starting point rests on the development of a linear regression model.
- Philosophical differences exist among statisticians and applied econometric analysts.
- Attention centered on the signs and magnitudes of the coefficients are inherent in the model.
- Model specifications in applied econometrics are non-unique.
- Model specifications provide the blueprint for action.
Section 1.8

The Generic Multiple Regression Model
The Generic Multiple Regression Model

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, \ldots, n \]

\[ \frac{\partial Y_i}{\partial X_{ji}} = \beta_j \]

\[ Y = X\beta + \varepsilon \]

\[
\begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{bmatrix}_{nx1}
= 
\begin{bmatrix}
1 & X_{11} & X_{12} & \ldots & X_{1k} \\
1 & X_{21} & X_{22} & \ldots & X_{2k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & X_{n1} & X_{n2} & \ldots & X_{nk}
\end{bmatrix}_{nx(k+1)}
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_k
\end{bmatrix}_{(k+1)x1}
\]

\[
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_n
\end{bmatrix}_{nx1}
\]

Estimation of regression parameters:
- Least squares (No knowledge of the distribution of the error or disturbance terms is required.)
- Maximum likelihood (requires knowledge of the distribution of the error or disturbance terms)
- The use of the matrix notation enables a view of how the data is housed in econometric software programs.
Components of the Model

- Endogenous variables – dependent variables, values of which are determined within the system
- Exogenous variables – determined outside the system but influence the system by affecting the values of the endogenous variables
- Structural parameters – estimated using econometric techniques and relevant data
- Lagged endogenous variables
- Lagged exogenous variables
- Predetermined variables
The Disturbance (or Error) Term

Stochastic, a random variable
Statistical distribution often normal

Captures:
1. Omission of the influence of other variables
2. Measurement error

There is the recognition that any econometric model is a parsimonious stochastic representation of reality.
Assumptions

1. Model linear in parameters
2. \( E(\varepsilon) = 0 \)
3. \( E(\varepsilon\varepsilon^T) = \sigma^2 I \)
   - a. Homoscedasticity
   - b. No serial or autocorrelation
4. \( X \) is a \( nx(k+1) \) matrix of fixed numbers (or alternatively \( X \) can be stochastic but \( E(\varepsilon|x) = 0 \)).
   \[ \Rightarrow E[X^T\varepsilon] = X^TE[\varepsilon] = 0 \]
5. Rank of \( X = r(X) = k+1 < n \) (\( x \) is full column rank meaning that you cannot write one variable as a perfect linear combination of other variables.)
6. \( \varepsilon \sim N(0, \sigma^2 I) \)
   The normality of the error terms enables the use of parametric distributions (for example, the \( t \) or \( F \) distribution) in performing tests of hypotheses.
Econometric Models

- Special type of algebraic model, namely stochastic (as opposed to deterministic)
  - linear or nonlinear in parameters
  - inherently linear versus nonlinear models
  - additive stochastic disturbance term that plays the role of a chance medium

- In general, each equation of an econometric model, other than definitions, equilibrium conditions, and identities, is assumed to contain an additive disturbance term.
Inherently Linear Models versus Inherently Nonlinear Models

- An inherently linear model is a model that can be transformed to the linear form.

\[ Y = F(X_1, X_2, \ldots, X_k, \varepsilon) \]

\[ F(y) = \beta_0 + \beta_1 g_1(X_1, X_2, \ldots, X_k) + \beta_2 g_2(X_1, X_2, \ldots, X_k) + \ldots + \beta_k g_k(X_1, X_2, \ldots, X_k) + \varepsilon \]

For example, \( Y = \sigma_0 X_1^{\sigma_1} X_2^{\sigma_2} \ldots X_k^{\sigma_k} \exp(\varepsilon) \)

\[ \ln Y = \ln \sigma_0 + \sigma_1 \ln X_1 + \sigma_2 \ln X_2 + \ldots + \sigma_k \ln X_k + \varepsilon \]

- Polynomial model, exponential model, reciprocal model, semi-logarithmic model, interaction model
Perils, Problems, and Pitfalls

- **Degrees-of Freedom Problem**
  - Not enough observations to enable adequate or reasonable estimates of the model.

- **Collinearity Problem**
  - Tendency of the data to bunch or move together. For example, in time-series data, the variables tend to exhibit the same trends, cyclical and secular, over time. Interdependence exists among regressors or explanatory variables in the model.
  - A common problem in economic analysis.
  - A problem of linear dependence.

continued...
Perils, Problems, and Pitfalls

- Serial Correlation Problem
  - Predominantly a time-series problem. Underlying changes occur very slowly over time to the extent that the stochastic disturbance term represents conditions relevant to the model but not explicitly accounted for in it (such as omitted variables). Serial correlation represents the dependence of stochastic disturbance terms in different periods.

- Heteroscedasticity Problem
  - The variance of the error terms is not constant (typically a problem with cross-sectional data).

- Structural Change Problem
  - Parameters might not be constant over time.
Trade Offs

- Structural specification versus forecasting
- Functional form
- Single-equation versus multi-equation specifications
- Time and effort to construct models
- Choice of econometric or time-series models

Replication

“It will be remembered that the 70 translators of the Septuagint were shut up in 70 separate rooms with the Hebrew text and brought out with them, when they emerged, 70 identical translations. Would the same miracle be vouchsafed if 70 multiple correlators were shut up with the same statistical material?”

John Maynard Keynes (1940)
Example of a Structural Model

The effect of unionization on earnings hypothesis:

*Unionization increases real wages.*

\[ \ln w = 0.28 \ln Q + 0.77 Z + 0.84 \ln P + 0.228 U - 0.16 UZ \]

\[ (0.056) \quad (0.28) \quad (0.32) \quad (0.094) \quad (1.10) \]

\[-1.94 U \ln P + \text{CONSTANT} \]

\[ (0.99) \]

\[ R^2 = 0.891 \]

\( W = \) relative wages (average hourly compensation in unionized industries relative to compensation in nonunionized industries)

\( Q = \) relative value of output

\( Z = \) unemployment rate

\( P = \) measure of expected inflation

\( U = \) measure of the extent of union membership

\( UZ, U \ln P = \) interaction variables
Another Example of a Structural Model

Specification based on change rather than levels, for example, estimation of a demand equation for new automobiles in the United States

\[
\Delta A = 0.115 + 0.106\Delta I - 0.507\Delta S - 0.234\Delta(p/m) - 0.827PS
\]

\[R^2 = 0.85\]

\(A\) = annual retail sales of new passenger automobiles

\(\Delta A\) = change in annual retail sales

\(\Delta I\) = change in real disposable income

\(\Delta S\) = change in the stock of passenger automobiles

\(\Delta(p/m)\) = change in average retail price deflated by the average duration of automobile loans

\(PS\) = dummy variable to take account of years of severe production shortages
Section 1.9

Software Considerations
Concerns about Software

Owing to the quantitative nature of the economic modeling and forecasting process, “number crunching” is a fact of life.

- Importance of Software Packages
- Characteristics of Package Choice:
  - Ease of use
  - Provides relevant information to carry out tasks

**No single package is optimal** for every situation.
Use of SAS

Common SAS procedures that are relative to the contents of this course are as follows:

- MEANS procedure
- UNIVARIATE procedure
- CORR procedure
- REG procedure
- AUTOLOG procedure
- MODEL procedure
- QLIM procedure
Section 1.10

Communication and Aims for the Analyst
Communication

- A technician can run a program and get output.
- An analyst must interpret the findings from examination of this output.
- There are no bonus points to be given to terrific hackers but poor analysts.

Aims

1. Improve your ability in developing models to conduct structural analysis and to forecast with some accuracy.
2. Enhance your ability in interpreting and communicating the results so as to improve decision-making.

Bottom Line

1. The analyst transforms the economic model/idea to a mathematical/statistical one.
2. The technician estimates the model and obtains a mathematical/statistical answer.
3. The analyst transforms the mathematical/statistical answer to an economic one.