PART ONE
Preliminaries

Part 1 covers the essential components of microeconometric analysis – an economic specification, a statistical model and a data set.

Chapter 1 discusses the distinctive aspects of microeconometrics, and provides an outline of the book. It emphasizes that discreteness of data, and nonlinearity and heterogeneity of behavioral relationships are key aspects of individual-level microeconometric models. It concludes by presenting the notation and conventions used throughout the book.

Chapters 2 and 3 set the scene for the remainder of the book by introducing the reader to key model and data concepts that shape the analyses of later chapters.

A key distinction in econometrics is between essentially descriptive models and data summaries at various levels of statistical sophistication and models that go beyond associations and attempt to estimate causal parameters. The classic definitions of causality in econometrics derive from the Cowles Commission simultaneous equations models that draw sharp distinctions between exogenous and endogenous variables, and between structural and reduced form parameters. Although reduced form models are very useful for some purposes, knowledge of structural or causal parameters is essential for policy analyses. Identification of structural parameters within the simultaneous equations framework poses numerous conceptual and practical difficulties. An increasingly-used alternative approach based on the potential outcome model, also attempts to identify causal parameters but it does so by posing limited questions within a more manageable framework. Chapter 2 attempts to provide an overview of the fundamental issues that arise in these and other alternative frameworks. Readers who initially find this material challenging should return to this chapter after gaining greater familiarity with specific models covered later in the book.

The empirical researcher's ability to identify causal parameters depends not only on the statistical tools and models but also on the type of data available. An experimental framework provides a standard for establishing causal connections. However, observational, not experimental, data form the basis of much of econometric inference. Chapter 3 surveys the pros and cons of three main types of data: observational data, data from social experiments, and data from natural experiments. The strengths and weaknesses of conducting causal inference based on each type of data are reviewed.

CHAPTER 1
Overview

1.1. Introduction

This book provides a detailed treatment of **microeconometric analysis**, the analysis of individual-level data on the economic behavior of individuals or firms. A broader definition would also include grouped data. Usually regression methods are applied to cross-section or panel data.

Analysis of individual data has a long history. Ernst Engel (1857) was among the earliest quantitative investigators of household budgets. Allen and Bowley (1935), Houthakker (1957), and Prais and Houthakker (1955) made important contributions following the same research and modeling tradition. Other landmark studies that were also influential in stimulating the development of microeconometrics, even though they did not always use individual-level information, include those by Marschak and Andrews (1944) in production theory and by Wold and Jureen (1953), Stone (1953), and Tobin (1958) in consumer demand.

As important as the above earlier cited work is on household budgets and demand analysis, the material covered in this book has stronger connections with the work on discrete choice analysis and censored and truncated variable models that saw their first serious econometric applications in the work of McFadden (1973, 1984) and Heckman (1974, 1979), respectively. These works involved a major departure from the overwhelming reliance on linear models that characterized earlier work. Subsequently, they have led to significant methodological innovations in econometrics. Among the earlier textbook-level treatments of this material (and more) are the works of Maddala (1983) and Amemiya (1985). As emphasized by Heckman (2001), McFadden (2001), and others, many of the fundamental issues that dominated earlier work based on market data remain important, especially concerning the conditions necessary for identifiability of causal economic relations. Nonetheless, the style of microeconometrics is sufficiently distinct to justify writing a text that is exclusively devoted to it.

Modern microeconometrics based on individual-, household-, and establishment-level data owes a great deal to the greater availability of data from cross-section and longitudinal sample surveys and census data. In the past two decades, with the expansion of electronic recording and collection of data at the individual level, data volume has grown explosively. So too has the available computing power for analyzing large and complex data sets. In many cases event-level data are available; for example, marketing science often deals with purchase data collected by electronic scanners in supermarkets, and industrial organization literature contains econometric analyses of airline travel data collected by online booking systems. There are now new branches of economics, such as social experimentation and experimental economics, that generate "experimental" data. These developments have created many new modeling opportunities that are absent when only aggregated market-level data are available. Meanwhile the explosive growth in the volume and types of data has also given rise to numerous methodological issues. Processing and econometric analysis of such large microdatabases, with the objective of uncovering patterns of economic behavior, constitutes the core of microeconometrics. Econometric analysis of such data is the subject matter of this book.

Key precursors of this book are the books by Maddala (1983) and Amemiya (1985). Like them it covers topics that are presented only briefly, or not at all, in undergraduate and first-year graduate econometrics courses. Especially compared to Amemiya (1985) this book is more oriented to the practitioner. The level of presentation is nonetheless advanced in places, especially for applied researchers in disciplines that are less mathematically oriented than economics.

A relatively advanced presentation is needed for several reasons. First, the data are often discrete or censored, in which case **nonlinear methods** such as logit, probit, and Tobit models are used. This leads to statistical inference based on more difficult asymptotic theory.

Second, **distributional assumptions** for such data become critically important. One response is to develop highly parametric models that are sufficiently detailed to capture the complexities of data, but these models can be challenging to estimate. A more common response is to minimize parametric assumptions and perform statistical inference based on standard errors that are "robust" to complications such as heteroskedasticity and clustering. In such cases considerable knowledge can be needed to ensure valid statistical inference even if a standard regression package is used.

Third, economic studies often aim to determine **causation** rather than merely measure correlation, despite access to observational rather than experimental data. This leads to methods to isolate causation such as instrumental variables, simultaneous equations, measurement error correction, selection bias correction, panel data fixed effects, and differences-in-differences.

Fourth, microeconomic data are typically collected using cross-section and panel surveys, censuses, or social experiments. **Survey data** collected using these methods are subject to problems of complex survey methodology, departures from simple random sampling assumptions, and problems of sample selection, measurement errors, and incomplete, and/or missing data. Dealing with such issues in a way that can support valid population inferences from the estimated econometric models population requires use of advanced methods.

Finally, it is not unusual that two or more **complications occur simultaneously**, such as endogeneity in a logit model with panel data. Then a cookbook approach becomes very difficult to implement. Instead, considerable understanding of the theory underlying the methods is needed, as the researcher may need to read econometrics journal articles and adapt standard econometrics software.

1.2. Distinctive Aspects of Microeconometrics

We now consider several advantages of microeconometrics that derive from its distinctive features.

1.2.1. Discreteness and Nonlinearity

The first and most obvious point is that microeconometric data are usually at a low level of aggregation. This has a major consequence for the functional forms used to analyze the variables of interest. In many, if not most, cases linear functional forms turn out to be simply inappropriate. More fundamentally, disaggregation brings to the forefront **heterogeneity** of individuals, firms, and organizations that should be properly controlled (modeled) if one wants to make valid inferences about the underlying relationships. We discuss these issues in greater detail in the following sections.

Although aggregation is not entirely absent in microdata, as for example when household- or establishment-level data are collected, the level of aggregation is usually orders of magnitude lower than is common in macro analyses. In the latter case the process of aggregation leads to smoothing, with many of the movements in opposite directions canceling in the course of summation. The aggregated variables often show smoother behavior than their components, and the relationships between the aggregates frequently show greater smoothness than the components. For example, a relation between two variables at a micro level may be piecewise linear with many nodes. After aggregation the relationship is likely to be well approximated by a smooth function. Hence an immediate consequence of disaggregation is the absence of features of continuity and smoothness both of the variables themselves and of the relationships between them.

Usually individual- and firm-level data cover a huge range of variation, both in the cross-section and time-series dimensions. For example, average weekly consumption of (say) beef is highly likely to be positive and smoothly varying, whereas that of an individual household in a given week may be frequently zero and may also switch to positive values from time to time. The average number of hours worked by female workers is unlikely to be zero, but many individual females have zero market hours of work (corner solutions), switching to positive values at other times in the course of their labor market history. Average household expenditure on vacations is usually positive, but many individual households may have zero expenditure on vacations in any given year. Average per capita consumption of tobacco products will usually be positive, but many individuals in the population have never consumed these products and never will, irrespective of price and income considerations. As Pudney (1989) has observed, microdata exhibit "holes, kinks and corners." The holes correspond to nonparticipation in the activity of interest, kinks correspond to the switching behavior, and corners correspond to the incidence of nonconsumption or nonparticipation at specific points of time. That is, discreteness and nonlinearity of response are intrinsic to microeconometrics.

An important class of nonlinear models in microeconometrics deals with **limited dependent variables** (Maddala, 1983). This class includes many models that provide suitable frameworks for analyzing discrete responses and responses with limited range of variation. Such tools of analyses are of course also available for analyzing macrodata, if required. The point is that they are indispensable in microeconometrics and give it its distinctive feature.

1.2.2. Greater Realism

Macroeconometrics is sometimes based on strong assumptions; the representative agent assumption is a leading example. A frequent appeal is made to microeconomic reasoning to justify certain specifications and interpretations of empirical results. However, it is rarely possible to say explicitly how these are affected by aggregation over time and micro units. Alternatively, very extreme aggregation assumptions are made. For example, aggregates are said to reflect the behavior of a hypothetical representative agent. Such assumptions also are not credible.

From the viewpoint of microeconomic theory, quantitative analysis founded on microdata may be regarded as more realistic than that based on aggregated data. There are three justifications for this claim. First, the measurement of the variables involved in such hypotheses is often more direct (though not necessarily free from measurement error) and has greater correspondence to the theory being tested. Second, hypotheses about economic behavior are usually developed from theories of individual behavior. If these hypotheses are tested using aggregated data, then many approximations and simplifying assumptions have to be made. The simplifying assumption of a representative agent causes a great loss of information and severely limits the scope of an empirical investigation. Because such assumptions can be avoided in microeconometrics, and usually are, in principle the microdata provide a more realistic framework for testing microeconomic hypotheses. This is not a claim that the promise of microdata is necessarily achieved in empirical work. Such a claim must be assessed on a case-by-case basis. Finally, a realistic portrayal of economic activity should accommodate a broad range of outcomes and responses that are a consequence of individual heterogeneity and that are predicted by underlying theory. In this sense microeconomic data sets can support more realistic models.

Microeconometric data are often derived from household or firm surveys, typically encompassing a wide range of behavior, with many of the behavioral outcomes taking the form of discrete or categorical responses. Such data sets have many awkward features that call for special tools in the formulation and analysis that, although not entirely absent from macroeconometric work, nevertheless are less widely used.

1.2.3. Greater Information Content

The potential advantages of microdata sets can be realized if such data are informative. Because sample surveys often provide independent observations on thousands of cross-sectional units, such data are thought to be more informative than the standard, usually highly serially correlated, macro time series typically consisting of at most a few hundred observations.

As will be explained in the next chapter, in practice the situation is not so clear-cut because the microdata may be quite noisy. At the individual level many (idiosyncratic) factors may play a large role in determining responses. Often these cannot be observed, leading one to treat them under the heading of a random component, which can be a very large part of observed variation. In this sense randomness plays a larger role in microdata than in macrodata. Of course, this affects measures of goodness of fit of the regressions. Students whose initial exposure to econometrics comes through aggregate time-series analysis are often conditioned to see large R^2 values. When encountering cross-section regressions for the first time, they express disappointment or even alarm at the "low explanatory power" of the regression equation. Nevertheless, there remains a strong presumption that, at least in certain dimensions, large microdata sets are highly informative.

Another qualification is that when one is dealing with purely cross-section data, very little can be said about the intertemporal aspects of relationships under study. This particular aspect of behavior can be studied using panel and transition data.

In many cases one is interested in the behavioral responses of a specific group of economic agents under some specified economic environment. One example is the impact of unemployment insurance on the job search behavior of young unemployed persons. Another example is the labor supply responses of low-income individuals receiving income support. Unless microdata are used such issues cannot be addressed directly in empirical work.

1.2.4. Microeconomic Foundations

Econometric models vary in the explicit role given to economic theory. At one end of the spectrum there are models in which the a priori theorizing may play a dominant role in the specification of the model and in the choice of an estimation procedure. At the other end of the spectrum are empirical investigations that make much less use of economic theory.

The goal of the analysis in the first case is to identify and estimate fundamental parameters, sometimes called deep parameters, that characterize individual taste and preferences and/or technological relationships. As a shorthand designation, we call this the **structural approach**. Its hallmark is a heavy dependence on economic theory and emphasis on causal inference. Such models may require many assumptions, such as the precise specification of a cost or production function or specification of the distribution of error terms. The empirical conclusions of such an exercise may not be robust with respect to the departures from the assumptions. In Section 2.4.4 we shall say more about this approach. At the present stage we simply emphasize that if the structural approach is implemented with aggregated data, it will yield estimates of the fundamental parameters only under very stringent (and possibly unrealistic) conditions. Microdata sets provide a more promising environment for the structural approach, essentially because they permit greater flexibility in model specification.

The goal of the analysis in the second case is to model relationship(s) between response variables of interest conditionally on variables the researcher takes as given, or exogenous. More formal definitions of **endogeneity** and **exogeneity** are given in Chapter 2. As a shorthand designation, we call this a **reduced form approach**. The essential point is that reduced form analysis does not always take into account all causal interdependencies. A regression model in which the focus is on the prediction of y given regressors $\mathbf{x}$, and not on the causal interpretation of the regression parameters, is often referred to as a reduced form regression. As will be explained in Chapter 2, in general the parameters of the reduced form model are functions of structural parameters. They may not be interpretable without some information about the structural parameters.

1.2.5. Disaggregation and Heterogeneity

It is sometimes said that many problems and issues of macroeconometrics arise from serial correlation of macro time series, and those of microeconometrics arise from heteroskedasticity of individual-level data. Although this is a useful characterization of the modeling effort in many microeconometric analyses, it needs amplification and is subject to important qualifications. In a range of microeconometric models, modeling of dynamic dependence may be an important issue.

The benefits of disaggregation, which were emphasized earlier in this section, come at a cost: As the data become more disaggregated the importance of controlling for interindividual heterogeneity increases. Heterogeneity, or more precisely unobserved heterogeneity, plays a very important role in microeconometrics. Obviously, many variables that reflect interindividual heterogeneity, such as gender, race, educational background, and social and demographic factors, are directly observed and hence can be controlled for. In contrast, differences in individual motivation, ability, intelligence, and so forth are either not observed or, at best, imperfectly observed.

The simplest response is to ignore such heterogeneity, that is, to absorb it into the regression disturbance. After all this is how one treats the myriad small unobserved factors. This step of course increases the unexplained part of the variation. More seriously, ignoring persistent interindividual differences leads to **confounding** with other factors that are also sources of persistent interindividual differences. Confounding is said to occur when the individual contributions of different regressors (predictor variables) to the variation in the variable of interest cannot be statistically separated. Suppose, for example, that the factor x_1 (schooling) is said to be the source of variation in y (earnings), when another variable x_2 (ability), which is another source of variation, does not appear in the model. Then that part of total variation that is attributable to the second variable may be incorrectly attributed to the first variable. Intuitively, their relative importances are confounded. A leading source of confounding bias is the incorrect omission of regressors from the model and the inclusion of other variables that are proxies for the omitted variable.

Consider, for example, the case in which a program participation (0/1 dummy) variable D is included in the regression mean function with a vector of regressors $\mathbf{x}$,

$$y = \mathbf{x}'\boldsymbol{\beta} + \alpha D + u, \tag{1.1}$$

where u is an error term. The term "treatment" is used in biological and experimental sciences to refer to an administered regimen involving participants in some trial. In econometrics it commonly refers to participation in some activity that may impact an outcome of interest. This activity may be randomly assigned to the participants or may be self-selected by the participant. Thus, although it is acknowledged that individuals choose their years of schooling, one still thinks of years of schooling as a "treatment" variable. Suppose that program participation is taken to be a discrete variable. The coefficient α of the "treatment variable" measures the average impact of the program participation ($D = 1$), conditional on covariates. If one does not control for unobserved heterogeneity, then a potential ambiguity affects the interpretation of the results. If d is found to have a significant impact, then the following question arises: Is α significantly different from zero because D is correlated with some unobserved variable that affects y or because there is a causal relationship between D and y? For example, if the program considered is university education, and the covariates do not include a measure of ability, giving a fully causal interpretation becomes questionable. Because the issue is important, more attention should be given to how to control for unobserved heterogeneity.

In some cases where dynamic considerations are involved the type of data available may put restrictions on how one can control for heterogeneity. Consider the example of two households, identical in all relevant respects except that one exhibits a systematically higher preference for consuming good A. One could control for this by allowing individual utility functions to include a heterogeneity parameter that reflects their different preferences. Suppose now that there is a theory of consumer behavior that claims that consumers become addicted to good A, in the sense that the more they consume of it in one period, the greater is the probability that they will consume more of it in the future. This theory provides another explanation of persistent interindividual differences in the consumption of good A. By controlling for heterogeneous preferences it becomes possible to test which source of persistence in consumption – preference heterogeneity or addiction – accounts for different consumption patterns. This type of problem arises whenever some dynamic element generates persistence in the observed outcomes. Several examples of this type of problem arise in various places in the book.

A variety of approaches for modeling heterogeneity coexist in microeconometrics. A brief mention of some of these follows, with details postponed until later.

An extreme solution is to ignore all unobserved interindividual differences. If unobserved heterogeneity is uncorrelated with observed heterogeneity, and if the outcome being studied has no intertemporal dependence, then the aforementioned problems will not arise. Of course, these are strong assumptions and even with these assumptions not all econometric difficulties disappear.

One approach for handling heterogeneity is to treat it as a **fixed effect** and to estimate it as a coefficient of an individual specific 0/1 dummy variable. For example, in a cross-section regression, each micro unit is allowed its own dummy variable (intercept). This leads to an extreme proliferation of parameters because when a new individual is added to the sample, a new intercept parameter is also added. Thus this approach will not work if our data are cross sectional. The availability of multiple observations per individual unit, most commonly in the form of panel data with T time-series observations for each of the N cross-section units, makes it possible to either estimate or eliminate the fixed effect, for example by first differencing if the model is linear and the fixed effect is additive. If the model is nonlinear, as is often the case, the fixed effect will usually not be additive and other approaches will need to be considered.

A second approach to modeling unobserved heterogeneity is through a **random effects** model. There are a number of ways in which the random effects model can be formulated. One popular formulation assumes that one or more regression parameters, often just the regression intercept, varies randomly across the cross section. In another formulation the regression error is given a component structure, with an individual specific random component. The random effects model then attempts to estimate the parameters of the distribution from which the random component is drawn. In some cases, such as demand analysis, the random term can be interpreted as random preference variation. Random effects models can be estimated using either cross-section or panel data.

1.2.6. Dynamics

A very common assumption in cross-section analysis is the absence of intertemporal dependence, that is, an absence of dynamics. Thus, implicitly it is assumed that the observations correspond to a stochastic equilibrium, with the deviation from the equilibrium being represented by serially independent random disturbances. Even in microeconometrics for some data situations such an assumption may be too strong. For example, it is inconsistent with the presence of serially correlated unobserved heterogeneity. Dependence on lagged dependent variables also violates this assumption.

The foregoing discussion illustrates some of the potential limitations of a single cross-section analysis. Some limitations may be overcome if repeated cross sections are available. However, if there is dynamic dependence, the least problematic approach might well be to use panel data.

1.3. Book Outline

The book is split into six parts. Part 1 presents the issues involved in microeconometric modeling. Parts 2 and 3 present general theory for estimation and statistical inference for nonlinear regression models. Parts 4 and 5 specialize to the core models used in applied microeconometrics for, respectively, cross-section and panel data. Part 6 covers broader topics that make considerable use of material presented in the earlier chapters.

The book outline is summarized in Table 1.1. The remainder of this section details each part in turn.

1.3.1. Part 1: Preliminaries

Chapters 2 and 3 expand on the special features of the **microeconometric** approach to modeling and **microeconomic data structures** within the more general statistical

Table 1.1. *Book Outline*

Part and Chapter	Background[a]	Example
1. Preliminaries		
1. Overview	–	
2. Causal and Noncausal Models	–	Simultaneous equations models
3. Microeconomic Data Structures	–	Observational data
2. Core Methods		
4. Linear Models	–	Ordinary least squares
5. Maximum Likelihood and Nonlinear Least-Squares Estimation	–	m-estimation or extremum estimation
6. Generalized Method of Moments and Systems Estimation	5	Instrumental variables
7. Hypothesis Tests	5	Wald, score, and likelihood ratio tests
8. Specification Tests and Model Selection	5,7	Conditional moment test
9. Semiparametric Methods	–	Kernel regression
10. Numerical Optimization	5	Newton–Raphson iterative method
3. Simulation-Based Methods		
11. Bootstrap Methods	7	Percentile t-method
12. Simulation-Based Methods	5	Maximum simulated likelihood
13. Bayesian Methods	–	Markov chain Monte Carlo
4. Models for Cross-Section Data		
14. Binary Outcome Models	5	Logit, probit for $y = (0, 1)$
15. Multinomial Models	5,14	Multinomial logit for $y = (1, \ldots, m)$
16. Tobit and Selection Models	5,14	Tobit for $y = \max(y^*, 0)$
17. Transition Data: Survival Analysis	5	Cox proportional hazards for $y = \min(y^*, c)$
18. Mixture Models and Unobserved Heterogeneity	5,17	Unobserved heterogeneity
19. Models for Multiple Hazards	5,17	Multiple hazards
20. Models of Count Data	5	Poisson for $y = 0, 1, 2, \ldots$
5. Models for Panel Data		
21. Linear Panel Models: Basics	–	Fixed and random effects
22. Linear Panel Models: Extensions	6,21	Dynamic and endogenous regressors
23. Nonlinear Panel Models	5,6,21,22	Panel logit, Tobit, and Poisson
6. Further Topics		
24. Stratified and Clustered Samples	5	Data $(y_{ij}, \mathbf{x}_{ij})$ correlated over j
25. Treatment Evaluation	5,21	Regressor $d = 1$ if in program
26. Measurement Error Models	5	Logit model with measurement errors
27. Missing Data and Imputation	5	Regression with missing observations

[a] The background gives the essential chapter needed in addition to the treatment of ordinary and weighted LS in Chapter 4. Note that the first panel data chapter (Chapter 21) requires only Chapter 4.

arena of regression analysis. Many of the issues raised in these chapters are pursued throughout the book as the reader develops the necessary tools.

1.3.2. Part 2: Core Methods

Chapters 4–10 detail the main general methods used in classical estimation and statistical inference. The results given in Chapter 5 in particular are extensively used throughout the book.

Chapter 4 presents some results for the **linear regression model**, emphasizing those issues and methods that are most relevant for the rest of the book. Analysis is relatively straightforward as there is an explicit expression for linear model estimators such as ordinary least squares.

Chapters 5 and 6 present **estimation theory** that can be applied to nonlinear models for which there is usually no explicit solution for the estimator. Asymptotic theory is used to obtain the distribution of estimators, with emphasis on obtaining robust standard error estimates that rely on relatively weak distributional assumptions. A quite general treatment of estimation, along with specialization to nonlinear least-squares and maximum likelihood estimation, is presented in Chapter 5. The more challenging generalized method of moments estimator and specialization to instrumental variables estimation are given separate treatment in Chapter 6.

Chapter 7 presents **classical hypothesis testing** when estimators are nonlinear and the hypothesis being tested is possibly nonlinear in parameters. **Specification tests** in addition to hypothesis tests are the subject of Chapter 8.

Chapter 9 presents **semiparametric estimation** methods such as kernel regression. The leading example is flexible modeling of the conditional mean. For the patents example, the nonparametric regression model is $\mathrm{E}[y|x] = g(x)$, where the function $g(\cdot)$ is unspecified and is instead estimated. Then estimation has an infinite-dimensional component $g(\cdot)$ leading to a nonstandard asymptotic theory. With additional regressors some further structure is needed and the methods are called semiparametric or seminonparametric.

Chapter 10 presents the **computational methods** used to compute a parameter estimate when the estimator is defined implicitly, usually as the solution to some first-order conditions.

1.3.3. Part 3: Simulation-Based Methods

Chapters 11–13 consider methods of estimation and inference that rely on simulation. These methods are generally more computationally intensive and, currently, less utilized than the methods presented in Part 2.

Chapter 11 presents the **bootstrap method** for statistical inference. This yields the empirical distribution of an estimator by obtaining new samples by simulation, such as by repeated resampling with replacement from the original sample. The bootstrap can provide a simple way to obtain standard errors when the formulas from asymptotic theory are complex, as is the case for some two-step estimators. Furthermore, if implemented appropriately, the bootstrap can lead to better statistical inference in small samples.

Chapter 12 presents **simulation-based estimation methods**, developed for models that involve an integral over a probability distribution for which there is no closed-form solution. Estimation is still possible by making multiple draws from the relevant distribution and averaging.

Chapter 13 presents **Bayesian methods**, which combine a distribution for the observed data with a specified prior distribution for parameters to obtain a posterior distribution of the parameters that is the basis for estimation. Recent advances make computation possible even if there is no closed-form solution for the posterior distribution. Bayesian analysis can provide an approach to estimation and inference that is quite different from the classical approach. However, in many cases only the Bayesian tool kit is adopted to permit classical estimation and inference for problems that are otherwise intractable.

1.3.4. Part 4: Models for Cross-Section Data

Chapters 14–20 present the main nonlinear models for **cross-section data**. This part is the heart of the book and presents advanced topics such as models for limited dependent variables and sample selection. The classes of models are defined by the range of values taken by the dependent variable.

Binary data models for dependent variable that can take only two possible values, say $y = 0$ or $y = 1$, are presented in Chapter 14. In Chapter 15 an extension is made to **multinomial** models, for dependent variable that takes several discrete values. Examples include employment status (employed, unemployed, and out of the labor force) and mode of transportation to work (car, bus, or train). Linear models can be informative but are not appropriate, as they can lead to predicted probabilities outside the unit interval. Instead logit, probit, and related models are used.

Chapter 16 presents models with **censoring, truncation, sample selection**. Examples include annual hours of work, conditional on choosing to work, and hospital expenditures, conditional on being hospitalized. In these cases the data are incompletely observed with a bunching of observations at $y = 0$ and with the remaining $y > 0$. The model for the observed data can be shown to be nonlinear even if the underlying process is linear, and linear regression on the observed data can be very misleading. Simple corrections for censoring, truncation, or sample selection such as the Tobit model exist, but these are very dependent on distributional assumptions.

Models for **duration data** are presented in Chapters 17–19. An example is length of unemployment spell. Standard regression models include the exponential, Weibull, and Cox proportional hazards model. Additionally, as in Chapter 16, the dependent variable is often incompletely observed. For example, the data may be on the length of a current spell that is incomplete, rather than the length of a completed spell.

Chapter 20 presents **count data** models. Examples include various measures of health utilization such as number of doctor visits and number of days hospitalized. Again the model is nonlinear, as counts and hence the conditional mean are nonnegative. Leading parametric models include the Poisson and negative binomial.

1.3.5. Part 5: Models for Panel Data

Chapters 21–23 present methods for **panel data**. Here the data are observed in several time periods for each of the many individuals in the sample, so the dependent variable and regressors are indexed by both individual and time. Any analysis needs to control for the likely positive correlation of error terms in different time periods for a given individual. Additionally, panel data can provide sufficient data to control for unobserved time-invariant individual-specific effects, permitting identification of causation under weaker assumptions than those needed if only cross-section data are available.

The basic linear panel data model is presented in Chapter 21, with emphasis on **fixed effects** and **random effects** models. Extensions of linear models to permit lagged dependent variables and endogenous regressors are presented in Chapter 22. Panel methods for the nonlinear models of Part 4 are presented in Chapter 23.

The panel data methods are placed late in the book to permit a unified self-contained treatment. Chapter 21 could have been placed immediately after Chapter 4 and is written in an accessible manner that relies on little more than knowledge of least-squares estimation.

1.3.6. Part 6: Further Topics

This part considers important topics that can generally relate to any and all models covered in Parts 4 and 5. Chapter 24 deals with modeling of clustered data in several different models. Chapter 25 discusses treatment evaluation. Treatment evaluation is a general term that can cover a wide variety of models in which the focus is on measuring the impact of some "treatment" that is either exogenously or randomly assigned to an individual on some measure of interest, denoted an "outcome variable." Chapter 26 deals with the consequences of measurement errors in outcome and/or regressor variables, with emphasis on some leading nonlinear models. Chapter 27 considers some methods of handling missing data in linear and nonlinear regression models.

1.4. How to Use This Book

The book assumes a basic understanding of the linear regression model with matrix algebra. It is written at the mathematical level of the first-year economics Ph.D. sequence, comparable to Greene (2003).

Although some of the material in this book is covered in a first-year sequence, most of it appears in second-year econometrics Ph.D. courses or in data-oriented microeconomics field courses such as labor economics, public economics, or industrial organization. This book is intended to be used as both an econometrics text and as an adjunct for such field courses. More generally, the book is intended to be useful as a reference work for applied researchers in economics, in related social sciences such as sociology and political science, and in epidemiology.

For readers using this book as a reference work, the models chapters have been written to be as self-contained as possible. For the specific models presented in Parts 4

Table 1.2. *Outline of a 20-Lecture 10-Week Course*

Lectures	Chapter	Topic
1–3	4, Appx. A	Review of linear models and asymptotic theory
4–7	5	Estimation: m-estimation, ML, and NLS
8	10	Estimation: numerical optimization
9–11	14, 15	Models: binary and multinomial
12–14	16	Models: censored and truncated
15	6	Estimation: GMM
16	7	Testing: hypothesis tests
17–19	21	Models: basic linear panel
20	9	Estimation: semiparametric

and 5 it will generally be sufficient to read the relevant chapter in isolation, except that some command of the general estimation results in Chapter 5 and in some cases Chapter 6 will be necessary. Most chapters are structured to begin with a discussion and example that is accessible to a wide audience.

For instructors using this book as a course text it is best to introduce the basic nonlinear cross-section and linear panel data models as early as possible, skipping many of the methods chapters. The most commonly used nonlinear cross-section models are presented in Chapters 14–16; these require knowledge of maximum likelihood and least-squares estimation, presented in Chapter 5. Chapter 21 on linear panel data models requires even less preparation, essentially just Chapter 4.

Table 1.2 provides an outline for a one-quarter second-year graduate course taught at the University of California, Davis, immediately following the required first-year statistics and econometrics sequence. A quarter provides sufficient time to cover the basic results given in the first half of the chapters in this outline. With additional time one can go into further detail or cover a subset of Chapters 11–13 on computationally intensive estimation methods (simulation-based estimation, the bootstrap, which is also briefly presented in Chapter 7, and Bayesian methods); additional cross-section models (durations and counts) presented in Chapters 17–20; and additional panel data models (linear model extensions and nonlinear models) given in Chapters 22 and 23.

At Indiana University, Bloomington, a 15-week semester-long field course in microeconometrics is based on material in most of Parts 4 and 5. The prerequisite courses for this course cover material similar to that in Part 2.

Some exercises are provided at the end of each chapter after the first three introductory chapters. These exercises are usually learning-by-doing exercises; some are purely methodological whereas others entail analysis of generated or actual data. The level of difficulty of the questions is mostly related to the level of difficulty of the topic.

1.5. Software

There are many software packages available for data analysis. Popular packages with strong microeconometric capabilities include LIMDEP, SAS, and STATA, all of which offer an impressive range of canned routines and additionally support user-defined procedures using a matrix programming language. Other packages that are also widely used include EVIEWS, PCGIVE, and TSP. Despite their time-series orientation, these can support some cross-section data analysis. Users who wish to do their own programming also have available a variety of options including GAUSS, MATLAB, OX, and SAS/IML. The latest detailed information about these packages and many others can be efficiently located via an Internet browser and a search engine.

1.6. Notation and Conventions

Vector and matrix algebra are used extensively.

Vectors are defined as column vectors and represented using lowercase bold. For example, for linear regression the regressor vector $\mathbf{x}$ is a $K \times 1$ column vector with jth entry x_j and the parameter vector $\boldsymbol{\beta}$ is a $K \times 1$ column vector with jth entry β_j, so

$$\underset{(K \times 1)}{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix} \quad \text{and} \quad \underset{(K \times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}.$$

Then the linear regression model $y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$ is expressed as $y = \mathbf{x}'\boldsymbol{\beta} + u$. At times a subscript i is added to denote the typical ith observation. The linear regression equation for the ith observation is then

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i.$$

The sample is one of N observations, $\{(y_i, \mathbf{x}_i), i = 1, \ldots, N\}$. In this book observations are usually assumed to be independent over i.

Matrices are represented using uppercase bold. In matrix notation the sample is $(\mathbf{y}, \mathbf{X})$, where $\mathbf{y}$ is an $N \times 1$ vector with ith entry y_i and $\mathbf{X}$ is a matrix with ith row $\mathbf{x}_i'$, so

$$\underset{(N \times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \quad \text{and} \quad \underset{(N \times \dim(\mathbf{x}))}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix}.$$

The linear regression model upon stacking all N observations is then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{u}$ is an $N \times 1$ column vector with ith entry u_i.

Matrix notation is compact but at times it is clearer to write products of matrices as summations of products of vectors. For example, the OLS estimator can be equivalently written in either of the following ways:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i.$$

Table 1.3. *Commonly Used Acronyms and Abbreviations*

Linear	OLS	Ordinary least squares
	GLS	Generalized least squares
	FGLS	Feasible generalized least squares
	IV	Instrumental variables
	2SLS	Two-stage least squares
	3SLS	Three-stage least squares
Nonlinear	NLS	Nonlinear least squares
	FGNLS	Feasible generalized nonlinear least squares
	NIV	Nonlinear instrumental variables
	NL2SLS	Nonlinear two-stage least squares
	NL3SLS	Nonlinear three-stage least squares
General	LS	Least squares
	ML	Maximum likelihood
	QML	Quasi-maximum likelihood
	GMM	Generalized method of moments
	GEE	Generalized estimating equations

Generic notation for a parameter is the $q \times 1$ vector $\boldsymbol{\theta}$. The regression parameters are represented by the $K \times 1$ vector $\boldsymbol{\beta}$, which may equal $\boldsymbol{\theta}$ or may be a subset of $\boldsymbol{\theta}$ depending on the context.

The book uses many abbreviations and acronyms. Table 1.3 summarizes abbreviations used for some common estimation methods, ordered by whether the estimator is developed for linear or nonlinear regression models. We also use the following: dgp (data-generating process), iid (independently and identically distributed), pdf (probability density function), cdf (cumulative distribution function), L (likelihood), $\ln L$ (log-likelihood), FE (fixed effects), and RE (random effects).

CHAPTER 2

Causal and Noncausal Models

2.1. Introduction

Microeconometrics deals with the theory and applications of methods of data analysis developed for microdata pertaining to individuals, households, and firms. A broader definition might also include regional- and state-level data. Microdata are usually either cross sectional, in which case they refer to conditions at the same point in time, or longitudinal (panel) in which case they refer to the same observational units over several periods. Such observations are generated from both nonexperimental setups, such as censuses and surveys, and quasi-experimental or experimental setups, such as social experiments implemented by governments with the participation of volunteers.

A microeconometric model may be a full specification of the probability distribution of a set of microeconomic observations; it may also be a partial specification of some distributional properties, such as moments, of a subset of variables. The mean of a single dependent variable conditional on regressors is of particular interest.

There are several objectives of microeconometrics. They include both data description and causal inference. The first can be defined broadly to include moment properties of response variables, or regression equations that highlight associations rather than causal relations. The second category includes causal relationships that aim at measurement and/or empirical confirmation or refutation of conjectures and propositions regarding microeconomic behavior. The type and style of empirical investigations therefore span a wide spectrum. At one end of the spectrum can be found very highly structured models, derived from detailed specification of the underlying economic behavior, that analyze **causal** (behavioral) or **structural relationships** for interdependent microeconomic variables. At the other end are **reduced form** studies that aim to uncover correlations and associations among variables, without necessarily relying on a detailed specification of all relevant interdependencies. Both approaches share the common goal of uncovering important and striking relationships that could be helpful in understanding microeconomic behavior, but they differ in the extent to which they rely on economic theory to guide their empirical investigations.

As a subdiscipline microeconometrics is newer than macroeconometrics, which is concerned with modeling of market and aggregate data. A great deal of the early work in applied econometrics was based on aggregate time-series data collected by government agencies. Much of the early work on statistical demand analysis up until about 1940 used market rather than individual or household data (Hendry and Morgan, 1996). Morgan's (1990) book on the history of econometric ideas makes no reference to microeconometric work before the 1940s, with one important exception. That exception is the work on household budget data that was instigated by concern with the living standards of the less well-off in many countries. This led to the collection of household budget data that provided the raw material for some of the earlier microeconometric studies such as those pioneered by Allen and Bowley (1935). Nevertheless, it is only since the 1950s that microeconometrics has emerged as a distinctive and recognized subdiscipline. Even into the 1960s the core of microeconometrics consisted of demand analyses based on household surveys.

With the award of the year 2000 Nobel Prize in Economics to James Heckman and Daniel McFadden for their contributions to microeconometrics, the subject area has achieved clear recognition as a distinct subdiscipline. The award cited Heckman "for his development of theory and methods for analyzing selective samples" and McFadden "for his development of theory and methods for analyzing discrete choice." Examples of the type of topics that microeconometrics deals with were also mentioned in the citation: "... what factors determine whether an individual decides to work and, if so, how many hours? How do economic incentives affect individual choices regarding education, occupation or place of residence? What are the effects of different labor-market and educational programs on an individual's income and employment?"

Applications of microeconometric methods can be found not only in every area of microeconomics but also in other cognate social sciences such as political science, sociology, and geography.

Beginning with the 1970s and especially within the past two decades revolutionary advances in our capacity for handling large data sets and associated computations have taken place. These, together with the accompanying explosion in the availability of large microeconomic data sets, have greatly expanded the scope of microeconometrics. As a result, although empirical demand analysis continues to be one of the most important areas of application for microeconometric methods, its style and content have been heavily influenced by newer methods and models. Further, applications in economic development, finance, health, industrial organization, labor and public economics, and applied microeconomics generally are now commonplace, and these applications will be encountered at various places in this book.

The primary focus of this book is on the newer material that has emerged in the past three decades. Our goal is to survey concepts, models, and methods that we regard as standard components of a modern microeconometrician's tool kit. Of course, the notion of standard methods and models is inevitably both subjective and elastic, being a function of the presumed clientele of this book as well as the authors' own backgrounds. There may also be topics we regard as too advanced for an introductory book such as this that others would place in a different category.

Microeconometrics focuses on the complications of nonlinear models and on obtaining estimates that can be given a structural interpretation. Much of this book, especially Parts 2–4, presents methods for nonlinear models. These nonlinear methods overlap with many areas of applied statistics including biostatistics. By contrast, the distinguishing feature of econometrics is the emphasis placed on causal modeling. This chapter introduces the key concepts related to causal (and noncausal) modeling, concepts that are germane to both linear and nonlinear models.

Sections 2.2 and 2.3 introduce the key concepts of structure and exogeneity. Section 2.4 uses the linear simultaneous equations model as a specific illustration of a structural model and connects it with the other important concepts of reduced form models. Identification definitions are given in Section 2.5. Section 2.6 considers single-equation structural models. Section 2.7 introduces the potential outcome model and compares the causal parameters and interpretations in the potential outcome model with those in the simultaneous equations model. Section 2.8 provides a brief discussion of modeling and estimation strategies designed to handle computational and data challenges.

2.2. Structural Models

Structure consists of

1. a set of variables $\mathbf{W}$ ("data") partitioned for convenience as $[\mathbf{Y}\ \mathbf{Z}]$;
2. a joint probability distribution of $\mathbf{W}$, $F(\mathbf{W})$;
3. an a priori ordering of $\mathbf{W}$ according to hypothetical cause-and-effect relationships and specification of a priori restrictions on the hypothesized model; and
4. a parametric, semiparametric, or nonparametric specification of functional forms and the restrictions on the parameters of the model.

This general description of a structural model is consistent with a well-established Cowles Commission definition of a structure. For example, Sargan (1988, p. 27) states:

> A model is the specification of the probability distribution for a set of observations. A structure is the specification of the parameters of that distribution. Therefore, a structure is a model in which all the parameters are assigned numerical values.

We consider the case in which the modeling objective is to explain the values of observable vector-valued variable $\mathbf{y}$, $\mathbf{y}' = (y_1, \ldots, y_G)$. Each element of $\mathbf{y}$ is a function of some other elements of $\mathbf{y}$ and of explanatory variables $\mathbf{z}$ and a purely random disturbance u. Note that the variables $\mathbf{y}$ are assumed to be interdependent. By contrast, interdependence between $\mathbf{z}_i$ is not modeled. The ith observation satisfies the set of implicit equations

$$\mathbf{g}\left(\mathbf{y}_i, \mathbf{z}_i, \mathbf{u}_i | \boldsymbol{\theta}\right) = \mathbf{0}, \tag{2.1}$$

where $\mathbf{g}$ is a known function. We refer to this as the **structural model**, and we refer to $\boldsymbol{\theta}$ as structural parameters. This corresponds to property 4 given earlier in this section.

Assume that there is a unique solution for $\mathbf{y}_i$ for every $(\mathbf{z}_i, \mathbf{u}_i)$. Then we can write the equations in an explicit form with $\mathbf{y}$ as function of $(\mathbf{z}, \mathbf{u})$:

$$\mathbf{y}_i = \mathbf{f}(\mathbf{z}_i, \mathbf{u}_i | \boldsymbol{\pi}). \tag{2.2}$$

This is referred to as the **reduced form** of the structural model, where $\boldsymbol{\pi}$ is a vector of reduced form parameters that are functions of $\boldsymbol{\theta}$. The reduced form is obtained by solving the structural model for the endogenous variables $\mathbf{y}_i$, given $(\mathbf{z}_i, \mathbf{u}_i)$. The reduced form parameters $\boldsymbol{\pi}$ are functions of $\boldsymbol{\theta}$.

If the objective of modeling is inference about elements of $\boldsymbol{\theta}$, then (2.1) provides a direct route. This involves estimation of the structural model. However, because elements of $\boldsymbol{\pi}$ are functions of $\boldsymbol{\theta}$, (2.2) also provides an indirect route to inference on $\boldsymbol{\theta}$. If $\mathbf{f}(\mathbf{z}_i, \mathbf{u}_i | \boldsymbol{\pi})$ has a known functional form, and if it is additively separable in $\mathbf{z}_i$ and $\mathbf{u}_i$, such that we can write

$$\mathbf{y}_i = \mathbf{g}(\mathbf{z}_i | \boldsymbol{\pi}) + \mathbf{u}_i = \mathrm{E}[\mathbf{y}_i | \mathbf{z}_i] + \mathbf{u}_i, \tag{2.3}$$

then the regression of $\mathbf{y}$ on $\mathbf{z}$ is a natural prediction function for $\mathbf{y}$ given $\mathbf{z}$. In this sense the reduced form equation has a useful role for making conditional predictions of $\mathbf{y}_i$ given $(\mathbf{z}_i, \mathbf{u}_i)$. To generate predictions of the left-hand-side variable for assigned values of the right-hand-side variables of (2.2) requires estimates of $\boldsymbol{\pi}$, which may be computationally simpler.

An important extension of (2.3) is the **transformation model**, which for scalar y takes the form

$$\Lambda(y) = \mathbf{z}'\boldsymbol{\pi} + \mathbf{u}, \tag{2.4}$$

where $\Lambda(y)$ is a transformation function (e.g., $\Lambda(y) = \ln(y)$ or $\Lambda(y) = y^{1/2}$). In some cases the transformation function may depend on unknown parameters. A transformation model is distinct from a regression, but it too can be used to make estimates of $\mathrm{E}[y|\mathbf{z}]$. An important example is the accelerated failure time model analyzed in Chapter 17.

One of the most important, and potentially controversial, steps in the specification of the structural model is property 3, in which an a priori ordering of variables into causes and effects is assigned. In essence this involves drawing a distinction between those variables whose variation the model is designed to explain and those whose variation is externally determined and hence lie outside the scope of investigation. In microeconometrics, examples of the former are years of schooling and hours worked; examples of the latter are gender, ethnicity, age, and similar demographic variables. The former, denoted $\mathbf{y}$, are referred to as **endogenous** and the latter, denoted $\mathbf{z}$, are called **exogenous** variables.

Exogeneity of a variable is an important simplification because in essence it justifies the decision to treat that variable as ancillary, and not to model that variable because the parameters of that relationship have no direct bearing on the variable under study. This important notion needs a more formal definition, which we now provide.

2.3. Exogeneity

We begin by considering the representation of a general finite dimensional parametric case in which the joint distribution of $\mathbf{W}$, with parameters $\boldsymbol{\theta}$ partitioned as $(\boldsymbol{\theta}_1\ \boldsymbol{\theta}_2)$, is factored into the conditional density of $\mathbf{Y}$ given $\mathbf{Z}$, and the marginal distribution of $\mathbf{Z}$, giving

$$f_J(\mathbf{W}|\boldsymbol{\theta}) = f_C(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}) \times f_M(\mathbf{Z}|\boldsymbol{\theta}). \tag{2.5}$$

A special case of this result occurs if

$$f_J(\mathbf{W}|\boldsymbol{\theta}) = f_C(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta}_1) \times f_M\left(\mathbf{Z}|\boldsymbol{\theta}_2\right),$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are functionally independent. Then we say that $\mathbf{Z}$ is exogenous with respect to $\boldsymbol{\theta}_1$; this means that knowledge of $f_M(\mathbf{Z}|\boldsymbol{\theta}_2)$ is not required for inference on $\boldsymbol{\theta}_1$, and hence we can validly condition the distribution of $\mathbf{Y}$ on $\mathbf{Z}$.

Models can always be reparameterized. So next consider the case in which the model is reparameterized in terms of parameters φ, with one-to-one transformation of $\boldsymbol{\theta}$, say $\varphi = h(\boldsymbol{\theta})$, where φ is partitioned into (φ_1, φ_2). This reparametrization may be of interest if, for example, φ_1 is structurally invariant to a class of policy interventions. Suppose φ_1 is the parameter of interest. In such a case one is interested in the exogeneity of $\mathbf{Z}$ with respect to φ_1. Then, the condition for exogeneity is that

$$f_J(\mathbf{W}|\varphi) = f_C\left(\mathbf{Y}|\mathbf{Z}, \varphi_1\right) \times f_M\left(\mathbf{Z}|\varphi_2\right), \tag{2.6}$$

where φ_1 is independent of φ_2.

Finally consider the case in which the interest is in a parameter λ that is a function of φ, say $h(\varphi)$. Then for exogeneity of $\mathbf{Z}$ with respect to λ, we need two conditions: (i) λ depends only on φ_1, i.e., $\lambda = h(\varphi_1)$, and hence only the conditional distribution is of interest; and (ii) φ_1 and φ_2 are "variation free" which means that the parameters of the joint distribution are not subject to cross-restrictions, i.e. $(\varphi_1, \varphi_2) \in \boldsymbol{\Phi}_1 \times \boldsymbol{\Phi}_2 = \{\varphi_1 \in \Phi_1, \varphi_2 \in \Phi_2\}$.

The factorization in (2.5)-(2.6) plays an important role in the development of the exogeneity concept. Of special interest in this book are the following three concepts related to exogeneity: (1) weak exogeneity; (2) Granger noncausality; (3) strong exogeneity.

Definition 2.1 (*Weak Exogeneity*): Z is **weakly exogenous for** λ if (i) and (ii) hold.

If the marginal model parameters are uninformative for inference on λ, then inference on λ can proceed on the basis of the conditional distribution $f(\mathbf{Y}|\mathbf{Z}, \varphi_1)$ alone. The operational implication is that weakly exogenous variables can be taken as given if one's main interest is in inference on λ or φ_1. This does not mean that there is no statistical model for $\mathbf{Z}$; it means that the parameters of that model play no role in the inference on φ_1, and hence are irrelevant.

2.3.1. Conditional Independence

Originally, the Granger causality concept was defined in the context of prediction in a time-series environment. More generally, it can be interpreted as a form of **conditional independence** (Holland, 1986, p. 957).

Partition $\mathbf{z}$ into two subsets $\mathbf{z}_1$ and $\mathbf{z}_2$; let $\mathbf{W} = [\mathbf{y},\ \mathbf{z}_1,\ \mathbf{z}_2]$ be the matrices of variables of interest. Then $\mathbf{z}_1$ and $\mathbf{y}$ are conditionally independent given $\mathbf{z}_2$ if

$$f(\mathbf{y}|\mathbf{z}_1, \mathbf{z}_2) = f(\mathbf{y}|\mathbf{z}_2). \tag{2.8}$$

This is stronger than the **mean independence** assumption, which would imply

$$\mathrm{E}[\mathbf{y}|\mathbf{z}_1, \mathbf{z}_2] = \mathrm{E}[\mathbf{y}|\mathbf{z}_2]. \tag{2.9}$$

Then $\mathbf{z}_1$ has no predictive value for $\mathbf{y}$, after conditioning on $\mathbf{z}_2$. In a predictive sense this means that $\mathbf{z}_1$ does not Granger-cause $\mathbf{y}$.

In a time-series context, $\mathbf{z}_1$ and $\mathbf{z}_2$ would be mutually exclusive lagged values of subsets of $\mathbf{y}$.

Definition 2.2 (*Strong Exogeneity*): $\mathbf{z}_1$ is **strongly exogenous** for φ if it is weakly exogenous for φ and does not Granger-cause $\mathbf{y}$ so (2.8) holds.

2.3.2. Exogenizing Variables

Exogeneity is a strong assumption. It is a property of random variables relative to parameters of interest. Hence a variable may be validly treated as exogenous in one structural model but not in another; the key issue is the parameters that are the subject of inference. Arbitrary imposition of this property will have some undesirable consequences that will be discussed in Section 2.4.

The exogeneity assumption may be justified by a priori theorizing, in which case it is a part of the maintained hypothesis of the model. It may in some cases be justified as a valid approximation, in which case it may be subject to testing, as discussed in Section 8.4.3. In cross-section analysis it may be justified as being a consequence of a natural experiment or a quasi-experiment in which the value of the variable is determined by an external intervention; for example, government or regulatory authority may determine the setting of a tax rate or a policy parameter. Of special interest is the case in which an external intervention results in a change in the value of an important policy variable. Such a natural experiment is tantamount to exogenization of some variable. As we shall see in Chapter 3, this creates a quasi-experimental opportunity to study the impact of a variable in the absence of other complicating factors.

2.4. Linear Simultaneous Equations Model

An important special case of the general structural model specified in (2.1) is the linear simultaneous equation model developed by the Cowles Commission econometricians. Comprehensive treatment of this model is available in many textbooks (e.g., Sargan, 1988). The treatment here is brief and selective; also see Section 6.9.6. The objective is to bring into the discussion several key ideas and concepts that have more general relevance. Although the analysis is restricted to linear models, many insights are routinely applied to nonlinear models.

2.4.1. The SEM Setup

The **linear simultaneous equations model** (SEM) setup is as follows:

$$\begin{aligned} y_{1i}\beta_{11} + \cdots + y_{Gi}\beta_{1G} + z_{1i}\gamma_{11} + \cdots + z_{Ki}\gamma_{1K} &= u_{1i} \\ \vdots \qquad\qquad\qquad\qquad\qquad \vdots \quad &= \vdots \\ y_{1i}\beta_{G1} + \cdots + y_{Gi}\beta_{GG} + z_{1i}\gamma_{G1} + \cdots + z_{Ki}\gamma_{GK} &= u_{Gi}, \end{aligned}$$

where i is the observation subscript.

A clear a priori distinction or preordering is made between endogenous variables, $\mathbf{y}_i' = (y_{1i}, \ldots, y_{Gi})$, and exogenous variables, $\mathbf{z}_i' = (z_{1i}, \ldots, z_{Ki})$. By definition the exogenous variables are uncorrelated with the purely random disturbances $(u_{1i}, \ldots, u_{Gi})$. In its unrestricted form every variable enters every equation.

In matrix notation, the G-equation SEM for the ith equation is written as

$$\mathbf{y}_i'\mathbf{B} + \mathbf{z}_i'\mathbf{\Gamma} = \mathbf{u}_i', \tag{2.10}$$

where $\mathbf{y}_i$, $\mathbf{B}$, $\mathbf{z}_i$, $\mathbf{\Gamma}$, and $\mathbf{u}_i$ have dimensions $G \times 1$, $G \times G$, $K \times 1$, $K \times G$, and $G \times 1$, respectively. For specified values of $(\mathbf{B}, \mathbf{\Gamma})$ and $(\mathbf{z}_i, \mathbf{u}_i)$ G linear simultaneous equations can in principle be solved for $\mathbf{y}_i$.

The standard assumptions of SEM are as follows:

1. $\mathbf{B}$ is nonsingular and has rank G.
2. $\text{rank}[\mathbf{Z}] = K$. The $N \times K$ matrix $\mathbf{Z}$ is formed by stacking $\mathbf{z}_i'$, $i = 1, \ldots, N$.
3. $\text{plim}\, N^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{\Sigma_{zz}}$ is a symmetric $K \times K$ positive definite matrix.
4. $\mathbf{u}_i \sim \mathcal{N}[\mathbf{0}, \mathbf{\Sigma}]$; that is, $\mathrm{E}[\mathbf{u}_i] = \mathbf{0}$ and $\mathrm{E}[\mathbf{u}_i\mathbf{u}_i'] = \Sigma = [\sigma_{ij}]$, where $\mathbf{\Sigma}$ is a symmetric $G \times G$ positive definite matrix.
5. The errors in each equation are serially independent.

In this model the structure (or structural parameters) consists of $(\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Sigma})$. Writing

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1' \\ \cdot \\ \cdot \\ \mathbf{y}_N' \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{z}_1' \\ \cdot \\ \cdot \\ \mathbf{z}_N' \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \mathbf{u}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_N' \end{bmatrix}$$

allows us to express the **structural model** more compactly as

$$\mathbf{YB} + \mathbf{Z\Gamma} = \mathbf{U}, \tag{2.11}$$

where the arrays $\mathbf{Y}, \mathbf{B}, \mathbf{Z}, \mathbf{\Gamma}$, and $\mathbf{U}$ have dimensions $N \times G$, $G \times G$, $N \times K$, $K \times G$, and $N \times G$, respectively. Solving for all the endogenous variables in terms of all the exogenous variables, we obtain the **reduced form of the SEM**:

$$\mathbf{Y} + \mathbf{Z}\boldsymbol{\Gamma}\mathbf{B}^{-1} = \mathbf{U}\mathbf{B}^{-1},$$
$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V}, \tag{2.12}$$

where $\boldsymbol{\Pi} = -\boldsymbol{\Gamma}\mathbf{B}^{-1}$ and $\mathbf{V} = \mathbf{U}\mathbf{B}^{-1}$. Given Assumption 4, $\mathbf{v}_i \sim \mathcal{N}[\mathbf{0}, \mathbf{B}^{-1\prime}\boldsymbol{\Sigma}\mathbf{B}^{-1}]$.

In the SEM framework the structural model has primacy for several reasons. First, the equations themselves have interpretations as economic relationships such as demand or supply relations, production functions, and so forth, and they are subject to restrictions of economic theory. Consequently, $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are parameters that describe economic behavior. Hence a priori theory can be invoked to form expectations about the sign and size of individual coefficients. By contrast, the unrestricted reduced form parameters are potentially complicated functions of the structural parameters, and as such it may be difficult to evaluate them postestimation. This consideration may have little weight if the goal of econometric modeling is prediction rather than inference on parameters with behavioral interpretation.

Consider, without loss of generality, the first equation in the model (2.11), with y_1 as the dependent variable. In addition, some of the remaining $G-1$ endogenous variables and $K-1$ exogenous variables may be absent from this equation. From (2.12) we see that in general the endogenous variables $\mathbf{Y}$ depend stochastically on $\mathbf{V}$, which in turn is a function of the structural errors $\mathbf{U}$. Therefore, in general plim $N^{-1}\mathbf{Y}'\mathbf{U} \neq \mathbf{0}$. Generally, the application of the least-squares estimator in the simultaneous equation setting yields inconsistent estimates. This is a well-known and basic result from the simultaneous equations literature, often referred to as the "simultaneous equations bias" problem. The vast literature on simultaneous equations models deals with identification and consistent estimation when the least-squares approach fails; see Sargan (1988) and Schmidt (1976), and Section 6.9.6.

The reduced form of SEM expresses every endogenous variable as a linear function of all exogenous variables and all structural disturbances. The reduced form disturbances are linear combinations of the structural disturbances. From the reduced form for the ith observation

$$\mathrm{E}[\mathbf{y}_i|\mathbf{z}_i] = \mathbf{z}_i'\boldsymbol{\Pi}, \tag{2.13}$$
$$\mathrm{V}[\mathbf{y}_i|\mathbf{z}_i] = \boldsymbol{\Omega} \equiv \mathbf{B}^{-1\prime}\boldsymbol{\Sigma}\mathbf{B}^{-1}. \tag{2.14}$$

The reduced form parameters $\boldsymbol{\Pi}$ are derived parameters defined as functions of the structural parameters. If $\boldsymbol{\Pi}$ can be consistently estimated then the reduced form can be used to make predictive statements about variations in $\mathbf{Y}$ due to exogenous changes in $\mathbf{Z}$. This is possible even if $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are not known. Given the exogeneity of $\mathbf{Z}$, the full set of reduced form regressions is a multivariate regression model that can be estimated consistently by least squares. The reduced form provides a basis for making conditional predictions of $\mathbf{Y}$ given $\mathbf{Z}$.

The restricted reduced form is the unrestricted reduced form model subject to restrictions. If these are the same restrictions as those that apply to the structure, then structural information can be recovered from the reduced form.

In the SEM framework, the unknown structural parameters, the nonzero elements of $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Sigma}$, play a key role because they reflect the causal structure of the model. The interdependence between endogenous variables is described by $\mathbf{B}$, and the responses of endogenous variables to exogenous shocks in $\mathbf{Z}$ is reflected in the parameter matrix $\boldsymbol{\Gamma}$. In this setup the causal parameters of interest are those that measure the direct marginal impact of a change in an explanatory variable, y_j or z_k on the outcome of interest y_l, $l \neq j$, and functions of such parameters and data. The elements of $\boldsymbol{\Sigma}$ describe the dispersion and dependence properties of the random disturbances, and hence they measure some properties of the way the data are generated.

2.4.2. Causal Interpretation in SEM

A simple example will illustrate the causal interpretation of parameters in SEM. The structural model has two continuous endogenous variables y_1 and y_2, a single continuous exogenous variable z_1, one stochastic relationship linking y_1 and y_2, and one definitional identity linking all three variables in the model:

$$\begin{aligned} y_1 &= \gamma_1 + \beta_1 y_2 + u_1, \quad 0 < \beta_1 < 1, \\ y_2 &= y_1 + z_1. \end{aligned}$$

In this model u_1 is a stochastic disturbance, independent of z_1, with a well-defined distribution. The parameter β_1 is subject to an inequality constraint that is also a part of the model specification. The variable z_1 is exogenous and therefore its variation is induced by external sources that we may regard as interventions. These interventions have a direct impact on y_2 through the identity and also an indirect one through the first equation. The impact is measured by the reduced form of the model, which is

$$\begin{aligned} y_1 &= \frac{\gamma_1}{1-\beta_1} + \frac{\beta_1}{1-\beta_1} z_1 + \frac{1}{1-\beta_1} u_1 \\ &= \mathrm{E}[y_1 | z_1] + v_1, \\ y_2 &= \frac{\gamma_1}{1-\beta_1} + \frac{1}{1-\beta_1} z_1 + \frac{1}{1-\beta_1} u_1 \\ &= \mathrm{E}[y_2 | z_1] + v_1, \end{aligned}$$

where $v_1 = u_1/(1-\beta_1)$. The reduced form coefficients $\beta_1/(1-\beta_1)$ and $1/(1-\beta_1)$ have a causal interpretation. Any externally induced unit change in z_1 will cause the value of y_1 and y_2 to change by these amounts. Note that in this model y_1 and y_2 also respond to u_1. In order not to confound the impact of the two sources of variation we require that z_1 and u_1 are independent.

Also note that

$$\begin{aligned} \frac{\partial y_1}{\partial y_2} = \beta_1 &= \frac{\beta_1}{1-\beta_1} \div \frac{1}{1-\beta_1} \\ &= \frac{\partial y_1}{\partial z_1} \div \frac{\partial y_2}{\partial z_1}. \end{aligned}$$

In what sense does β_1 measure the causal effect of y_2 on y_1? To see a possible difficulty, observe that y_1 and y_2 are interdependent or jointly determined, so it is unclear in what sense y_2 "causes" y_1. Although z_1 (and u_1) is the ultimate cause of changes in the reduced form sense, y_2 is a proximate or an intermediate cause of y_1. That is, the first structural equation provides a snapshot of the impact of y_2 on y_1, whereas the reduced form gives the (equilibrium) impact after allowing for all interactions between the endogenous variables to work themselves out. In a SEM framework even endogenous variables are viewed as causal variables, and their coefficients as causal parameters. This approach can cause puzzlement for those who view causality in an experimental setting where independent sources of variation are the causal variables. The SEM approach makes sense if y_2 has an independent and exogenous source of variation, which in this model is z_1. Hence the marginal response coefficient β_1 is a function of how y_1 and y_2 respond to a change in z_1, as the immediately preceding equation makes clear.

Of course this model is but a special case. More generally, we may ask under what conditions will the SEM parameters have a meaningful causal interpretation. We return to this issue when discussing identification concepts in Section 2.5.

2.4.3. Extensions to Nonlinear and Latent Variable Models

If the simultaneous model is **nonlinear in parameters** only, the structural model can be written as

$$\mathbf{YB}(\boldsymbol{\theta}) + \mathbf{Z\Gamma}(\boldsymbol{\theta}) = \mathbf{U}, \tag{2.15}$$

where $\mathbf{B}(\boldsymbol{\theta})$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ are matrices whose elements are functions of the structural parameters $\boldsymbol{\theta}$. An explicit reduced form can be derived as before.

If **nonlinearity** is instead **in variables** then an explicit (analytical) reduced form may not be possible, although linearized approximations or numerical solutions of the dependent variables, given $(\mathbf{z}, \mathbf{u})$, can usually be obtained.

Many microeconometric models involve **latent** or **unobserved variables** as well as observed endogenous variables. For example, search and auction theory models use the concept of reservation wage or reservation price, choice models invoke indirect utility, and so forth. In the case of such models the structural model (2.1) may be replaced by

$$\mathbf{g}\left(\mathbf{y}_i^*, \mathbf{z}_i, \mathbf{u}_i | \boldsymbol{\theta}\right) = \mathbf{0}, \tag{2.16}$$

where the latent variables $\mathbf{y}_i^*$ replace the observed variables $\mathbf{y}_i$. The corresponding reduced form solves for $\mathbf{y}_i^*$ in terms of $(\mathbf{z}_i, \mathbf{u}_i)$, yielding

$$\mathbf{y}_i^* = \mathbf{f}\left(\mathbf{z}_i, \mathbf{u}_i | \boldsymbol{\pi}\right). \tag{2.17}$$

This reduced form has limited usefulness as $\mathbf{y}_i^*$ is not fully observed. However, if we have functions $\mathbf{y}_i = \mathbf{h}(\mathbf{y}_i^*)$ that relate observable with latent counterparts of $\mathbf{y}_i$, then the reduced form in terms of observables is

$$\mathbf{y}_i = \mathbf{h}\left(\mathbf{f}\left(\mathbf{z}_i, \mathbf{u}_i | \boldsymbol{\pi}\right)\right). \tag{2.18}$$

See Section 16.8.2 for further details.

When the structural model involves nonlinearities in variables, or when latent variables are involved, an explicit derivation of the functional form of this reduced form may be difficult to obtain. In such cases practitioners use approximations. By citing mathematical or computational convenience, a specific functional form may be used to relate an endogenous variable to all exogenous variables, and the result would be referred to as a "reduced form type relationship."

2.4.4. Interpretations of Structural Relationships

Marschak (1953, p. 26) in an influential essay gave the following definition of a structure:

> Structure was defined as a set of conditions which did not change while observations were being made but which might change in future. If a specified change of structure is expected or intended, prediction of variables of interest to the policy maker requires some knowledge of past structure. . . . In economics, the conditions that constitute a structure are (1) a set of relations describing human behavior and institutions as well as technological laws and involving, in general, nonobservable random disturbances and nonobservable random errors of measurement; (2) the joint probability distribution of these random quantities.

Marschak argued that the structure was fundamental for a quantitative evaluation or tests of economic theory and that the choice of the best policy requires knowledge of the structure.

In the SEM literature a structural model refers to "autonomous" (not "derived") relationships. There are other closely related concepts of a structure. One such concept refers to "deep parameters," by which is meant technology and preference parameters that are invariant to interventions.

In recent years an alternative usage of the term structure has emerged, one that refers to econometric models based on the hypothesis of dynamic stochastic optimization by rational agents. In this approach the starting point for any structural estimation problem is the first-order necessary conditions that define the agent's optimizing behavior. For example, in a standard problem of maximizing utility subject to constraints, the behavioral relations are the deterministic first-order marginal utility conditions. If the relevant functional forms are explicitly stated, and stochastic errors of optimization are introduced, then the first-order conditions define a behavioral model whose parameters characterize the utility function – the so-called deep or policy-invariant parameters. Examples are given in Sections 6.2.7 and 16.8.1.

Two features of this **highly structural approach** should be mentioned. First, they rely on a priori economic theory in a serious manner. Economic theory is not used simply to generate a list of relevant variables that one can use in a more or less arbitrarily specified functional form. Rather, the underlying economic theory has a major (but not exclusive) role in specification, estimation, and inference. The second feature is that identification, specification, and estimation of the resulting model can be very complicated, because the agent's optimization problem is potentially very complex, especially if dynamic optimization under uncertainty is postulated and discreteness and discontinuities are present; see Rust (1994).

2.5. Identification Concepts

The goal of the SEM approach is to consistently estimate $(\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Sigma})$ and conduct statistical inference. An important precondition for consistent estimation is that the model should be identified. We briefly discuss the important twin concepts of **observational equivalence** and **identifiability** in the context of parametric models.

Identification is concerned with determination of a parameter given sufficient observations. In this sense, it is an asymptotic concept. Statistical uncertainty necessarily affects any inference based on a finite number of observations. By hypothetically considering the possibility that sufficient number of observations are available, it is possible to consider whether it is logically possible to determine a parameter of interest either in the sense of its point value or in the sense of determining the set of which the parameter is an element. Therefore, identification is a fundamental consideration and logically occurs prior to and is separate from statistical estimation. A great deal of econometric literature on identification focuses on point identification. This is also the emphasis of this section. However, **set identification**, or **bounds identification**, is an important approach that will be used in selected places in this book (e.g., Chapters 25 and 27; see Manski, 1995).

Definition 2.3 (*Observational Equivalence*): Two structures of a model defined as joint probability distribution function $\Pr[\mathbf{x}|\boldsymbol{\theta}]$, $\mathbf{x} \in \mathbf{W}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, are **observationally equivalent** if $\Pr[\mathbf{x}|\boldsymbol{\theta}^1] = \Pr[\mathbf{x}|\boldsymbol{\theta}^2]\ \forall\ \mathbf{x} \in \mathbf{W}$.

Less formally, if, given the data, two structural models imply identical joint probability distributions of the variables, then the two structures are observationally equivalent. The existence of multiple observationally equivalent structures implies the failure of identification.

Definition 2.4 (*Identification*): A structure $\boldsymbol{\theta}^0$ is **identified** if there is no other observationally equivalent structure in $\boldsymbol{\Theta}$.

A simple example of nonidentification occurs when there is perfect collinearity between regressors in the linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. Then we can identify the linear combination $\mathbf{C}\boldsymbol{\beta}$, where $\text{rank}[\mathbf{C}] < \text{rank}[\boldsymbol{\beta}]$, but we cannot identify $\boldsymbol{\beta}$ itself.

This definition concerns uniqueness of the structure. In the context of the SEM we have given, this definition means that identification requires that there is a unique triple $(\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Sigma})$ consistent with the observed data. In SEM, as in other cases, identification involves being able to obtain unique estimates of structural parameters given the sample moments of the data. For example, in the case of the reduced form (2.12), under the stated assumptions the least-squares estimator provides unique estimates of $\mathbf{\Pi}$, that is, $\widehat{\mathbf{\Pi}} = [\mathbf{Z}'\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{Y}$, and identification of $\mathbf{B}, \mathbf{\Gamma}$ requires that there is a solution for the unknown elements of $\mathbf{\Gamma}$ and $\mathbf{B}$ from the equations $\mathbf{\Pi} + \mathbf{\Gamma}\mathbf{B}^{-1} = \mathbf{0}$, given a priori restrictions on the model. A unique solution implies just identification of the model.

A complete model is said to be identified if all the model parameters are identified. It is possible that for some models only a subset of parameters is identified. In some situations it may be important to be able to identify some function of parameters, and not necessarily all the individual parameters. Identification of a function of parameters means that function can be recovered uniquely from $F(\mathbf{W}|\mathbf{\Theta})$.

How does one ensure that the structures of alternative model specifications can be "ruled out"? In SEM the solution to this problem depends on augmenting the sample information by a priori restrictions on ($\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Sigma}$). The a priori restrictions must introduce sufficient additional information into the model to rule out the existence of other observationally equivalent structures.

The need for a priori restrictions is demonstrated by the following argument. First note that given the assumptions of Section 2.4.1 the reduced form, defined by ($\mathbf{\Pi}$, $\mathbf{\Omega}$), is always unique. Initially suppose there are no restrictions on ($\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Sigma}$). Next suppose that there are two observationally equivalent structures ($\mathbf{B}_1$, $\mathbf{\Gamma}_1$, $\mathbf{\Sigma}_1$) and ($\mathbf{B}_2$, $\mathbf{\Gamma}_2$, $\mathbf{\Sigma}_2$). Then

$$\mathbf{\Pi} = -\mathbf{\Gamma}_1\mathbf{B}_1^{-1} = -\mathbf{\Gamma}_2\mathbf{B}_2^{-1}. \tag{2.19}$$

Let $\mathbf{H}$ be a $G \times G$ nonsingular matrix. Then $\mathbf{\Gamma}_1\mathbf{B}_1^{-1} = \mathbf{\Gamma}_1\mathbf{H}\mathbf{H}^{-1}\mathbf{B}_1^{-1} = \mathbf{\Gamma}_2\mathbf{B}_2^{-1}$, which means that $\mathbf{\Gamma}_2 = \mathbf{\Gamma}_1\mathbf{H}$, $\mathbf{B}_2 = \mathbf{B}_1\mathbf{H}$. Thus the second structure is a linear transformation of the first.

The SEM solution to this problem is to introduce restrictions on ($\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Sigma}$) such that we can rule out the existence of linear transformations that lead to observationally equivalent structures. In other words, the restrictions on ($\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Sigma}$) must be such that there is no matrix $\mathbf{H}$ that would yield another structure with the same reduced form; given ($\mathbf{\Pi}$, $\mathbf{\Omega}$) there will be a unique solution to the equations $\mathbf{\Pi} = -\mathbf{\Gamma}\mathbf{B}^{-1}$ and $\mathbf{\Omega} \equiv (\mathbf{B}^{-1})'\mathbf{\Sigma}\mathbf{B}^{-1}$.

In practice a variety of restrictions can be imposed including (1) normalizations, such as setting diagonal elements of $\mathbf{B}$ equal to 1, (2) zero (exclusion) and linear homogeneous and inhomogeneous restrictions, and (3) covariance and inequality restrictions. Details of the necessary and sufficient conditions for identification in linear and nonlinear models can be found in many texts including Sargan (1988).

Meaningful imposition of identifying restrictions requires that the a priori restrictions imposed should be valid a posteriori. This idea is pursued further in several chapters where identification issues are considered (e.g., Section 6.9).

Exclusion restrictions essentially state that the model contains some variables that have zero impact on some endogenous variables. That is, certain directions of causation are ruled out a priori. This makes it possible to identify other directions of causation. For example, in the simple two-variable example given earlier, z_1 did not enter the y_1-equation, making it possible to identify the direct impact of y_2 on y_1. Although exclusion restrictions are the simplest to apply, in parametric models identification can also be secured by inequality restrictions and covariance restrictions.

If there are no restrictions on $\mathbf{\Sigma}$, and the diagonal elements of $\mathbf{B}$ are normalized to 1, then a **necessary condition** for identification is the **order condition**, which states that the number of excluded exogenous variables must at least equal the number of included endogenous variables. A **sufficient condition** is the **rank condition** given in many texts that ensures for the jth equation parameters $\mathbf{\Pi}\mathbf{\Gamma}_j = -\mathbf{B}_j$ yields a unique solution for $(\mathbf{\Gamma}_j, \mathbf{B}_j)$ given $\mathbf{\Pi}$.

Given identification, the term **just (exact) identification** refers to the case when the order condition is exactly satisfied; **overidentification** refers to the case when the number of restrictions on the system exceeds that required for exact identification.

Identification in nonlinear SEM has been discussed in Sargan (1988), who also gives references to earlier related work.

2.6. Single-Equation Models

Without loss of generality consider the first equation of a linear SEM subject to normalization $\beta_{11} = 1$. Let $y = y_1$, let $\mathbf{y}_1$ denote the endogenous components of $\mathbf{y}$ other than y_1, and let $\mathbf{z}_1$ denote the exogenous components of $\mathbf{z}$ with

$$y = \mathbf{y}_1'\boldsymbol{\alpha} + \mathbf{z}_1'\boldsymbol{\gamma} + u. \tag{2.20}$$

Many studies skip the formal steps involved in going from a system to a single equation and begin by writing the regression equation

$$y = \mathbf{x}'\boldsymbol{\beta} + u,$$

where some components of $\mathbf{x}$ are endogenous (implicitly $\mathbf{y}_1$) and others are exogenous (implicitly $\mathbf{z}_1$). The focus lies then on estimating the impact of changes in key regressor(s) that may be endogenous or exogenous, depending on the assumptions. Instrumental variable or two-stage least-squares estimation is the most obvious estimation strategy (see Sections 4.8, 6.4, and 6.5).

In the SEM approach it is natural to specify at least some of the remaining equations in the model, even if they are not the focus of inquiry. Suppose $\mathbf{y}_1$ has dimension 1. Then the first possibility is to specify the structural equation for y_1 and for the other endogenous variables that may appear in this structural equation for y_1. A second possibility is to specify the reduced form equation for y_1. This will show exogenous variables that affect y_1 but do not directly affect y. An advantage is that in such a setting instrumental variables emerge naturally. However, in recent empirical work using instrumental variables in a single-equation setting, even the formal step of writing down a reduced form for the endogenous right-hand-side variable is avoided.

2.7. Potential Outcome Model

Motivation for causal inference in econometric models is especially strong when the focus is on the impact of public policy and/or private decision variables on some specific outcomes. Specific examples include the impact of transfer payments on labor supply, the impact of class size on student learning, and the impact of health insurance on utilization of health care. In many cases the causal variables themselves reflect individual decisions and hence are potentially endogenous. When, as is usually the case, econometric estimation and inference are based on **observational data**, identification of and inference on causal parameters pose many challenges. These challenges can become potentially less serious if the causal issues are addressed using data from a controlled **social experiment** with a proper statistical design. Although such experiments have been implemented (see Section 3.3 for examples and details) they are generally expensive to organize and run. Therefore, it is more attractive to implement causal modeling using data generated by a **natural experiment** or in a quasi-experimental setting. Section 3.4 discusses the pros and cons of these data structures; but for present purposes one should think of a natural or **quasi experiment** as a setting in which some causal variable changes exogenously and independently of other explanatory variables, making it relatively easier to identify causal parameters.

A major obstacle for causality modeling stems from the *fundamental problem of causal inference* (Holland, 1986). Let X be the hypothesized cause and Y the outcome. By manipulating the value of X we can change the value of Y. Suppose the value of X is changed from x_1 to x_2. Then a measure of the causal impact of the change on Y is formed by comparing the two values of Y: y_2, which results from the change, and y_1, which would have resulted had no change in x occurred. However, if X did change, then the value of Y, in the absence of the change, would not be observed. Hence nothing more can be said about causal impact without some hypothesis about what value Y would have assumed in the absence of the change in X. The latter is referred to as a **counterfactual**, which means hypothetical unobserved value. Briefly stated, all causal inference involves comparison of a factual with a counterfactual outcome. In the conventional econometric model (e.g., SEM) a counterfactual does not need to be explicitly stated.

A relatively newer strand in the microeconometric literature – **program evaluation** or **treatment evaluation** – provides a statistical framework for the estimation of causal parameters. In the statistical literature this framework is also known as the **Rubin causal model (RCM)** in recognition of a key early contribution by Rubin (1974, 1978), who in turn cites R.A. Fisher as originator of the approach. Although, following recent convention, we refer to this as the Rubin causal model, Neyman (Splawa-Neyman) also proposed a similar statistical model in an article published in Polish in 1923; see Neyman (1990). Models involving counterfactuals have been independently developed in econometrics following the seminal work of Roy (1951). In the remainder of this section the salient features of RCM will be analyzed.

Causal parameters based on counterfactuals provide statistically meaningful and operational definitions of causality that in some respects differ from the traditional Cowles foundation definition. First, in ideal settings this framework leads to considerable simplicity of econometric methods. Second, this framework typically focuses on the *fewer* causal parameters that are thought to be most relevant to policy issues that are examined. This contrasts with the traditional econometric approach that focuses simultaneously on all structural parameters. Third, the approach provides additional insights into the properties of causal parameters estimated by the standard structural methods.

2.7.1. The Rubin Causal Model

The term "treatment" is used interchangeably with "cause." In medical studies of new drug evaluation, involving groups of those who receive the treatment and those who do not, the drug response of the treated is compared with that of the untreated. A measure of causal impact is the average difference in the outcomes of the treated and the nontreated groups. In economics, the term treatment is used very broadly. Essentially it covers variables whose impact on some outcome is the object of study. Examples of treatment–outcome pairs include schooling and wages, class size and scholastic performance, and job training and earnings. Note that a treatment need not be exogenous, and in many situations it is an endogenous (choice) variable.

Within the framework of a **potential outcome model (POM),** which assumes that every element of the target population is potentially exposed to the treatment, the triple (y_{1i}, y_{0i}, D_i), $i = 1, \ldots, N$, forms the basis of treatment evaluation. The categorical variable D takes the values 1 and 0, respectively, when treatment is or is not received; y_{1i} measures the response for individual i receiving treatment, and y_{0i} measures that when not receiving treatment. That is,

$$y_i = \begin{cases} y_{1i} \text{ if } D_i = 1, \\ y_{0i} \text{ if } D_i = 0. \end{cases} \tag{2.21}$$

Since the receipt and nonreceipt of treatment are mutually exclusive states for individual i, only one of the two measures is available for any given i, the unavailable measure being the counterfactual. The effect of the cause D on outcome of individual i is measured by $(y_{1i} - y_{0i})$. The average causal effect of $D_i = 1$, relative to $D_i = 0$, is measured by the **average treatment effect** (**ATE**):

$$\text{ATE} = \mathrm{E}[y|D = 1] - \mathrm{E}[y|D = 0], \tag{2.22}$$

where expectations are with respect to the probability distribution over the target population. Unlike the conventional structural model that emphasizes marginal effects, the POM framework emphasizes ATE and parameters related to it.

The experimental approach to the estimation of ATE-type parameters involves a **random assignment** of treatment followed by a comparison of the outcomes with a set of nontreated cases that serve as controls. Such an experimental design is explained in greater detail in Chapter 3. Random assignment implies that individuals exposed to treatment are chosen randomly, and hence the treatment assignment does not depend on the outcome and is uncorrelated with the attributes of treated subjects. Two major simplifications follow. The treatment variable can be treated as exogenous and its coefficient in a linear regression will not suffer from omitted variable bias if some relevant variables are unavoidably omitted from the regression. Under certain conditions, discussed at greater length in Chapters 3 and 25, the mean difference between the outcomes of the treated and the control groups will provide an estimate of ATE. The payoff to the well-designed experiment is the relative simplicity with which causal statements can be made. Of course, to ensure high statistical precision for the treatment effect estimate, one should still control for those attributes that also independently influence the outcomes.

Because random assignment of treatment is generally not feasible in economics, estimation of ATE-type parameters must be based on observational data generated under nonrandom treatment assignment. Then the consistent estimation of ATE will be threatened by several complications that include, for example, possible correlation between the outcomes and treatment, omitted variables, and endogeneity of the treatment variable. Some econometricians have suggested that the absence of randomization comprises the major impediment to convincing statistical inference about causal relationships.

The potential outcome model can lead to causal statements if the counterfactual can be clearly stated and made operational. An explicit statement of the counterfactual, with a clear implication of what should be compared, is an important feature of this model. If, as may be the case with observational data, there is lack of a clear distinction between observed and counterfactual quantities, then the answer to the question of who is affected by the treatment remains unclear. ATE is a measure that weights and combines marginal responses of specific subpopulations. Specific assumptions are required to operationalize the counterfactual. Information on both treated and untreated units that can be observed is needed to estimate ATE. For example, it is necessary to identify the untreated group that proxies the treated group if the treatment were not applied. It is not necessarily true that this step can always be implemented. The exact way in which the treated are selected involves issues of sampling design that are also discussed in Chapters 3 and 25.

A second useful feature of the POM is that it identifies opportunities for causal modeling created by natural or quasi-experiments. When data are generated in such settings, and provided certain other conditions are satisfied, causal modeling can occur without the full complexities of the SEM framework. This issue is analyzed further in Chapters 3 and 25.

Third, unlike the structural form of the SEM where all variables other than that being explained can be labeled as "causes," in the POM not all explanatory variables can be regarded as causal. Many are simply attributes of the units that must be controlled for in regression analysis, and attributes are not causes (Holland, 1986). Causal parameters must relate to variables that are actually or potentially, and directly or indirectly, subject to intervention.

Finally, identifiability of the ATE parameter may be an easier research goal and hence feasible in situations where the identifiability of a full SEM may not be (Angrist, 2001). Whether this is so has to be determined on a case-by-case basis. However, many available applications of the POM typically employ a limited, rather than full, information framework. However, even within the SEM framework the use of a limited information framework is also feasible, as was previously discussed.

2.8. Causal Modeling and Estimation Strategies

In this section we briefly sketch some of the ways in which econometricians approach the modeling of causal relationships. These approaches can be used within both SEM and POM frameworks, but they are typically identified with the former.

2.8.1. Identification Frameworks

Full-Information Structural Models

One variant of this approach is based on the parametric specification of the joint distribution of endogenous variables conditional on exogenous variables. The relationships are not necessarily derived from an optimizing model of behavior. Parametric restrictions are placed to ensure identification of the model parameters that are the target of statistical inference. The entire model is estimated simultaneously using maximum likelihood or moments-based estimation. We call this approach the **full-information structural approach**. For well-specified models this is an attractive approach but in general its potential limitation is that it may contain some equations that are poorly specified. Under joint estimation the effects of localized misspecification may also affect other estimates.

Statistically we may interpret the full-information approach as one in which the joint probability distribution of endogenous variables, given the exogenous variables, forms the basis of inference about causality. The jointness may derive from contemporaneous or dynamic interdependence between endogenous variables and/or the disturbances on the equations.

Limited-Information Structural Models

By contrast, when the central object of statistical inference is estimation of one or two key parameters, a **limited-information** approach may be used. A feature of this approach is that, although one equation is the focus of inference, the joint dependence between it and other endogenous variables is exploited. This requires that explicit assumptions are made about some features of the model that are not the main object of inference. Instrumental variable methods, sequential multistep methods, and limited information maximum likelihood methods are specific examples of this approach. To implement the approach one typically works with one (or more) structural equations and some implicitly or explicitly stated reduced form equations. This contrasts with the full-information approach where all equations are structural. The limited-information approach is often computationally more tractable than the full-information one.

Statistically we may interpret the limited-information approach as one in which the joint distribution is factored into the product of a conditional model for the endogenous variable(s) of interest, say $\mathbf{y}_1$, and a marginal model for other endogenous variables, say $\mathbf{y}_2$, which are in the set of the conditioning variables, as in

$$f(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = g(\mathbf{y}_1|\mathbf{x}, \mathbf{y}_2, \boldsymbol{\theta}_1)h(\mathbf{y}_2|\mathbf{x}, \boldsymbol{\theta}_2), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{2.23}$$

Modeling may be based on the component $g(\mathbf{y}_1|\mathbf{x}, \mathbf{y}_2, \boldsymbol{\theta}_1)$ with minimal attention to $h(\mathbf{y}_2|\mathbf{x}, \boldsymbol{\theta}_2)$ if $\boldsymbol{\theta}_2$ are regarded as **nuisance parameters.** Of course, such a factorization is not unique, and hence the limited-information approach can have several variants.

Identified Reduced Forms

A third variant of the SEM approach works with an **identified reduced form**. Here too one is interested in structural parameters. However, it may be convenient to estimate these from the reduced form subject to restrictions. In time series the identified vector autoregressions provide an example.

2.8.2. Identification Strategies

There are numerous potential ways in which the identification of key model parameters can be jeopardized. Omitted variables, functional form misspecifications, measurement errors in explanatory variables, using data unrepresentative of the population, and ignoring endogeneity of explanatory variables are leading examples. Microeconometrics contains many specific examples of how these challenges can be tackled. Angrist and Krueger (2000) provide a comprehensive survey of popular identification strategies in labor economics, with emphasis on the POM framework. Most of the issues are developed elsewhere in the book, but a brief mention is made here.

Exogenization

Data are sometimes generated by natural experiments and quasi-experiments. The idea here is simply that a policy variable may exogenously change for some subpopulation while it remains the same for other subpopulations. For example, minimum wage laws in one state may change while they remain unchanged in a neighboring state. Such events naturally create treatment and control groups. If the natural experiment approximates a randomized treatment assignment, then exploiting such data to estimate structural parameters can be simpler than estimation of a larger simultaneous equations model with endogenous treatment variables. It is also possible that the treatment variable in a natural experiment can be regarded as exogenous, but the treatment itself is not randomly assigned.

Elimination of Nuisance Parameters

Identification may be threatened by the presence of a large number of nuisance parameters. For example, in a cross-section regression model the conditional mean function $\mathrm{E}[y_i|\mathbf{x}_i]$ may involve an individual specific fixed effect α_i, assumed to be correlated with the regression error. This effect cannot be identified without many observations on each individual (i.e., panel data). However, with just a short panel it could be eliminated by a transformation of the model. Another example is the presence of timeinvariant unobserved exogenous variables that may be common to groups of individuals.

An example of a transformation that eliminates fixed effects is taking differences and working with the differences-in-differences form of the model.

Controlling for Confounders

When variables are omitted from a regression, and when omitted factors are correlated with the included variables, a confounding bias results. For example, in a regression with earnings as a dependent variable and schooling as an explanatory variable, individual ability may be regarded as an omitted variable because only imperfect proxies for it are typically available. This means that potentially the coefficient of the schooling variable may not be identified. One possible strategy is to introduce **control variables** in the model; the general approach is called the **control function approach**. These variables are an attempt to approximate the influence of the omitted variables. For example, various types of scholastic achievement scores may serve as controls for ability.

Creating Synthetic Samples

Within the POM framework a causal parameter may be unidentified because no suitable comparison or control group can provide the benchmark for estimation. A potential solution is to create a synthetic sample that includes a comparison group that are proxies for controls. Such a sample is created by **matching** (discussed in Chapter 25). If treated samples can be augmented by well-matched controls, then identification of causal parameters can be achieved in the sense that a parameter related to ATE can be estimated.

Instrumental Variables

If identification is jeopardized because the treatment variable is endogenous, then a standard solution is to use valid instrumental variables. This is easier said than done. The choice of the instrumental variable as well as the interpretation of the results obtained must be done carefully because the results may be sensitive to the choice of instruments. The approach is analyzed in Sections 4.8, 4.9, 6.4, 6.5, and 25.7, as well as in several other places in the book as the need arises. Again a natural experiment may provide a valid instrument.

Reweighting Samples

Sample-based inferences about the population are only valid if the sample data are representative of the population. The problem of sample selection or biased sampling arises when the sample data are not representative, in which case the population parameters are not identified. This problem can be approached as one that requires correction for sample selection (Chapter 16) or one that requires reweighting of the sample information (Chapter 24).

2.9. Bibliographic Notes

2.1 The 2001 Nobel lectures by Heckman and McFadden are excellent sources for both historical and current information about the developments in microeconometrics. Heckman's lecture is remarkable for its comprehensive scope and offers numerous insights into many aspects of microeconometrics. His discussion of heterogeneity has many points of contact with several topics covered in this book.

2.2 Marschak (1953) gives a classic statement of the primacy of structural modeling for policy evaluation. He makes an early mention of the idea of parameter invariance.

2.3 Engle, Hendry, and Richard (1983) provide definitions of weak and strong exogeneity in terms of the distribution of observable variables. They make links with previous literature on exogeneity concepts.

2.4 and **2.5** The term "identification" was used by Koopmans (1949). Point identification in linear parametric models is covered in most textbooks including those by Sargan (1988) who gives a comprehensive and succint treatment, Davidson and MacKinnon (2004), and Greene (2003). Gouriéroux and Monfort (1989, chapter 3.4) provide a different perspective using Fisher and Kullback information measures. Bounds identification in several leading cases is developed in Manski (1995).

2.6 Heckman (2000) provides a historical overview and modern interpretations of causality in the traditional econometric model. Causality concepts within the POM framework are carefully and incisively analyzed by Holland (1986), who also relates them to other definitions. A sample of the statisticians' viewpoints of causality from a historical perspective can be found in Freedman (1999). Pearl (2000) gives insightful schematic exposition of the idea of "treating causation as a summary of behavior under interventions," as well as numerous problems of inferring causality in a nonexperimental situation.

2.7 Angrist and Krueger (1999) survey solutions to identification pitfalls with examples from labor economics.

CHAPTER 3

Microeconomic Data Structures

3.1. Introduction

This chapter surveys issues concerning the potential usefulness and limitations of different types of microeconomic data. By far the most common data structure used in microeconometrics is survey or census data. These data are usually called **observational data** to distinguish them from **experimental data**.

This chapter discusses the potential limitation of the aforementioned data structures. The inherent limitations of observational data may be further compounded by the manner in which the data are collected, that is, by the sample frame (the way the sample is generated), sample design (simple random sample versus stratified random sample), and sample scope (cross-section versus longitudinal data). Hence we also discuss sampling issues in connection with the use of observational data. Some of this terminology is new at this stage but will be explained later in this chapter.

Microeconometrics goes beyond the analysis of survey data under the assumptions of simple random sampling. This chapter considers extensions. Section 3.2 outlines the structure of multistage sample surveys and some common forms of departure from random sampling; a more detailed analysis of their statistical implications is provided in later chapters. It also considers some commonly occurring complications that result in the data not being necessarily representative of the population. Given the deficiencies of observational data in estimating causal parameters, there has been an increased attempt at exploiting experimental and quasi-experimental data and frameworks. Section 3.3 examines the potential of data from social experiments. Section 3.4 considers the modeling opportunities arising from a special type of observational data, generated under quasi-experimental conditions, that naturally provide treated and untreated subjects and hence are called natural experiments. Section 3.5 covers practical issues of microdata management.