CHAPTER 1. INTRODUCTION

Throughout the social sciences, scholars are aware that models estimated using observational data are likely affected by endogeneity for one or more predictors. Indeed, social scientists often construct models that explicitly specify reciprocal relationships or feedback loops among multiple outcomes. But many social scientists may underestimate the consequences of endogeneity resulting from such "nonrecursive models" or be unaware of how to correctly model or otherwise address it. As a result, it is not unusual for researchers to adopt an estimation strategy that effectively ignores such possible endogeneity and results in biased estimates. Equally problematic is the estimation or interpretation of results from nonrecursive models in which the assumptions of the model are either statistically or theoretically inappropriate.

This monograph provides an overview of methods appropriate for the analysis of nonrecursive simultaneous equation models. Simultaneous equation models have at least two equations and stand in contrast to the more common instance in which social scientists posit a single equation with a single outcome. A simultaneous equation model is nonrecursive if (1) two of the outcomes in the model affect one another (a reciprocal relationship) or there is a feedback loop at some point in the system of equations (i.e., a causal path can be traced from one variable back to itself), and/or (2) at least some disturbances are correlated. A model is termed fully recursive if there is no posited reciprocal relationship or feedback loop, *and* there are assumed to be no relationships among the error terms of the equations.

We introduce the specification, identification, and estimation of simultaneous equation models, how to assess the quality of the estimates, and how to correctly interpret results, with a focus on nonrecursive models. In nonrecursive models, identification of an equation (demonstrating that unique values for the parameters can be estimated) often requires the incorporation of a variable that is correlated with the problematic variable but uncorrelated with the disturbance term of the equation: This is referred to as an instrumental variable. Another focus of the monograph will be introducing proper selection, use, and assessment of instrumental variables. We will emphasize that properly selecting instrumental variables is important regardless of which estimator is used.

Throughout the monograph, we blend two complementary perspectives on simultaneous equation nonrecursive models. First, simultaneous equations are addressed in the structural equation modeling (SEM) with latent variables literature (e.g., Bollen, 1989b; Kaplan, 2009).[1] The SEM literature stresses specification using path diagrams, full-information estimation, and global assessment of model fit. The SEM approach to simultaneous equations is limited, however, by a relative neglect of nonrecursive models, a lack of assessment of individual equations, and little discussion of the quality of the instrumental variables used to identify nonrecursive equations. Many SEM treatments of simultaneous equations focus almost exclusively on full-information estimation strategies such as maximum likelihood. But a limited-information approach is useful for several tests that assess the quality of the individual equations in the model. A focus on maximum likelihood alone leaves researchers without the tools to properly evaluate the assumptions underlying their nonrecursive models.

A complementary approach to simultaneous equation models comes from the econometrics tradition (e.g., Greene, 2008; Kennedy, 2008; Wooldridge, 2002, 2009).[2] The econometric literature highlights the link between simultaneous equation models and assumption violations of traditional regression techniques, stresses the identification of nonrecursive models using instrumental variables, and focuses on limited-information estimators to a greater extent. But because the econometrics literature focuses on individual equations, it does not stress the interpretation of results as part of a multiequation model. Furthermore, the range of possible assessments of instrumental variables are rarely consolidated and compared.

Throughout this monograph, we take the position that limited-information estimators such as two-stage least squares are not outdated methods that can be safely ignored by researchers using structural equation software packages. Instead, we argue that a need for a clear and rigorous understanding of nonrecursive models has developed. With increasing ease of software programs, social scientists can now estimate nonrecursive models without fully understanding their unique attributes. Many full-information software programs do not provide equation-by-equation assessment of the quality of the model proposed. Important information is contained in the reduced-form equations of such models, and this information is often opaque to

[1]Simultaneous equation models differ from SEMs in that measurement error in the variables is rarely considered in simultaneous equation models, and there is no attempt to estimate latent variables (see Bollen, 1989b, pp. 80–150). The distinction is not hard and fast and can be viewed as largely a matter of emphasis.

[2]Confusion can arise from terminology. For example, some economists will refer to simultaneous equation models as we define them as structural equation models, although their approach is different from the latent variable SEM approach.

researchers adopting a full-information estimation approach. By integrating the econometrics and SEM approaches to simultaneous equation models, this monograph provides a "back to basics" approach that is directly relevant to researchers wishing to estimate nonrecursive models. The monograph is oriented around five steps we advocate for modeling: specification, identification, estimation, assessment, and interpretation. Related volumes in the series expand on particular steps: for example, Berry (1984) provides extensive coverage of identification.

In our presentation, we assume knowledge of multiple regression analysis. Readers will further benefit if they have some familiarity with SEM. Good presentations of SEM are available in Bollen (1989b) and Kaplan (2009). Models will be presented and discussed as path diagrams, equations, or matrix equations. Good introductions to matrix algebra can be found in Gill (2006, chaps. 3 and 4), Fox (2009), and Namboodiri (1984). We present examples using some of the available software, with a focus on software such as SAS and Stata that allow estimation of individual equations.