Chapter 22

On Exogeneity

David Kaplan

22.1. Introduction

When using linear statistical models to estimate substantive relationships, a distinction is made between endogenous variables and exogenous variables. Alternative names for these variables are *dependent* and *independent,* or *criterion* and *predictor.* In the case of multiple linear regression, one variable is designated as the endogenous variable, and the remaining variables are designated as exogenous. In multivariate regression, a set of endogenous variables is chosen and related to one or more exogenous variables. In the case of structural equation modeling, there is typically a set of endogenous variables that are related to each other and also related to a set of exogenous variables.

More often than not, the choice of endogenous and exogenous variables is guided by the research question of interest, with little consideration given to statistical consequences of that choice. Moreover, an inspection of standard textbooks in the social and behavioral sciences reveals confusing definitions of endogenous and exogenous variables. For example, Cohen and Cohen (1983) write,

> Exogenous variables are measured variables that are not caused by any other variable in the model except (possibly) other exogenous variables. They have essentially the same meaning as independent variables in ordinary regression analysis except that they explicitly include the assumption that they are not causally dependent on the endogenous variables in the model. Endogenous variables are, in part, effects of exogenous variables and do not have a causal effect on them. (p. 375)

In another example, Bollen (1989) writes,

> The terms exogenous and endogenous are model specific. It may be that an exogenous variable in one model is endogenous in another. Or, a variable shown as exogenous, in reality, may be influenced by a variable in the model. Regardless of these possibilities, the convention is to refer to variables as exogenous or endogenous based on their representation in a particular model. (p. 12)

And finally, from an econometric perspective, Wonnacott and Wonnacott (1979) write with regard to treating income (denoted as I in their definition) as an exogenous variable,

> An important distinction must be made between two kinds of variables in our system. By assumption, I is an exogenous variable. Since its value is determined from *outside* the system, it will often be referred to as *predetermined;* however it should be recognized that a predetermined variable may be either fixed *or* random. The essential point is that its values are determined elsewhere. (pp. 257–258)

The above definitions of exogenous variables are typical of those found in most social science

AUTHOR'S NOTE: The author is grateful to Professor Aris Spanos for valuable comments on an earlier draft of this chapter.

statistics textbooks.[1] Nevertheless, these and other similar definitions are problematic for a number of reasons. First, these definitions do not articulate precisely what it means to say that exogenous variables are not dependent on the endogenous variables in the model. For example, with the Cohen and Cohen (1983) definition, if an exogenous variable is possibly dependent on other exogenous variables, then these "dependent" exogenous variables are actually endogenous, and what is being described by this definition is a system of structural equations. Second, Bollen's (1989) definition, although accurately describing the standard convention, seems to confuse a variable's representation in a model with exogeneity. However, the location of a variable in a model does not necessarily render the variable exogenous. In other words, Bollen's definition implies that simply stating that a variable is exogenous makes it so. Moreover, Bollen defines *exogeneity* with respect to a model and not with respect to the statistical structure of the data used to test the model. Third, in the Wonnacott and Wonnacott (1979) definition, the notion of "outside the system" is never really developed. Implied by these different definitions is a confusion between theoretical exogeneity versus statistical exogeneity and the consequences for the former when the latter does not hold.

From our discussion so far, it is clear that these common definitions of exogeneity do not provide a complete picture of the subtleties or seriousness of the problem. A more complete study of the problem of exogeneity comes from the work of Richard (1982) and his colleagues within the domain of econometrics. This chapter, therefore, provides a didactic introduction to the econometric notion of exogeneity as it pertains to linear regression with a brief discussion of the problem with respect to structural equation modeling, multilevel modeling, and growth curve modeling. It is the goal of this chapter to highlight the seriousness of examining exogeneity assumptions carefully when specifying statistical models—particularly if models are to be used for prediction or the evaluation of policies or interventions. Attention will focus primarily on the concept of weak exogeneity and informal methods for testing whether weak exogeneity holds. The more restrictive concept of strong exogeneity will be similarly introduced along with the notion of Granger noncausality, which will require incorporating a dynamic component into the simple linear regression model. Super exogeneity will be introduced along with related concepts of parameter constancy and invariance. Methods for testing strong and super exogeneity will be outlined. Weak, strong, and super exogeneity will be linked to the uses of a statistical model for inference, forecasting, and policy analysis, respectively.

The organization of this chapter is as follows. In Section 22.2, the general problem of exogeneity is introduced. In Section 22.3, the concept of weak exogeneity will be defined in the case of simple linear regression. Here, the auxiliary concepts of *parameters of interest* and *variation freeness* will be introduced. This section will also discuss exogeneity in the context of structural equation modeling. In Section 22.4, we will consider the conditions under which weak exogeneity can be assumed to hold, as well as conditions where it is likely to be violated. We will also consider three indirect but related tests of weak exogeneity. In Section 22.5, we will introduce a temporal component to the model that will lead to the concept of Granger noncausality and, in turn, to strong exogeneity. We will discuss these concepts as they pertain to the use of statistical models for prediction. In Section 22.6, we will consider the problem of super exogeneity and the concepts of parameter constancy and invariance. We will consider these concepts in light of their implications for evaluating interventions or policies. Finally, Section 22.7 will conclude with a discussion of the implications of the exogeneity assumption for the standard practice of statistical modeling, briefly touching on the implications of the exogeneity assumption for two other popular statistical methodologies in the social and behavioral sciences. Throughout this chapter, concepts will be grounded in substantive problems within the field of education and education policy.

22.2. The Problem of Exogeneity

It was noted in Section 22.1 that definitions of exogenous and endogenous variables encountered in standard social science statistics textbooks are often confusing. In this section, we consider the problem of defining exogeneity more carefully, relying on work in econometric theory. A collection of seminal papers on the problem of exogeneity can be found in Ericsson and Irons (1994), and a brief discussion of the problem was introduced to the structural equation modeling literature by Kaplan (2000).

To begin, it is typical to invoke the heuristic that an exogenous variable is one whose cause is determined from "outside the system under investigation." This heuristic is implied in the Wonnacott and

[1] It is also quite common to find that textbooks avoid a definition of exogenous variables altogether.

Wonnacott (1979) definition of an exogenous variable given above. Usually, the notion of a variable being generated from "outside the system" is another way of stating that there is zero covariance between the regressor and the disturbance term. However, such a heuristic is problematic upon close inspection because it does not explicitly define what "outside the system" actually means.

As a way of demonstrating the problem with this heuristic, consider the counterexample given by Hendry (1995) of a fixed-regressor model. To provide a substantive motivation for these ideas, consider the problem of estimating the relationship between reading proficiency in young children as a function of parental reading activities (e.g., how often each week parents read to their children). We may represent this relationship by the simple model

$$y_t = \beta x_t + u_t, \tag{1}$$

where y represents reading proficiency, x represents the parental reading activities, β is the regression coefficient, and u is the disturbance term, which is assumed to be $NID(0, \sigma_u^2)$. The subscript t denotes the particular time point of measurement—a distinction that might be needed with the analysis of panel data.

Typically, parental reading activities are treated as fixed. That is, at time t, levels of parental involvement in reading are assumed to be set and remain the same from that point on. If this assumption were true, then conditional estimation of reading proficiency given parental involvement in reading activities would be valid. However, it is probably not the case in practice that parental reading activities are fixed but rather are likely to be a function of past parental reading activities. That is, perhaps the mechanism that generates parental reading activities at time t is better represented by a first-order autoregressive model,

$$x_t = \gamma x_{t-1} + v_t, \tag{2}$$

where we will assume that $|\gamma| < 1$, ensuring a stable autoregressive process. Even if it were the case that the model in equation (2) generated parental reading activities *prior* to generating reading proficiency, that is still not a sufficient condition to render parental reading activities exogenous in this example. The reason is that such a condition does not preclude current disturbances in equation (1) to be related to past disturbances in equation (2)—namely,

$$u_t = \varphi v_{t-1} + \varepsilon_t. \tag{3}$$

If equation (3) holds for $\varphi \neq 0$, then

$$\begin{aligned} E(x_t, u_t) &= E[(\gamma x_{t-1} + v_t)(\varphi v_{t-1} + \varepsilon_t)] \\ &= \gamma\varphi\sigma_v^2, \end{aligned} \tag{4}$$

and, therefore, x_t is correlated with u_t and hence is not exogenous.

This simple counterexample serves to illustrate the subtleties of the problem of exogeneity. Despite treating parental reading activities as a fixed regressor and assuming that it is generated "from outside the system," the fact is that the true mechanism that generates current values of the regressor yields a model in which the regressor is correlated with the disturbance term, suggesting that it is generated from inside the system as far as the model is concerned. Therefore a rigorous definition of exogeneity is required that does not depend on the particular model under study but rather is based on the true structure of the system under investigation (Hendry, 1995).

22.3. Weak Exogeneity

Having shown that the concept of exogeneity is more subtle than standard definitions imply, we can begin our formal discussion of the problem by introducing the concept of weak exogeneity, which will serve to set the groundwork for subsequent discussions of other forms of exogeneity. To fix ideas, consider a matrix of variables denoted as $\mathbf{z}$ of order $N \times r$, where N is the sample size and r is the number of variables. Under the assumption of independent observations, the joint distribution of $\mathbf{z}$ is given as

$$f(\mathbf{z}|\boldsymbol{\theta}) = f(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\mathbf{z}_i|\boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\theta}$ is a vector of parameters of the joint distribution of $\mathbf{z}$. Most statistical modeling requires a partitioning of $\mathbf{z}$ into endogenous variables to be modeled and exogenous variables that are assumed to account for the variation and covariation in the endogenous variables. Denote by $\mathbf{y}$ the $N \times p$ matrix of endogenous variables and denote by $\mathbf{x}$ an $N \times q$ matrix of exogenous variables where $r = p + q$. We can rewrite equation (1) in terms of the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ and the marginal distribution of $\mathbf{x}$. That is, equation (1) can be related to the conditional distribution in the following decomposition:

$$f(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{x}, \boldsymbol{\omega}_1) f(\mathbf{x}, \boldsymbol{\omega}_2), \tag{6}$$

where $\boldsymbol{\omega}_1$ are the parameters associated with the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$, and $\boldsymbol{\omega}_2$ are the parameters associated with the marginal distribution of $\mathbf{x}$. The parameter spaces of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are denoted as $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, respectively.

It is clear that factoring the joint distribution in equation (5) into the product of the conditional distribution and marginal distribution in equation (6) presents no loss of information. However, standard statistical modeling almost always focuses on the conditional distribution in equation (6). Indeed, the conditional distribution is often referred to as the *regression function.* That being the case, then focusing on the conditional distribution assumes that the marginal distribution can be taken as given (Ericsson, 1994). The issue of exogeneity concerns the implications of this assumption for the parameters of interest.

22.3.1. Variation Freeness

Another important concept as it relates to the problem of exogeneity is that of *variation freeness.* Specifically, variation freeness means that for any value of $\boldsymbol{\omega}_2$ in $\boldsymbol{\Omega}_2$, $\boldsymbol{\omega}_1$ can take on any value in $\boldsymbol{\Omega}_1$ and vice versa (Spanos, 1986). In other words, it is assumed that the pair $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ belong to the product of their respective parameter spaces—namely, $(\boldsymbol{\Omega}_1 \times \boldsymbol{\Omega}_2)$—and that the parameter space $\boldsymbol{\Omega}_1$ is not restricted by $\boldsymbol{\omega}_2$ and vice versa. Thus, knowing the value of a parameter in the marginal model provides no information regarding the range of values that a parameter in the conditional model can take. Alternatively, restricting $\boldsymbol{\omega}_2$ in any way that ensures that $\boldsymbol{\omega}_2$ is in $\boldsymbol{\Omega}_2$ does not restrict $\boldsymbol{\omega}_1$ in any way that does not allow it to take all possible values in $\boldsymbol{\Omega}_1$.

As an example of variation freeness, consider a simple regression model with one endogenous variable y and one exogenous variable x. The parameters of interest of the conditional distribution are $\boldsymbol{\omega}_1 \equiv (\beta_0, \beta_1, \sigma_u^2)$, and the parameters of the marginal distribution are $\boldsymbol{\omega}_2 \equiv (\mu_x, \sigma_x^2)$. Furthermore, note that $\beta_1 = \sigma_{xy}/\sigma_x^2$, where σ_{xy} denotes the covariance of x and y. Following Ericsson (1994), if σ_{xy} varies proportionally with σ_x^2, then σ_x^2, which is in $\boldsymbol{\omega}_2$, carries no information relevant for the estimation of $\beta_1 = \sigma_{xy}/\sigma_x^2$, which is in $\boldsymbol{\omega}_1$. Therefore, $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are variation free. An example in which variation freeness could be violated is in cases where a parameter in the conditional model is constrained to be equal to a parameter in the marginal model—however, such cases are rare in the social and behavioral sciences. Below we will show an example in which the condition of variation freeness does not hold.

22.3.2. Parameters of Interest

Variation freeness does not guarantee that one can ignore the marginal model when interest centers on the parameters of the conditional model. As in Ericsson (1994), if interest centers on estimating the conditional and marginal means, then both the conditional and marginal models are needed.[2] This requires us to focus the issue of variation freeness on the *parameters of interest*—namely, those parameters that are a function of the parameters of the conditional model only. More formally, the parameters of interest $\boldsymbol{\Psi}$ are a function of $\boldsymbol{\omega}_1$; that is, $\boldsymbol{\Psi} = g(\boldsymbol{\omega}_1)$.

22.3.3. A Definition of Weak Exogeneity

The above concepts of factorization, parameters of interest, and variation freeness lead to a definition of weak exogeneity. Specifically, following Richard (1982; see also Ericsson, 1994; Spanos, 1986), a variable x is weakly exogenous for the parameters of interest (say, $\boldsymbol{\Psi}$) if and only if there exists a reparameterization of $\boldsymbol{\theta}$ as $\boldsymbol{\omega}$ with $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$, such that

(i) $\boldsymbol{\Psi} = g(\boldsymbol{\omega}_1)$—that is, $\boldsymbol{\Psi}$ is a function of $\boldsymbol{\omega}_1$ only—and
(ii) $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are variation free—that is, $(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \in \boldsymbol{\Omega}_1 \times \boldsymbol{\Omega}_2$.

22.3.4. Weak Exogeneity and the Problem of Nominal Regressors[3]

It is quite common in the social and behavioral sciences for models to contain regressor variables whose scales are nominal. Examples of such variables include demographic features of individuals such as gender or race. In other cases, nominal variables may represent orthogonal components of an experimental design, such as assignment to a treatment or control group. In both cases, the regressors are fixed, nonstochastic constants to be contrasted with stochastic random variables such as socioeconomic status or the amount of parental reading activities. In both cases, data are often submitted to some regression analysis software package for estimation. In the case of experimental design variables, data are

[2] In point of fact, however, one can "recover" the marginal mean of $\mathbf{x}$ from the constant in a regression.

[3] The author is grateful to Professor Aris Spanos for clarifying this issue.

often submitted to an analysis-of-variance (ANOVA) package. Experimental design textbooks often include a discussion of how ANOVA can be viewed as a special case of the "linear regression" model (see, e.g., Kirk, 1995). The similarity of ANOVA and the linear regression model generates a problem with respect to our discussion of exogeneity. Specifically, given our discussion of exogeneity to this point, a fair question may be to what extent nominal variables, such as gender, race, or experimental design arrangements, are "exogenous" for statistical estimation. In what sense are these variables generated from "outside the system"?

That the question of the "exogeneity" of nominal regressors is raised at all is suggestive of a conflation of ideas typically represented in statistical textbooks in the social sciences—specifically, the merging of the so-called *Gauss linear model* and the *linear regression model* (Spanos, 1999). Indeed, the similarity of the notation of both models contributes to the confusion.

Briefly, the origins of the Gauss linear model came about as an attempt to explain lawful relationships in planetary orbits using less than perfectly accurate measuring instruments. In that context, the Gauss linear model represented an "experimental design" situation in which the xs were fixed, nonstochastic constants albeit subject to observational error. Only the outcome variable y was considered to be a random variable. Indeed, according to Spanos (1999), the original linear model, as proposed by Legendre (1805), did not rest on any formal probabilistic arguments whatsoever. Rather, probabilistic arguments regarding the structure of the errors were added by Gauss and Laplace to justify the statistical optimality of the least squares approach to parameter estimation. Specifically, if it could be assumed that the errors were normal, independent, and identically distributed, then the least squares approach attained certain optimal properties. Later, Fisher applied the Gauss linear model to experimental designs and added the idea of randomization.

What is important for our discussion is that the Gauss linear model was not explicitly rooted in probabilistic notions of random variables, leading, in turn, to notions of conditional versus marginal distributions. It was Galton, with assistance from Karl Pearson, who later proposed the linear regression model, unaware that it was in any way related to the Gauss linear model. The hope was to use the rigorous "lawlike" modeling ideas of Gauss to support Galton's emerging theories of heredity and eugenics (Spanos, 1999). However, it was G. U. Yule (1897) who demonstrated that the same method of least squares used to estimate the Gauss linear model could also be used to estimate Galton's linear regression model (Mulaik, 1985). In this case, y and x were assumed to be jointly normal random variables, and βx was defined as the conditional expectation of y given x, where x is the realization of a stochastic random variable X.

Defining the conditional expectation requires being able to factor the joint distribution into the conditional and marginal distributions, and this requires stochastic random regressors (Spanos, 1999). Therefore, from the standpoint of our discussion of exogeneity, nominal regressors such as race, gender, or experimental design variables do not lead to any conceptual difficulty. When such variables are of interest, one has specified a Gauss linear model. The notion of the conditional distribution does not enter into the discussion because factoring the joint distribution into the conditional and marginal distributions is only possible in the case of stochastic random regressors. In the context of the linear regression model, however, nonstochastic variables enter the conditional mean via the marginal means of the stochastic variables; that is, the constant term is a function of the nonstochastic variables and is therefore not constant.[4]

22.3.5. An Extension to Structural Equation Modeling

It may be of interest to examine how the problem of weak exogeneity extends to structural equation modeling. We focus on structural equation modeling because it had its origins primarily in econometrics (see Kaplan, 2000, for a brief history), and certain aspects of its development are relevant to our discussion of exogeneity. We consider the problem of exogeneity with reference to other methodologies in Section 22.7.

To examine the relevance of weak exogeneity for structural equation models, we should revisit the distinction between the *structural form* and the *reduced-form* specifications of a structural equation model. The structural form of the general structural equation model is denoted as (e.g., Jöreskog, 1973)

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \tag{7}$$

[4] To see this, consider the addition of a nonstochastic variable (say, gender) to a regression model with other stochastic regressors. Heterogeneity in the mean of y and the mean of x induced by gender can be modeled as $\mu_y = a(gender)$, and $\mu_x = d(gender)$, where a and d are parameters. Expressed in terms of the regression function, $\mu_y = \beta_0 + \beta_1 \mu_x$. After substitution, $a(gender) = \beta_0 + \beta_1 d(gender)$, from which we obtain $\beta_0 = (a - \beta_1 d)gender$. Thus, the constant term is a function of a nonstochastic variable.

where $\mathbf{y}$ is a vector of endogenous variables, $\boldsymbol{\alpha}$ is a vector of structural intercepts, $\mathbf{B}$ is a matrix of coefficients relating endogenous variables to each other, $\boldsymbol{\Gamma}$ is a matrix relating endogenous variables to exogenous variables, $\mathbf{x}$ is a vector of exogenous variables, and $\boldsymbol{\zeta}$ is a vector of disturbance terms. In structural equation modeling, the structural parameters of interest are $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is the covariance matrix of the disturbance terms.

As noted above, equation (7) represents the structural form of the model. The specification of fixed or freed elements in $\mathbf{B}$ and/or $\boldsymbol{\Gamma}$ denotes a priori restrictions, presumably reflecting an underlying hypothesis regarding the mechanism that yields values of $\mathbf{y}$. The standard approach to structural equation modeling requires that certain assumptions be met for application of standard estimation procedures such as maximum likelihood. Specifically, it is generally assumed that the conditional distribution of the endogenous variables, given the exogenous variables, is multivariate normally distributed. Violations of this assumption can, in principle, be addressed via alternative estimation methods that explicitly capture the nonnormality of the data, such as Browne's asymptotic distribution-free estimator (Browne, 1984) or Muthén's weighted least squares estimator for categorical data (Muthén, 1984). If this or other assumptions are violated, then the standard likelihood ratio chi-square test, estimates, and standard errors will be incorrect. A fuller discussion of the assumptions of structural equation modeling can be found in Kaplan (2000).

With regard to the assumption of exogeneity, a perusal of extant textbooks and substantive literature on structural equation modeling suggests that the exogeneity of the predictor variables, as defined above, is not formally addressed—an exception being Kaplan (2000). Indeed, the extant literature reveals that only theoretical considerations are given when delimiting a variable as "exogenous."[5]Assessing exogeneity in terms of the statistical structure of the data requires that we revisit the reduced-form specification of a structural equation model.

22.3.6. The Reduced-Form Specification Revisited

In classic econometric treatments of structural equation modeling, the reduced form plays a central role in establishing the identification of structural parameters. The reduced-form specification of a structural model is derived from rewriting the structural form so that the endogenous variables are on one side of the equation, and the exogenous variables are on the other side. Specifically, considering equation (7), we have

$$\begin{aligned}\mathbf{y} &= \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta},\\ &= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta},\\ &= \boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1\mathbf{x} + \boldsymbol{\zeta}^*, \qquad (8)\end{aligned}$$

where it is assumed that $(\mathbf{I} - \mathbf{B})$ is non-singular. In equation (8), $\boldsymbol{\Pi}_0$ is the vector of reduced-form intercepts, $\boldsymbol{\Pi}_1$ is the matrix of reduced-form slope coefficients, and $\boldsymbol{\zeta}^*$ is the vector of reduced-form disturbances, where $\text{Var}(\boldsymbol{\zeta}^*) = \boldsymbol{\Psi}^*$. Establishing the identification of the structural parameters requires determining if they can be solved uniquely from the reduced-form parameters (Fisher, 1966). An inspection of equation (8) reveals that the reduced form is nothing more than the multivariate general linear model. From here, equation (8) can be used to assess weak exogeneity. Specifically, from the context of the reduced form of the model, the parameters of the conditional model are $\boldsymbol{\omega}_1 \equiv (\boldsymbol{\Pi}_0, \boldsymbol{\Pi}_1, \boldsymbol{\Psi}^*)$, and the parameters of the marginal model are $\boldsymbol{\omega}_2 \equiv (\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\Sigma}_x)$, where $\boldsymbol{\mu}_x$ is the mean vector of $\mathbf{x}$, and $\boldsymbol{\Sigma}_\mathbf{x}$ is the covariance matrix of $\mathbf{x}$.

22.4. Assessing Weak Exogeneity

Recall that weak exogeneity concerns the extent to which the parameters of the marginal distribution of the exogenous variables are related to the parameters of the conditional distribution. In this section we consider three inextricably related ways in which the assumption of weak exogeneity can be violated: (a) violation of the joint normality of variables; (b) violation of the linearity assumption; and (c) violation of the assumption of homoskedastic errors.

22.4.1. Assessing Joint Normality

For simplicity, consider once again the simple linear regression model discussed in Section 22.2. It is known that within the class of elliptically symmetric multivariate distributions, the bivariate normal distribution possesses a conditional variance (*skedasticity*) that can be shown not to depend on the exogenous variables (Spanos, 1999). To see this, consider the bivariate normal distribution for two random variables y and x.

[5] See, for example, Bollen's (1989) definition discussed earlier.

The conditional and marginal densities of the bivariate normal distribution can be written respectively as

$$\begin{aligned}(y|x) &\cong N((\beta_0 + \beta_1 x), \sigma_u^2),\\ x &\cong N[\mu_x, \sigma_x^2],\\ \beta_0 &= \mu_y - \beta_1\mu_x, \quad \beta_1 = \frac{\sigma_{xy}}{\sigma_x^2},\\ \sigma_u^2 &= \sigma_y^2 - \left(\frac{\sigma_{xy}}{\sigma_x^2}\right)^2, \end{aligned} \tag{9}$$

where $\beta_0 + \beta_1\mu_x$ is the conditional mean of y given x, σ_u^2 is the conditional variance of y given x, μ_x is the marginal mean of x, and σ_x^2 is the marginal variance of x. Let

$$\begin{aligned}\boldsymbol{\theta} &= (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}),\\ \boldsymbol{\omega}_1 &= (\beta_0, \beta_1, \sigma_u^2),\\ \boldsymbol{\omega}_2 &= (\mu_x, \sigma_x^2). \end{aligned} \tag{10}$$

Note that for the bivariate normal distribution (and, by extension, the multivariate normal distribution), x is weakly exogenous for the estimation of the parameters in $\boldsymbol{\omega}_1$ because the parameters of the marginal distribution contained in the set $\boldsymbol{\omega}_2$ do not appear in the set of the parameters for the conditional distribution $\boldsymbol{\omega}_1$. In other words, the choice of values of the parameters in $\boldsymbol{\omega}_2$ does not restrict in any way the range of values that the parameters in $\boldsymbol{\omega}_1$ can take.

The bivariate normal distribution, as noted above, belongs to the class of elliptically symmetric distributions. Other distributions in this family include the Student's t, the logistic, and the Pearson Type III distributions. To demonstrate the problem with violating the assumption of bivariate normality, we can consider the case in which the joint distribution can be characterized by a bivariate Student's t-distribution (i.e., symmetric but leptokurtic). The conditional and marginal densities under the bivariate Student's t can be written as (see Spanos, 1999)

$$\begin{aligned}(y|x) \cong St\Bigg(&(\beta_0 + \beta_1 x),\\ &\frac{\nu\sigma_u^2}{\nu - 1}\left\{1 + \frac{1}{\nu\sigma_x^2}[x - \mu_x]^2\right\}\nu + 1\Bigg),\\ x \cong St&[\mu_x, \sigma_x^2; \nu], \end{aligned} \tag{11}$$

where ν are the degrees of freedom. Let

$$\begin{aligned}\boldsymbol{\theta} &= (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}),\\ \boldsymbol{\omega}_1 &= (\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2),\\ \boldsymbol{\omega}_2 &= (\mu_x, \sigma_x^2). \end{aligned} \tag{12}$$

Notice that the parameters of the marginal distribution $\boldsymbol{\omega}_2$ appear with the parameters of conditional distributions $\boldsymbol{\omega}_1$. Thus, by definition, x is not weakly exogenous for the estimation of the parameters in $\boldsymbol{\omega}_1$.

From this discussion, it is clear that one simple test of exogeneity is to assess the assumption of joint normality of y and x by using, say, Mardia's coefficient of multivariate skewness and kurtosis (Mardia, 1970). If the joint distribution is something other than normal, then parameter estimation must occur under the correct distributional form, and hence proper inferences may require estimation of the parameters of the marginal distribution as well as the conditional distribution. Because it is probably the case that joint normality does not hold in practice, this last point is extremely critical for the standard approach to statistical modeling in the behavioral sciences and will be taken up in more detail in Section 22.7.

22.4.2. Assessing the Assumption of Linearity

Joint normality of y and x is clearly central to establishing weak exogeneity. A consequence of the joint normality assumption is that the regression function $E(y|x, \theta) = \beta_0 + \beta_1' x$ is linear in x (Spanos, 1986). This follows from two properties of the normal distribution: (a) that a linear transformation of a normally distributed random variable is normal and (b) that a subset of normally distributed random variables is normal (Spanos, 1986). Therefore, deviations from linearity indirectly point to violations of normality and hence to violations of the weak exogeneity of x. Nonlinear relationships that cannot be transformed into linear relationships through well-behaved transformations will result in biased and inconsistent estimates of the parameters of the regression model. Assessing linearity can be accomplished through informal inspection of plots or more formally by using Kolmogorov-Gabor polynomials or the RESET method, both described in Spanos (1986). Should linearity be rejected, it may be possible to address the problem through normalizing transformations on y and/or x.

22.4.3. Assessing the Assumption of Homoskedastic Errors

The assumption of the joint normality of y and x also implies the assumption of homoskedastic errors. This is because, from the properties of the normal distribution, the conditional variance (skedasticity) function $\mathrm{Var}(y|x) = \sigma_y^2 - \sigma_{xy}^2/\sigma_x^2$ is free of x, where σ_{xy}^2 is the squared covariance of y and x. Thus, heteroskedasticity calls into question the assumption of weak exogeneity of x because it implies a relationship between the parameters of the marginal distribution and the conditional distribution. In addition, ordinary least squares estimation that ignores heteroskedasticity will result in unbiased but inefficient estimates of the regression coefficients. Most software packages contain easy-to-use options for obtaining residual scatter plots to assess the assumption of homoskedasticity. A direct test of the hypothesis of homoskedasticity was proposed by White (1980) and is available in many statistical software packages. Assessing the assumption of homoskedasticity in the context of structural equation modeling and multilevel modeling introduces additional complexities that will be addressed in Section 22.7.

22.5. Granger Noncausality and Strong Exogeneity

Our discussion of weak exogeneity in Section 22.3 did not specify a temporal structure for the data. Although the concept of weak exogeneity can be motivated by using models with lagged variables (Ericsson, 1994), it is not necessary to do so. The concept of weak exogeneity is applicable to cross-sectional data as well as to temporal data. However, to introduce the concepts of Granger noncausality and strong exogeneity, we must expand our models to account for the dynamic structure of the phenomenon under study. These extensions have important consequences for the statistical analysis of panel data when one wishes to properly model dynamic relationships and to use these models for forecasting or prediction.

To begin, consider an extension of our substantive problem of estimating the relationship between reading proficiency and parental involvement in reading activities. Let $\mathbf{z}_t$ be the vector of variables y_t and x_t. The basic problem now is that there is a dependence of current values of $\mathbf{z}$ on past values of $\mathbf{z}$, denoted as $\mathbf{z}_{t-1}$ with elements y_{t-1} and x_{t-1}. Therefore, the decomposition in equation (5) is no longer valid given the true dynamic structure of the process. Instead, we now need to condition on the past history of the process—namely,

$$f(\mathbf{z}_t|\mathbf{z}_{t-1};\mathbf{\Theta}). \tag{13}$$

The conditioning in equation (13) leads to a decomposition represented as a first-order vector autoregressive model of the form

$$\mathbf{z}_t = \boldsymbol{\pi}\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{14}$$

from which it follows that

$$y_t = \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 y_{t-1} + u_t, \tag{15}$$

$$x_t = \pi_1 x_{t-1} + \pi_2 y_{t-1} + \nu_t. \tag{16}$$

From our substantive perspective, equation (15) models current reading scores as a function of current and past parental involvement as well as past reading scores. Equation (16) models current parental involvement as a function of past parental involvement and past reading scores.

The above specification in equations (15) and (16) makes sense substantively insofar as feedback from previous reading scores might influence the amount of current parental involvement in reading activities. In other words, parents may notice improvement in their child's reading proficiency and feel reinforced for their reading activities. The question here, however, concerns whether parental involvement can be considered exogenous to reading proficiency and be used to predict future reading proficiency. In this case, we observe that weak exogeneity is not sufficient for the conditional model to be used to develop predictions of y because, as in our counterexample in Section 22.2, past values of y predict current values of x unless $\pi_2 = 0$. The condition that $\pi_2 = 0$ yields the condition of Granger noncausality (Granger, 1969). Granger noncausality essentially means that only lagged values of x enter into equation (15).

Weak exogeneity along with Granger noncausality yields the condition of *strong exogeneity.* The condition of strong exogeneity allows x_t (parental reading activities) to be treated as fixed at time t for the prediction of future values of y (reading proficiency) using the model in equation (15). Should Granger noncausality not hold (i.e., $\pi_2 \neq 0$), then valid prediction of future values of y would require the joint analysis of the conditional model in equation (15) and the marginal model in equation (16). In other words, the feedback inherent in the model when $\pi_2 \neq 0$ would have to be taken into account when interest centers on prediction.

22.5.1. Testing Strong Exogeneity and Granger Noncausality

Testing for strong exogeneity is relatively straightforward. First, it should be noted again that strong exogeneity requires weak exogeneity. Thus, if weak exogeneity does not hold, then neither does strong exogeneity. However, strong exogeneity also requires Granger noncausality. Thus, should y Granger cause x, then strong exogeneity does not hold. The simple test for Granger noncausality is given in equation (16), where the null hypothesis of Granger noncausality is given by $\pi_2 = 0$.[6]

22.6. Super Exogeneity

An important application of statistical models in the social and behavioral sciences is in the evaluation of interventions or policies related to the exogenous variables. For example, consider the question of the relationship between per pupil class time spent using Internet technology and classroom-level academic achievement. If interest centers on achievement as a function of time spent using Internet technology, then it is assumed that the parameters of the achievement equation (the conditional model) are invariant to changes in the parameters of the marginal distribution of classroom Internet access time.

One set of policies related to classroom Internet connections and access time may have to do with the so called *e-rate*. The e-rate initiative was put forth during the Clinton administration as a means of providing discounted telecommunication services to schools and libraries. A specific goal of the program was to ameliorate the so-called "digital divide" that separates suburban middle- to upper-middle-class schools from lower-middle-class and inner-city poor schools with respect to access to technology in the classroom. Changes in e-rate policy should, if successful, induce shifts in the distribution of classroom Internet connections. The question is whether a shift in the parameters of the marginal distribution of classroom Internet connections changes the fundamental relationship between the number of classroom Internet connections and classroom achievement.

Formally, invariance concerns the extent to which the parameters of the conditional distribution do not change when there are changes in the parameters of the marginal distribution. As pointed out by Ericsson (1994), *invariance* is not to be confused with *variation freeness*, as discussed under the topic of weak exogeneity. Using the e-rate example, let $\boldsymbol{\omega}_1$ be the parameters of the conditional model describing the relationship between classroom achievement and time spent on classroom Internet activities, and let $\boldsymbol{\omega}_2$ be the parameters of the marginal distribution of time spent on Internet activities. Following Engle and Hendry (1993), assume for simplicity that two scalar parameters are related via the function

$$\omega_{1t} = \varphi\omega_{2t}, \qquad (17)$$

where φ is an unknown scalar. Variation freeness suggests that over the period where ω_2 is constant, there is no information in ω_2 that is helpful in the estimation of ω_1. However, it can be seen that ω_1 is not invariant to changes in ω_2—that is, shifts in the parameters of the marginal distribution over some period of time lead to shifts in the parameters of the conditional distribution. By contrast, invariance implies that

$$\omega_1 = \varphi_t\omega_{2t}, \quad \forall t. \qquad (18)$$

In terms of our substantive example, equation (18) implies that changes in the parameters of the marginal distribution of classroom time spent on Internet activities due to, say, e-rate policy changes do not change its relationship to academic achievement. Invariance of these parameters, combined with the assumption of weak exogeneity, yields the condition of *super exogeneity.*[7]

22.6.1. Testing Super Exogeneity

There are two common tests for super exogeneity (Ericsson, 1994), but note that super exogeneity also requires that the assumption of weak exogeneity holds. Thus, if weak exogeneity is shown not to hold, then super exogeneity is refuted. The first of the two common tests for super exogeneity is to establish the constancy of $\boldsymbol{\omega}_1$ (the parameters of the conditional model) and the nonconstancy of $\boldsymbol{\omega}_2$ (the parameters of the marginal model). Parameter constancy simply means that the parameters of interest take on the same value over time. *Parameter constancy* is to be contrasted with *invariance* as discussed above, which refers to parameters that do not change as a function of changes in a policy or changes due to interventions.

Continuing, if the parameters of the conditional model remain constant regardless of the nonconstancy of the parameters of the marginal model, then super exogeneity holds. Methods for establishing constancy have been given by Chow (1960). Briefly, the Chow test requires deciding on a possible breakpoint of interest over the period of the analysis based on substantive considerations. Once that breakpoint is decided, then

[6] Clearly, this hypothesis will not hold exactly. Issues of power and the size of the alternative hypothesis $\pi_2 \neq 0$ become relevant as they pertain to the accuracy of forecasts when Granger noncausality does not hold.

[7] Strong exogeneity is not a precondition for super exogeneity (see Hendry, 1995).

a regression model for the series prior to and after the breakpoint is specified. Let β_1 and β_1 and $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$ be the regression coefficients and disturbance variances for the models before and after the breakpoint, respectively. The Chow test is essentially an F-type test of the form

$$CH = \left(\frac{RSS_T - RSS_1 - RSS_2}{RSS_1 + RSS_2}\right)\left(\frac{T - 2k}{k}\right), \quad (19)$$

where T is the number of time periods, k is the number of regressors, and RSS_T, RSS_1, and RSS_2 are the residual sum of squares for the total sample period, subperiod 1, and subperiod 2, respectively. The test in equation (19) can be used to test $\mathrm{H}_0 : \beta_1 = \beta_2$ and $\sigma^2_{u_1} = \sigma^2_{u_2}$ and is distributed under H_0 as $CH \approx F(k, T - 2k)$. Limitations with the Chow test have been discussed in Spanos (1986).

The second test extends beyond the first in the following way. Here, the goal is to model the marginal process in such a way as to render it empirically constant over time (Ericsson, 1994). This can be accomplished by adding dummy variables that account for "seasonal" changes or interventions occurring over time in the marginal process. This exercise amounts to changing or intervening with the marginal process. Once these additional variables are shown to render the marginal model constant, they are then added to the conditional model. If the variables that rendered the marginal model constant are found to be nonsignificant in the conditional model, then this demonstrates the invariance of the conditional model to changes in the process of the marginal model (Engle & Hendry, 1993; Ericsson, 1994).

Returning to the e-rate example, consider the simple model that relates the number of Internet connections to academic achievement. Here we wish to test super exogeneity because we would like to use the measure of Internet connections as a policy variable for forecasting changes in academic achievement as a function of changes in the number of Internet connections over time. To begin, we must test for the weak exogeneity of the number of Internet connections because weak exogeneity is necessary for super exogeneity to hold. Next, we would use, for example, a Chow test to establish the constancy of the conditional model parameters of interest to the nonconstancy of the marginal parameters. This is then followed by developing a model for the change in the number of Internet connections over time, by adding variables that describe this change. These could be dummy variables that measure points in time in which the e-rate policy was enacted or other variables that would describe how the average number of Internet connections in the classroom would have changed over time. These variables are then added to the model relating achievement to the number of Internet connections. Should these new variables be nonsignificant in the conditional model, then this demonstrates how the parameters relating achievement to the number of Internet connections are invariant to changes in the parameters of the marginal model.

22.6.2. An Aside: Inverted Regression and Super Exogeneity

Consider the hypothetical situation in which an investigator wishes to regress science achievement scores on attitudes toward science, both measured on a sample of eighth-grade students using the model in equation (1). Suppose further that both sets of scores are reliable and valid and that, for the sake of this example, the attitude measure is super exogenous for the achievement equation. This implies that the measure of attitudes toward science satisfies the assumption of weak exogeneity and that the parameters of interest are constant and invariant to changes in the marginal distribution of attitudes toward science. Now, suppose the investigator wishes to change the question and estimate the regression of attitudes toward science on science achievement scores. In this case, it would be a simple matter of inverting the regression coefficient, obtaining $1/\beta$ as the inverted regression coefficient. The question is whether the inverted model still retains the property of super exogeneity.

To answer this question, we need to consider the density function for the inverted model. Following Ericsson (1994), let the bivariate density for the inverted regression model of two random variables x and y be defined as

$$\begin{aligned} (x_t | y_t) &\approx N[(c + \delta y_t, \tau^2)], \\ y_t &\approx N(\mu_y, \sigma^2_y), \end{aligned} \quad (20)$$

where $\delta = \sigma_{xy}/\sigma^2_y$, $c = \mu_x - \pi\mu_y$, and $\tau^2 = \sigma^2_x - \sigma^2_{xy}/\sigma_y$. The model in equation (20) can be expressed in model form as

$$\begin{aligned} x_t &= c + \delta y_t + \nu_{2t} & \nu_{2t} &\approx N(0, \tau^2), \\ y_t &= \mu_y + \varepsilon_{yt} & \varepsilon_{yt} &\approx N(0, \sigma^2_y), \end{aligned} \quad (21)$$

where the usual regression assumptions hold for this model. When equation (20) is written in line with the factorization of density functions, the result is the form

$$F(\mathbf{z}_t | \boldsymbol{\theta}) = F_{x|y}(x_t | y_t, \boldsymbol{\varphi}_1) F_y(y_t | \boldsymbol{\varphi}_2), \quad (22)$$

where $\boldsymbol{\varphi} \equiv (\varphi_1', \varphi_2') = h(\boldsymbol{\theta})$, a one-to-one function. To see the problem with inverted regression, we need to recognize that there is a one-to-one mapping between the parameters of the un-inverted model $\boldsymbol{\omega}$ from Section 22.3 and the inverted model. Specifically, we note that because $\beta = \sigma_{xy}/\sigma_x^2$ and $\sigma_u^2 = \sigma_y^2 - \sigma_{xy}^2/\sigma_x^2$, then after some algebra, it can be shown that

$$\delta = \frac{\beta\sigma_x^2}{\tau^2 + \beta^2\sigma_x^2}. \qquad (23)$$

It can be seen from equation (23) that $\delta \neq 1/\beta$ unless $\sigma_u^2 = 0$. Moreover, from Ericsson (1994), we note that if x_t is super exogenous for β and σ_u^2, then even if β is constant, δ will vary due to variation in the marginal process of x_t via the parameter σ_x^2. In other words, super exogeneity is violated because the parameters of the inverted model are nonconstant even when the parameters of the uninverted model are constant (Ericsson, 1994, p. 18).

22.6.3. Super Exogeneity, the Lucas Critique, and Their Relevance for the Social and Behavioral Sciences

Super exogeneity plays an important philosophical role in economics and economic policy analysis. Specifically, super exogeneity protects economic policy analysis from the so-called "Lucas critique." It is beyond the scope of this chapter to delve into the history and details of the Lucas critique. Suffice to say that the Lucas critique concerns the use of econometric models for policy analysis because econometric models contain information that changes as a function of changes in the very phenomenon under study. The following quote of Lucas (1976) illustrates the problem:

> Given that the structure of an econometric model consists of optimal decision rules for economic agents, and that optimal decision rules vary systematically with changes in the structure of the series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models. (quoted in Hendry, 1995, p. 529)

In other words, "a model cannot be used for policy if implementing the policy would change the model on which that policy was based, since then the outcome of the policy would not be what the model had predicted" (Hendry, 1995, p. 172).

The types of models considered in econometric policy analysis differ in important ways from those considered in the other social and behavioral sciences. For example, typical models used in, say, sociology or education do not consist of specific representations of the optimal decision behavior of "agents" and so do not lend themselves to the exact problem described by the Lucas critique. Also, models used in the social and behavioral sciences do not specify "technical" equations of the output of the system under investigation. Nevertheless, because the Lucas critique fundamentally suggests a denial of the property of invariance (Hendry, 1995), it may still be relevant to models used for policy analysis in domains other than economics. For instance, returning to the example of the e-rate and its role in educational achievement, the Lucas critique would claim that the parameters representing the relationship between Internet connections and educational achievement are not invariant to changes in the marginal process induced by the e-rate policy. However, tests of super exogeneity outlined above are tests of the Lucas critique, and so it is possible to empirically evaluate the seriousness of this problem for policy analysis.

22.7. Summary and Implications

Close examination of typical social and behavioral science definitions of exogenous variables shows that they are fraught with ambiguities. Yet, exogeneity is clearly of such vital importance to applied statistical modeling that a much more rigorous conceptualization of the problem is required, including guidance as to methods of testing exogeneity. The purpose of this chapter was to provide a didactic introduction to econometric notions of exogeneity, motivating these concepts from the standpoint of simple linear regression and its extension to structural equation modeling. The problem of exogeneity, as developed in the econometrics literature, provides a depth of conceptualization and rigor that is argued in this chapter to be of value to the other social and behavioral sciences.

To summarize, each form of exogeneity relates to a particular use of a statistical model. Table 22.1 reviews the different forms of exogeneity, their specific requirements, and informal tests. To review, weak exogeneity relates to the use of a model for purposes of inference. It concerns the extent to which the parameters of the marginal distribution of the exogenous variable can be ignored when focusing on the conditional distribution of the endogenous variable given the exogenous variable. Should weak exogeneity not hold, then estimation must account for both the marginal and conditional distributions. Strong exogeneity

Table 22.1 Summary of Different Forms of Exogeneity

Form of Exogeneity	*Implications for*	*Assumptions*	*Informal/Formal Tests*
Weak exogeneity	Inference	Multivariate normality of the joint distribution; homoskedasticity; linearity	Mardia's measures; homoskedasticity and linearity tests
Strong exogeneity	Forecasting and prediction	Weak exogeneity and Granger noncausality	Weak exogeneity tests; test of coefficient on lagged endogenous variable (see equation (16))
Super exogeneity	Policy analysis	Weak exogeneity, parameter constancy, and parameter invariance	Chow test; nonsignificance in conditional model of variables that describe policy changes in the marginal model

supplements the requirement of weak exogeneity with the notion of Granger noncausality so that exogenous variables can be treated as fixed for purposes of forecasting and prediction. Should Granger noncausality not hold, then prediction and forecasting must account for the dynamic structure underlying the exogenous variables. Super exogeneity requires weak exogeneity to hold and concerns the invariance of the parameters of the conditional distribution given real-world changes in the parameters of the marginal distribution. If an intervention or policy leads to changes in the distribution of the marginal process but does not change the relationship described by the conditional model, then the exogenous variable is super exogenous for policy or intervention analysis.

22.7.1. Implications for Standard Statistical Practice

The impact of the exogeneity assumption on standard statistical practice in the social and behavioral sciences is profound. To begin, it is clear that the exogeneity problem is not unique to linear regression and structural equation models. Indeed, the problem is present in all statistical models in which a distinction is made between exogenous and endogenous variables, resulting in a partitioning of the joint distribution into the conditional and marginal distributions.

It is worth considering briefly how the problem of exogeneity might arise in other statistical models. Here we consider *multilevel modeling* (including growth curve modeling), a methodology that is enjoying widespread popularity in the social and behavioral sciences (see, e.g., Raudenbush & Bryk, 2002). Multilevel modeling is a powerful analytic methodology for the study of hierarchically organized social systems such as schools or businesses. In education, for example, multilevel modeling has yielded a much greater understanding of the organizational structure of schools as they support student learning. In this methodology, the so-called "Level 1" variables constitute endogenous outcomes such as student achievement that can be modeled as a function of student-level exogenous variables. Parameters of the Level 1 model include the intercept and the slope(s) that are allowed to vary over so-called "Level 2" units such as classrooms. Classroom level variation in the Level 1 coefficients can be modeled as a function of classroom exogenous variables such as measures of the amount of teacher training in specific subject matter skills. Variation over Level 3 units such as schools is also possible, and school-level variables can be included to explain this component of variation.

Future research should examine the problem of exogeneity in multilevel models. Suffice to say here that exogeneity enters into multilevel models at each level of the system. Statistical theory underlying multilevel modeling shows that these models have built-in heteroskedasticity problems that are resolved by specialized estimation methods. Yet, what remains to be determined is if the parameters of interest in multilevel models can be shown to be variation free with respect to the parameters of the student-level and school-level exogenous variables. Because multilevel models are used to supplement important discussions of education policy, assessing the weak exogeneity of policy-relevant variables is crucial.

A special case of multilevel modeling is *growth curve modeling,* a methodology that is also enjoying tremendous popularity in the social sciences and directly accounts for the dynamic features of panel data. In such models, the Level 1 endogenous variable is an outcome such as a reading proficiency score for a particular student measured over multiple occasions. This score is modeled as a function of a time dimension such as grade level, as well as possibly time-varying covariates such as parent involvement in reading activities. The parameters of the Level 1 model constitute the initial level and rate of change, and these are allowed to vary randomly over individuals, who are in turn modeled as a function of time-invariant exogenous variables such as race/ethnicity, gender, or perhaps experience in an early childhood intervention program. Variation in average initial level and rate of change can also be modeled as a function of Level 3 units such as classrooms or schools. The power of this methodology is that it allows one to study individual and group contributions to individual growth over time.

The problem of exogeneity enters growth curve models in a variety of ways. First, repeated measures on individuals can be a function of time-invariant variables. For example, in estimating growth in reading proficiency in the younger grades, time-invariant variables might include the IQ of the children (assumed to be stable over time), the income of the parents, and so on. Again, these variables are assumed to be exogenous.

Second, the repeated outcomes can be modeled as a function of time-varying covariates. Each time-varying covariate is presumed to be exogenous to its respective outcomes and is used to help explain, for example, seasonal trends in the data. However, time-varying variables can also be allowed to have a lagged effect on later outcomes. For example, a time-varying covariate such as parental reading activities at time t can be specified to influence reading achievement at time t as well as reading achievement at time $t + 1$. This represents the introduction of a lagged exogenous variable into the full-growth curve model, and so issues of strong exogeneity and Granger noncausality may be of relevance. In other words, the Level 1 model that characterizes achievement at time t as a function of time-varying covariates assumes that the time-varying covariate at time t is not a function of achievement at time $t - 1$. If this assumption does not hold, then the time-varying covariate is not strongly exogenous.

In addition to the fact that exogeneity represents an issue in a wide range of statistical models, it must also be recognized that most statistical software packages estimate the parameters of statistical models under the untested assumption that weak exogeneity holds. In other words, software packages that engage in *conditional estimation* (e.g., conditional maximum likelihood), conditional on the set of exogenous variables, do so assuming that there is no information in the marginal process that is relevant for the estimation of the conditional parameters. However, as noted above, weak exogeneity is only valid if the joint distribution of the variables is multivariate normal—a heroic assumption at best. Therefore, it is likely in practice that estimates derived under conditional estimation are incorrect. The only situation in which this is not a problem is in estimation of the Gauss linear model with nonstochastic regressors. Future research and software development should explore methods of estimation that account for the parameters of the marginal distribution along with the conditional distribution for a given specification of the form of the joint distribution of the data.

In the context of simple linear regression, informal testing of weak exogeneity via assessing joint normality and homoskedasticity is relatively straightforward. Indeed, most standard statistical software packages provide various direct and indirect tests of these assumptions. In the context of structural equation modeling, however, although considerable attention has been paid to the normality assumption (see, e.g., Kaplan, 2000, for a review), scant attention has been paid to assessing assumptions of linearity and homoskedasticity. This may be due to the fact that textbook treatments of structural equation modeling motivate the methodology from the viewpoint of the structural form of the model, and therefore it is not directly obvious how homoskedasticity could be assessed. However, if attention turns to the reduced form of the model as described in equation (8), then standard methods for assessing the normality assumption—including homoskedasticity and linearity—would be relatively easy to implement. Therefore, users of structural equation modeling should be encouraged to study plots and other diagnostics associated with the multivariate linear model to assess weak exogeneity.

The issue raised here is not so much how to assess weak exogeneity but rather how to proceed if the assumption of weak exogeneity does not hold. Recognition of the seriousness of the exogeneity assumption should lead to fruitful research that focuses on estimation methods under alternative specifications of the joint distribution of the data. In attempting to characterize the joint distribution of the data, all means of data exploration should be encouraged. There should be no concern about “finding a model in the data” because the joint distribution of the data is theory free[8] (Spanos, 1986). Theory information only becomes a problem when there is a factoring of the joint distribution into the conditional and marginal distributions insofar as that is the point in the modeling process, in which a substantive distinction is made between endogenous and exogenous variables and where parameters of interest are defined (see Spanos, 1999).

The implications of the strong exogeneity assumption for statistical practice are relevant if models are used for prediction and forecasting. In this case, weak exogeneity is still a necessary requirement, but in addition, it is imperative that Granger noncausality be established. Similarly, implications of the super-exogeneity assumption are relevant when models are used for policy or intervention evaluations. Super exogeneity also forces us to consider the requirement of parameter constancy and invariance—issues that have not received as much attention in the social and behavioral sciences as they should. Focusing on parameter constancy and invariance also forces us to consider whether there exist invariants in social and behavioral processes. Moreover, as pointed out by Ericsson (1994), parameter constancy is a central assumption of most estimation methods and hence is of vital importance to statistics generally.

22.7.2. Concluding Remarks

Our discussion throughout this chapter leads to the recognition that *exogeneity* is an adjective describing an assumed characteristic of a variable that is being chosen for theoretical reasons to be an exogenous variable. Weak exogeneity is the necessary condition underlying all forms of exogeneity, and hence this assumption is fundamental and requires empirical confirmation to ensure valid inferences. Additional assumptions are required to yield valid predictions or evaluations of policies or interventions.

Exogeneity resides at the nexus of the actual data-generating process (DGP) and the statistical model used to understand that process. In the simplest terms, the actual DGP is the real-life mechanism that generated the observed data. It is the reference point for both the theory and the statistical model. In the former case, the theory is put forth to explain the reality under investigation—for example, the organizational structure of schooling that generates student achievement. In the latter case, the statistical model is designed to capture the statistical features of that aspect of the actual DGP that we choose to study and measure (Spanos, 1986; see also Kaplan, 2000).

In addition to the role that exogeneity plays with regard to fundamental distinctions between theory, the DGP, and statistical models, exogeneity raises a number of other important philosophical questions that are central to the practice of statistical modeling in the social and behavioral sciences. One issue, for example, concerns the proper place of data mining as a premodeling strategy. We find that when attention focuses on characterizing the joint distribution of the data, then data mining has a central role to play. Another issue arising from our study of exogeneity concerns the dynamic reality of the phenomenon under investigation. Granger noncausality and strong exogeneity force us to consider exogenous variables as possibly being responsive to their own dynamic structure and that this must be correctly modeled to obtain accurate estimates for prediction and forecasting. Super exogeneity reminds us that our models are sensitive to real-life changes in the process under investigation. Finally, serious consideration of the problem of exogeneity forces us to reexamine statistical textbooks in the social and behavioral sciences to clarify ambiguous concepts and historical developments. It is hoped that reflecting on the importance of the exogeneity assumption will lead to a critical assessment of the methods of statistical modeling in the social and behavioral sciences.

References

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62–83.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28,* 591–605.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/ correlation for the behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum.

Engle, R. F., & Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics, 56,* 119–139.

[8] The exception being that theory enters into the choice of the variable set as well as methods of measurement. These issues are not trivial but are not central to our discussion of the role of theory as it pertains to the separation of variables into endogenous and exogenous variables.

Ericsson, N. R. (1994). Testing exogeneity: An introduction. In N. R. Ericsson & J. S. Irons (Eds.), *Testing exogeneity* (pp. 3–38). Oxford, UK: Oxford University Press.

Ericsson, N. R., & Irons, J. S. (Eds.). (1994). *Testing exogeneity.* Oxford, UK: Oxford University Press.

Fisher, F. (1966). *The identification problem in econometrics.* New York: McGraw-Hill.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37,* 424–438.

Hendry, D. F. (1995). *Dynamic econometrics.* Oxford, UK: Oxford University Press.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Academic Press.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Thousand Oaks, CA: Sage.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole.

Legendre, A. M. (1805). *Nouvelles méthods pour la détermination des orbites des comètes* (New methods for determining the orbits of comets). Paris: Firmin Didot.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Journal of Monetary Economics, 1*(Suppl.), 19–46.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57,* 519–530.

Mulaik, S. A. (1985). Exploratory statistics and empiricism. *Philosophy of Science, 52,* 410–430.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousands Oaks, CA: Sage.

Richard, J.-F. (1982). Exogeneity, causality, and structural invariance in econometric modeling. In G. C. Chow & P. Corsi (Eds.), *Evaluating the reliability of macro-economic models* (pp. 105–118). New York: John Wiley.

Spanos, A. (1986). *Statistical foundations of econometric modeling.* Cambridge, UK: Cambridge University Press.

Spanos, A. (1999). *Probability theory and statistical inference.* Cambridge, UK: Cambridge University Press.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48,* 817–838.

Wonnacott, R. J., & Wonnacott, T. H. (1979). *Econometrics* (2nd ed.). New York: John Wiley.

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society, 60,* 812–854.